

## **Facilitando el acceso al contenido del patrimonio cultural en Chequia: Archivos de Autoridad Nacionales y proyecto INTERMI**

*Spanish translation of the original paper: “Facilitating access to cultural heritage content in Czechia: National Authority Files and INTERMI project”.*

*Translated by: Jorge Fernández García, Biblioteca Nacional de España ,Madrid, Spain.*

*The text of this document has been translated into Spanish and differences from the original text may occur. This translation is provided for reference purposes only.*

*El texto de este documento ha sido traducido al español y puede haber diferencias con el texto original. Se proporciona esta traducción con el propósito de servir de referencia exclusivamente.*

### **Marie Balíková**

National Library of the Czech Republic, Prague, Czechia

E-mail address: Marie.Balikova@nkp.cz

### **Miroslav Kunt**

National Archives, Prague, Czechia.

E-mail address: Miroslav.Kunt@nacr.cz

### **Jana Šubová**

Cosmotron Bohemia, s. r. o., Hodonín, Czechia

E-mail address: subova@cosmotron.cz

### **Nadežda Andrejčíková**

Cosmotron Bohemia, s. r. o., Hodonín, Czechia

E-mail address: andrejcikova@cosmotron.cz



This is a Spanish translation of “ Facilitating access to cultural heritage content in Czechia: National Authority Files and INTERMI project” copyright © 2015 by Jorge Fernández García. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

---

## **Abstract:**

*El artículo informa sobre el proyecto Interoperabilidad en Instituciones de la Memoria (INTERMI) de Chequia realizado por la Biblioteca y el Archivo Nacional de la República Checa dentro del Programa de investigación aplicada y desarrollo de identidad cultural y nacional (NAKI) financiado por el Ministerio de Cultura de la República Checa. El propósito del proyecto es crear infraestructura científica y tecnológica para apoyar el procesamiento, y uso/reutilización del contenido del Patrimonio Cultural y poder compartirlo en forma de metadatos sobre objetos de información preservados en instituciones de la memoria, esto significa crear un modelo conceptual de conocimiento que reúna los criterios y necesidades de los usuarios de las instituciones de la memoria en Chequia. El proyecto está basado en un nuevo paradigma, que está centrado en el procesamiento de entidades ampliado por información semántica necesaria y en la manifestación de complejas relaciones entre ellas. El papel de los Archivos de Autoridad Nacional en la base de conocimientos de INTERMI será evaluada.*

**Keywords:** instituciones de la memoria, autoridades nacionales, patrimonio cultural, bases de conocimientos, modelos conceptuales, entidades, enfoque centrado en eventos

---

## **Introducción y contexto**

Bibliotecas, archivos, galerías y museos como instituciones de la memoria recopilan, organizan y ponen a disposición objetos/recursos de información en el ámbito del patrimonio nacional cultural, industrial, científico, y natural. El ámbito del patrimonio Cultural incluye también patrimonio documental y de archivos.

Ahora, en el periodo de la digitalización masiva, el principal objetivo de las tareas en las instituciones de la memoria es dar apoyo efectivo a la selección, agregación y fácil uso/reutilización del contenido digital.

En el Proyecto INTERMI, este objetivo común depende de la especificación de necesidades de la comunidad de usuarios, de la calidad de los metadatos relacionados a los objetos de información (identificación inequívoca de objetos en el nivel de metadatos incluidos), de la aplicación de los principios de la programación orientada a objetos y el enfoque centrado en eventos, de la aplicación de tecnologías de interoperabilidad semántica, lo que significa en el modelo conceptual y estructura aplicada de datos, de herramientas formales para la representación adecuada al entorno de la Web semántica (e.g. XML, RDF, OWL), y de la aplicación de los principios de Linked Open Data.

## **Especificaciones de las necesidades de la comunidad de usuarios y sus expectativas del proyecto INTERMI.**

Los usuarios de las instituciones de la memoria en Chequia (público general, catalogadores, indizadores, especialistas de la información, conservadores, archiveros y profesionales de todos los ámbitos) apoyan la creación y desarrollo de la infraestructura para la base de datos del Patrimonio Cultural; prefieren hacer disponible el contenido del patrimonio Cultural en la Web tanto legible para los humanos como en forma comprensible para las máquinas, significa compartir el Patrimonio Cultural de acuerdo a la Web Semántica

y a los principios de Linked Open Data. Son conscientes de lo que supone el desarrollo y la aplicación de normas, reglas y esquemas de metadatos de gran calidad y apropiados a la Web Semántica. Necesitan más información sobre entidades en los actuales Registros de Autoridad y están de acuerdo que es necesario pasar las autoridades a entidades de valor en la web, de términos a conceptos.

### **Objetos de información y su identificación a nivel de metadatos**

En el proyecto INTERMI, los objetos de información representan las entidades del mundo real, e.g. personas, instituciones, objetos tridimensionales, (i.e. objetos artísticos y técnicos, objetos de naturaleza inanimada), actividades, eventos y representaciones artísticas y otros logros, lugares etc. que representan temas para su colección de documentos y son materia para sus actividades. A la vez, la identificación de los objetos de información en el nivel de metadatos está basada tanto en nombres como autoridades de materia.

### **El papel de los ficheros de Autoridad Nacional en poner disponible el contenido digital en Chequia.**

Los ficheros de Autoridad Nacional contienen ficheros de nombres personales (639472 registros), ficheros de entidades y congresos (140002), Título Uniforme (2454), ficheros de términos de materia (36954), ficheros de nombres geográficos (27170) y fichero Género/forma (1783).

Los ficheros Nacionales de Autoridad fueron desarrollados originalmente con el objetivo de apoyar la catalogación compartida y la indexación de materiales bibliotecarios. Los nombres de autoridad fueron creados de acuerdo a las reglas AACR2 mientras las autoridades de materia fueron originalmente creadas de acuerdo a las reglas LCSH<sup>1</sup>.

Más tarde, con la llegada de la WWW, algunos nuevos principios fueron introducidos: post coordinación, autorización de unidades léxicas aisladas (no cadenas completas de encabezamientos de materia), la aplicación limitada de cadenas de encabezamientos de materia en registros bibliográficos. El principal objetivo de los registros de autoridad ha sido establecer formas autorizadas de nombres/denominaciones, sus variantes y evidencias de estas formas; han sido codificadas usando los formatos MARC.

### **Fortalezas de los ficheros de autoridad tradicionales.**

En los ficheros de autoridad tradicionales, se establecen puntos de acceso normalizados, contienen las mejores prácticas bien documentadas y están compartidas entre bibliotecas, así como otras instituciones. Las bibliotecas en Chequia participan en “Creación cooperativa y uso de ficheros de autoridad nacional” mientras los museos y galerías participan en el proyecto “Autoridades de Museo” y el proyecto “Registro de colecciones de Bellas Artes”. Todos estos proyectos están basados en las autoridades nacionales; autoridades de museo y el Registro de colecciones de Bellas Artes se limitan al nombre (personal y corporativo) de autoridades.

---

<sup>1</sup> Desde Mayo, 2015 han sido introducidas las reglas RDA

## **Debilidades de los ficheros de autoridad tradicionales**

### **Sistema Basado en registro**

Los ficheros tradicionales de autoridad de bibliotecas están basados en sistemas basados en registros. Están creados por encima de todo para los humanos, son comprensibles para los humanos, pero no están en formato comprensible para la máquina, sólo son legibles para la máquina.

### **Falta de granularidad de datos**

AACR2 y el formato MARC no permiten la suficiente granularidad de la información codificada. El formato MARC no ofrece suficiente granularidad en los elementos de datos que deben ser claramente definidos por la web semántica y propósitos LOD. “El registro MARC no fue creado como un conjunto de elementos de datos sino como un formato para el almacenamiento y visualización de texto de los registros del catálogo de la biblioteca.”<sup>2</sup>

### **Entidades del mismo tipo en diferentes ficheros de autoridad**

Otro problema es que algunas entidades son clasificadas en varios/diferentes ficheros de autoridad en relación directa con los campos MARC en los que han sido codificados. Algunas entidades como las actividades forman parte de dos ficheros de autoridad: fichero de congresos (seminarios, congresos, talleres) y fichero de términos de materia (guerras, batallas, etc.). Los edificios renombrados, forma parte del fichero de nombres corporativos. Los no nombrados están clasificados en el fichero de términos de materia. Estas entidades deberían estar clasificadas de acuerdo a sus atributos y a su capacidad para expresar las relaciones con otras entidades más que de acuerdo al campo MARC.

### **Poca información en los registros de autoridad**

Los registros de autoridad contienen poca información sobre la entidad: están concentradas en las formas preferidas de entidades nombre/denominación introducidas como encabezamientos principales y sus variantes introducidas como ver referencias. No hubo posibilidad para introducir toda la información necesaria para expresar propiedades/atributos y relaciones de entidad de una manera estructurada. Al contrario, era posible mencionar alguna información en campos de nota solo como texto libre.

### **Término describiendo tipo de entidad corporativa**

Para los encabezamientos de entidades corporativas, no es posible indicar un término para describir el tipo de entidad (e.g.: Ltd., pc ), sin embargo para los archiveros el tipo de entidad corporativa es uno de los atributos más importantes que distingue una entidad individual de otras similares.

## **Ficheros de autoridad nacional aplicados en el Proyecto INTERMI.**

Las autoridades nacionales fueron analizadas en profundidad y se decidió usarlas de acuerdo a los siguientes principios:

- Los puntos de acceso normalizados deben ser usados como formas preferentes o variantes de las denominaciones de las entidades.

---

<sup>2</sup> <http://futurelib.pbworks.com/w/page/29114548/MARC%20elements>

- Toda la información que proporciona contexto a las entidades en los registros de autoridad debe ser aceptada en el sistema INTERMI así como introducida de forma estructurada.
- La tendencia a tomar decisiones basadas puramente en el formato (formato Marc) deben ser evitadas siempre que sean posibles.
- El modelo conceptual propietario y el formato de estructura de datos tienen que ser creados para reunir todas las necesidades específicas de los usuarios de la comunidad.
- El modelo conceptual INTERMI debe estar basado en entidades (no en autoridades), significa que habrá una fuerte tendencia a desplazar de las autoridades a las entidades de valor en la web.

## **Modelo conceptual INTERMI**

El modelo conceptual INTERMI está basado en algunos principios del paradigma orientado a objetos, enfoque centrado en eventos y está diseñado como convertible a CIDOC CRM.

### **Principios del paradigma orientado a objetos**

- Granularidad de datos- los datos están divididos en elementos significativos más pequeños que permiten dividir mejor el tipo de datos para generalizarlos, así se apoya la flexibilidad de datos.
- Herencia de datos- la estructura jerárquica de los elementos necesita y soporta la herencia de atributos o propiedades.
- Reutilización de datos- los elementos están diseñados para ser reutilizados en más de una ubicación en el modelo y para permitir una mayor expansión del modelo.
- Polimorfismo- los elementos pueden funcionar como diferentes elementos dependiendo de su uso.

### **Enfoque centrado en eventos**

Los eventos representan acciones específicas que suceden en un tiempo y lugar específicos y son codificadas como objetos de información. Coincidimos que “la modelización de eventos es tan abstracta que puede ser usada para describir obras culturales y documentación de observaciones científicas”<sup>3</sup>

Existen dos tipos de eventos en el modelo conceptual INTERMI: eventos como clases de entidades y eventos que son usados para describir atributos complejos y relaciones describiendo otras entidades, autodenominadas “entidades descriptivas”. Eventos (clases de entidades) incluye un evento temporalmente nombrado (largo plazo o corto plazo; una vez o

---

<sup>3</sup> <http://www.cidoc-crm.org/docs/fin-paper.pdf>

repetido) y también entidades nombradas que están relacionadas con la actividad humana tales como la cultura, costumbres. “Entidades descriptivas” están conectadas a eventos en la historia o vida de la entidad (e.g. nacimiento, instauración, creación, destrucción).

El enfoque centrado en eventos significa que todos los atributos de entidad que sean posibles sean descritos como evento; pensamos que el enfoque centrado en eventos refleja el concepto de Linked Data.

## **Tecnologías de la información aplicadas en el proyecto INTERMI**

El principal objetivo del proyecto INTERMI es crear bases de datos de conocimiento y una ontología para objetos de información que suceden en el patrimonio cultural. La necesidad de afrontar cuestiones tecnológicas es evidente, por lo tanto tenemos que tratar con dos principales problemas:

- ¿Cómo representar los datos cubiertos por el modelo conceptual INTERMI y proporcionar la preservación de estos datos? Significa encontrar un modelo de datos internos para su almacenamiento que apoye la apertura del modelo conceptual INTERMI, rápido acceso y presentación flexible. Decidimos usar una estructura propia basada en XML que soporte las relaciones entre entidades. La base para INTERMI también contiene una poderosa base de datos de objetos.
- ¿Cómo representar los datos para sistemas externos (sistemas de gestión bibliotecaria, sistemas de gestión de datos usados en archivos y museos así como la Web)? Conceptos de Web semántica y Linked Data representan espacio adecuado para INTERMI debido a que ellos proveen tecnologías para la representación de datos en contexto no solo para la representación de contenido.

Tecnologías de web semántica consisten en :

- URI (identificador uniforme de recursos) usado para la identificación de objetos y conceptos abstractos ;
- RDF y RDF SCHEMA para la representación de información usando un modelo abstracto de metadatos, RDF permite el registro de declaraciones simples con tripletas (materia-predicado-objeto);
- OWL- lenguaje para creación de ontologías;
- SPARQL- lenguaje de consulta.

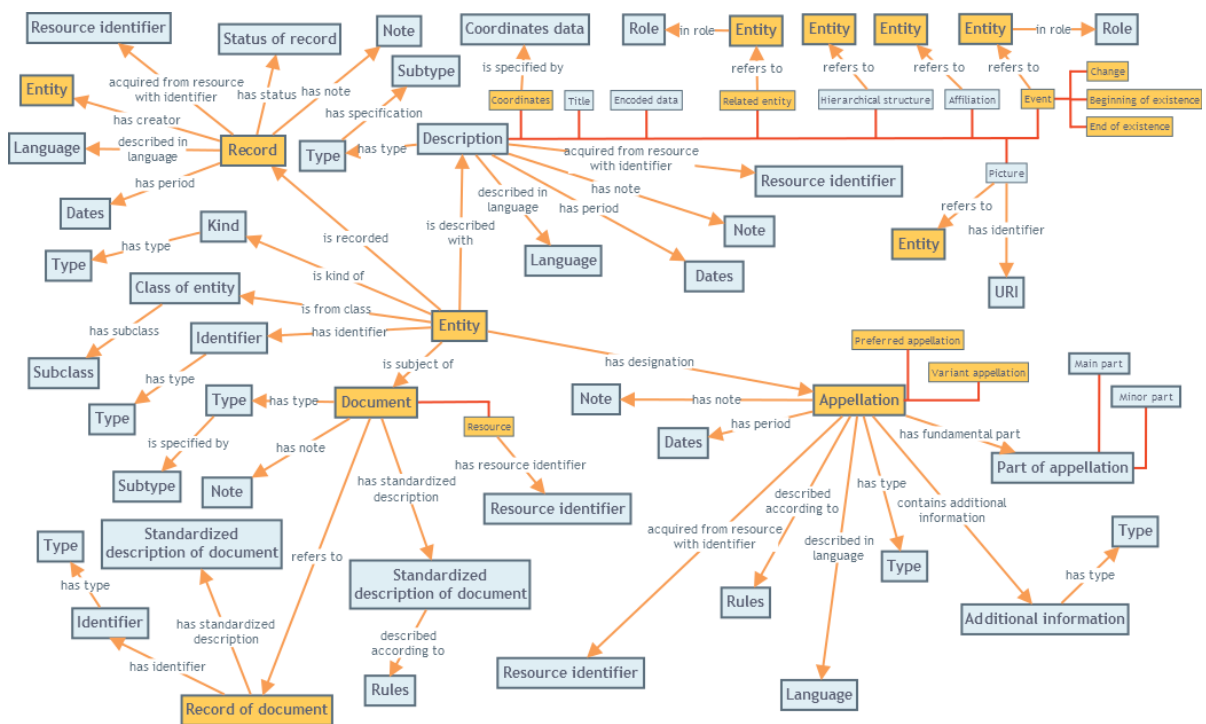
Varias normas relacionadas con el concepto de Linked Data y la Web Semántica fueron consideradas para buscar una o más normas adecuadas a la representación de datos. Como en el proyecto INTERMI, las siguientes normas deben ser mencionadas:

- SKOS- para la representación del tesoro- pueden ser aplicadas en entidades del tipo “Concepto General”
- FOAF- ontología para describir personas, actividades y relaciones- pueden ser entidades del tipo “Persona/criatura”.

Para mantener la continuidad, es necesario proveer datos en la estructura actual de la base de conocimiento INTERMI que se use frecuentemente en bibliotecas. – es decir, MARC21/XML para autoridades.

La presentación de datos INTERMI en diferentes estructuras específicas requiere una selección de datos que puedan ser usados en estructuras específicas (o normas). Esto también necesita más transformación de datos de la estructura interna a la externa.

Para los sistemas externos, el acceso a datos INTERMI será provisto usando servicios web (e.g. log –in de usuario, buscar registro, capturar registro).



**Figura 1 Modelo Conceptual INTERMI**

### Elementos clave del modelo conceptual INTERMI: entidades y relaciones

Entidad es el elemento más importante del modelo conceptual de INTERMI.

Las entidades representan objetos identificables en el mundo real. Una entidad es un concepto (individual y general) que tiene un identificador, una entidad tipo que describe la categoría de la entidad y es definida por un conjunto de características que permiten describir cualquier entidad usada en el modelo. Una de las tareas más difíciles fue definir clases de entidades y sus características. Basada en una intensiva discusión interdisciplinaria se decidió que el modelo INTERMI debía ser realizado en 7 clases de entidades: personal, familia, entidad corporativa, objeto geográfico, obra, evento y entidades de trabajo general. Cada clase está dividida en subclases que son subgrupos significativos de estas entidades y

necesitan ser representadas explícitamente debido a su significado para la aplicación de la base de conocimientos INTERMI.

La definición de una clase está basada en las características comunes de las entidades. El principal objetivo fue identificar fácilmente y describir las características de las entidades pertenecientes a las clases específicas. Sin embargo, está permitido usar características de una clase para describir la entidad categorizada primariamente en otras clases (ejemplos son castillos, presas etc.)

TIPOS DE ENTIDADES	NÚMERO	EJEMPLOS DE SUBCLASES DE ENTIDADES
PERSONA/CRATURA	4	PERSONAS REALES, PERSONAS NOMBRADAS EN OBRAS RELIGIOSAS, PERSONAS FICTICIAS Y LEGENDARIAS, ENTIDADES REALES O FICTICIAS NO HUMANAS
FAMILIA	3	FAMILIAS, RAMAS DE LA FAMILIA, FAMILIAS FICTICIAS
ENTIDAD CORPORATIVA	12	TERRITORIO CON SU PROPIA ADMINISTRACIÓN, PARTIDOS POLÍTICOS, MOVIMIENTOS, ASOCIACIONES, FIRMAS ...
OBJETO GEOGRÁFICO	9	ESTADOS, PAISES, REGIONES HISTÓRICAS CARACTERÍSTICAS GEOMORFOLÓGICAS
EVENTOS	3	ACCIONES ORGANIZADAS Y EVENTOS (CONFERENCIAS, TALLERES) DIAS IMPORTANTES, GUERRAS, BATALLAS
OBRAS	7	OBRAS LITERARIAS, OBRAS DE ARTE, DOCUMENTOS CONOCIDOS (LEYES, CONSTITUCIONES), ESTRUCTURAS EDIFICIOS RENOMBRADOS (CASTILLOS, IGLESIAS ...) PROGRAMAS , PROYECTOS, BECAS
CONCEPTO GENERAL	7	OBJETOS Y SUS PARTES FÍSICAS CATEGORÍAS Y GRUPOS DE PERSONAS NO NOMBRADAS Y ENTIDADES, ENTIDADES ABSTRACTAS, MATERIALES, TÉCNICAS

**Tabla 1. Tipos de entidades y ejemplos de subclases de entidades.**

### **Breves características de los tipos de entidades usados en el Proyecto INTERMI.**

#### **Tipos de entidades: Persona/Criatura**

Estas clases de entidades incluyen personas reales y personas nombradas en obras religiosas, personas ficticias y legendarias, y entidades reales o ficticias no humanas.

#### **Personas con más biografías y otras identidades.**

Existen dos posibilidades para describir una persona compleja con más identidades en el proyecto INTERMI. La primera (y preferida por las reglas INTERMI) es describir la persona con más identidades en un registro con una denominación preferida (más usada, más conocida etc.) mientras otras apelaciones están grabadas como apelaciones variantes. La segunda mencionada en las reglas INTERMI es describir la persona con identidad real o ficticia separadamente – en dos registros. Esto parece útil cuando los diferentes datos



biográficos usados para la identidad alternativa (como el curriculum vitae ficticio) existen. Es necesario crear relaciones entre todos los registros conectados a una entidad.

### **Seudónimo colectivo, anónimo y creador desconocido.**

Un grupo usando seudónimo colectivo se incluye en esta clase de entidades si el nombre que ellos usan es similar al de esa persona. Entidades específicas como creador anónimo y creador desconocido están incluidas también en el tipo de entidad “persona/criatura”.

### **Tipos de entidades: Familia**

“Familia” son un tipo de entidades que incluye familias y partes (ramas) de familias. En el proyecto INTERMI, se decidió crear clases separadas de estas entidades para describir características específicas de las familias en su complejidad. En cuanto a esta clase, hay que mencionar que un enfoque diferente es aplicado en archivos que conceden un concepto de familias como entidades (como un grupo de gente nombrado y organizado).

### **Tipos de entidades: entidad corporativa**

El tipo “Entidad corporativa” fue una de las clases más discutidas teniendo en cuenta los diferentes métodos de descripción aplicados en archivos y bibliotecas. Las reglas INTERMI adopta principios de archivos e incluye el tipo de entidad corporativa como uno de los atributos que determina la entidad. Aunque, este principio no cumple con las reglas RDA.

### **Tipos de entidades: objeto geográfico**

Objeto geográfico como entidad se identifica con un conjunto estable de coordenadas geográficas. Las etapas de desarrollo del objeto geográfico son normalmente descritas en un registro con denominaciones variantes especificadas por fechas indicando la fecha en que las denominaciones fueron usadas. Si es necesario, INTERMI permite la aplicación de principios de registros separados para cada fase importante del desarrollo de la entidad. (con las relaciones correspondientes entre registros).

### **Tipos de entidades: Obra**

Este tipo representa los más complejas entidades en el proyecto INTERMI. El concepto obra dentro de las comunidades de instituciones de la memoria varía enormemente. Los Requisitos Funcionales para Registros Bibliográficos (en lo sucesivo FRBR) define obra como una creación intelectual o artística perceptible. Una obra es una entidad abstracta; no hay un solo objeto material que uno pueda señalar como una obra. Reconocemos la obra a través de realizaciones individuales o expresiones de la obra, pero la obra misma existe solo en la comunidad de contenido entre las diferentes expresiones de la obra<sup>4</sup>.

Según la Catalogación de Objetos Culturales “una obra es una creación perceptible intelectual o artística limitada primariamente a objetos y estructuras hechas por humanos incluyendo obras construidas, obras de arte visual , y artefactos culturales...Obras incluyen arquitectura, arquitectura de paisaje, otras construcciones, objetos tales como pinturas, esculturas, murales, dibujos, impresos, mobiliario, cerámicas, herramientas, vestidos, textiles, otros objetos decorativos o utilitarios, o cualquier otro tipo de los miles de tipos de creaciones artísticas y de restos culturales. Representaciones de arte, instalaciones, y obras de sitio

---

<sup>4</sup> [http://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)

específico están incluidas. *Están excluidas obras literarias, musicales, artes escénicas, ciencia, religión, filosofía y otros intangibles culturales*<sup>5</sup>.”

Como estos son dos enfoques muy diferentes, era muy difícil definir este tipo de entidades. En los museos y galerías, el concepto FRBR de obra se encuentra con la incompreensión debido a los diferentes significados del concepto (como “obra de arte” producto de arte de gran carga creativa).

Cuando identificamos los tipos de entidades “Obra” definimos características básicas de la entidad “obra” como sigue: “Obra” es el resultado de la actividad intencional humana, y no representa un proceso. Usando este gran nivel de abstracción de la descripción permitimos identificar el tipo de entidades “obra” de forma aceptable para todas las comunidades de memoria de las instituciones. La descripción de la entidad obra se concentra en su estado actual. Una obra en su complejidad es descrita en un registro. Si es necesario, es posible crear registros separados de etapas históricas importantes o parte de las obras (con relaciones entre los registros). El tipo de entidad obra consiste en materiales que tienen algún material de representación, pero también de entidades que no existen en ninguna representación más. Este tipo incluye arte y obras literarias, documentos legales (leyes, etc), normas, productos, marcas, juegos, proyectos, edificios etc.

#### **Tipos de entidades: Evento**

Tipos de entidades “evento” incluyen un evento nombrado temporalmente (largo plazo o corto plazo; una vez o repetido) y también entidades nombradas que esta relacionadas a la actividad humana tales como cultura, hábitos populares.

De acuerdo, a las reglas INTERMI, cualquier repetición o interpretación del evento es descrita como una entidad separada. Sin embargo, es posible crear un registro colectivo (cubierta) que consista en información relacionada a la historia compleja de eventos. Este registro está relacionado a todas las repeticiones/interpretaciones.

#### **Tipos de entidades: concepto general**

Este tipo incluye conceptos generales para entidades específicas tales como categorías de gente, cosas, animales, plantas y también conceptos generales para entidades abstractas tales como características, áreas científicas, estilos de arte, etc.

#### **Relaciones entre las entidades en el modelo conceptual INTERMI**

Los modelos de relación son una asociación entre dos o más entidades para los que las ocurrencias de estas entidades deben corresponderse. Frecuentemente, una relación significativa existe entre dos tipos diferentes de entidad (e.g. creador- obra, familia-miembro de la familia, persona-objeto geográfico (lugar de nacimiento)). Hay dos tipos de relación en el modelo conceptual INTERMI : simples y complejas.

Para crear una relación simple en el modelo de conocimiento INTERMI las reglas indican crear un nexo a otra entidad (e.g. un nexo de persona a persona) con la posibilidad de especificar el papel de la entidad unidad (e.g. hermano, esposa). Es posible introducir información adicional sobre las relaciones tales como fechas y notas.

---

<sup>5</sup> [http://cco.vrafoundation.org/downloads/PartOne\\_GeneralGuidelines.pdf](http://cco.vrafoundation.org/downloads/PartOne_GeneralGuidelines.pdf)

Algunas relaciones son complejas: expresan las complejas relaciones en su completas necesidades para definir las y usarlas, sucesos “descriptivos” en la vida o la historia de la entidad (e.g. nacimiento, establecimiento, creación, destrucción) y un conjunto de relaciones que están conectadas.

### **Registro de entidad INTERMI- estructura de información sobre una entidad.**

La entidad INTERMI se describe/documenta en registros que contiene un conjunto de datos elementos relacionados que son procesados y almacenados juntos. El típico registro de entidad INTERMI incluye la siguiente información:

<b>IDENTIFICADOR</b>	<b>INTERPI ID</b> – ÚNICO OBLIGATORIO; MÁS IDENTIFICADORES ESTÁN PERMITIDOS E.G. ID de Autoridades Nacionales, Autoridades de Museo, VIAF ID
<b>Clase/subclase</b>	Obligatorio
<b>Lenguaje</b>	Lengua en que las notas y la información adicional son introducidas- idioma Checo.
<b>Reglas</b>	Usadas en descripciones de atributos e información sobre la entidad e.g. RDA, AACR2, CCO, Reglas Básicas (archivos), reglas INTERPI
<b>Fechas</b>	Simple en la forma del texto y complejas (repletas de información) que permiten tanto: visualización y la búsqueda.
<b>Notas</b>	Públicas y no públicas
<b>Breves características, descripción, historia</b>	Breves características: máximo dos frases, descripción, historia; máximo 12000 signos.
<b>Denominación</b>	Preferida y formas variantes, cada forma es identificada por reglas máximo es una forma preferida por una regla.
<b>Partes de denominación</b>	Parte principal, otra parte de la denominación, parte adicional de denominación (materias, geográficas, información cronológica)
<b>Clasificación</b>	Características de la entidad expresadas a través un término tópico, categoría
<b>Relaciones y eventos usados en la descripción de la entidad</b>	Relaciones simples que se expresan a través de un nexo a otra entidad (con expresiones del papel de la entidad relacionada, fechas de relación y notas) las relaciones complejas se expresan a través de eventos unidos a más tipos de entidades en diferentes papeles fechas y notas.  Tres tipos de entidades descriptivas son identificadas: comienzo, fin de la existencia de la entidad. Cambio de la entidad
<b>Representación visual de la entidad</b>	e.g. imagen
<b>Recurso para información sobre la entidad</b>	Se da fuente de información del registro entero o de una propiedad específica; se aplica la descripción normalizada.

## **Tabla 2 INTERMI registro de entidad**

### **Reglas del modelo de conocimiento INTERMI.**

Las reglas del modelo de conocimiento INTERMI proporcionan instrucciones y pautas sobre la formulación de datos para las instituciones de la memoria en Chequia. El objetivo no era crear una regla patentada INTERMI. Al mismo tiempo, queríamos mantener las mejores prácticas y la singularidad del enfoque de las memorias de las instituciones cuando sea necesario. Hemos analizado cuidadosamente las reglas utilizadas en las instituciones de la memoria (AACR2, ISAAR CPF, CCO, RDA) y llegamos a la conclusión de que no era posible aplicar ninguna de las reglas existentes en su totalidad. Decidimos aplicar RDA siempre que fuera posible. Sin embargo, para las áreas específicas de las entidades de descripción, e.g. descripción de denominación de entidad, descripción de fechas y definición de reglas para el mapeo de vocabularios controlados usados en el de instituciones de la memoria, se han formulado instrucciones específicas. Instrucciones específicas se han cumplido con e.g. las Reglas de Catalogación de Objetos Culturales (e.g. descripción de fechas) o las Reglas Básicas para los archivos en Chequia. (e.g. descripción de formas preferidas y variantes de entidades corporativas).

### **Reglas INTERMI para denominación de entidades**

Diferentes grupos de las comunidades de usuarios de la memoria en Chequia quieren tener diferentes formas preferidas para las denominaciones de la misma entidad (e.g. entidad de objeto general, entidad corporativa); esta idea es de importancia clave para las reglas INTERMI; permite crear más de una denominación preferida de entidad en un registro. Sin embargo, es necesario asignar cada denominación aquellas reglas de acuerdo con lo que se ha creado. Este principio posibilita proveer múltiples puntos de vista sobre los datos de una entidad de acuerdo a las reglas usadas en cada institución de la memoria y también es útil en el proceso de elaboración de un tesoro. Esperamos que el principio de denominación preferidas múltiple de una entidad sea la principal solución por llegar, pero en la aplicación práctica las comunidades de usuarios de las instituciones de la memoria utilizarán una denominación preferida – debido a que lo encuentran fácil y razonable.

### **Las reglas INTERMI para fechas en la descripción de entidad.**

Según las fechas usadas en la descripción de entidad hemos decidido adoptar principios de CCO. Estos principios están basados en fechas enteras en dos modalidades – para acceso al usuario y para acceso informático. Para acceso del usuario, fechas inciertas están descritas en formato texto (e.g. ca/aproximadamente 1984), para el acceso informático en formato de ISO 8601, donde la primera fecha delimita la fecha inicial potencial, la segunda la fecha de término potencial (e.g. 1890-1910).

### **Las reglas INTERMI para el mapeo de vocabularios controlados, registros etc. Usados en las instituciones de la memoria**

El objetivo de estas reglas INTERMI es resumir métodos utilizados para elaborar varios vocabularios para entidades objeto generales. Suponemos que la elaboración de reglas de vocabularios controlados debería ser importante para aquellas instituciones creando su propio vocabulario controlado específico en una forma que se ajuste la Web Semántica.

Mapeo/armonización del proceso depende de:

- La elección de un sistema controlado de términos que puedan servir como una base para comparar y evaluar términos de otros recursos terminológicos.
- La elección de una herramienta para la representación formal de recursos terminológicos (SKOS).
- La descripción y características (formales y semánticas) de términos usados en recursos terminológicos.
- La elección de métodos aplicados en procesos de mapeo/armonización, tales como basados en léxico, basados en conceptos, y ejemplos basados en mapeo.

El propósito del mapeo basado en léxico es averiguar el grado de similaridad de cadenas de texto

Pojem	Identifikátor		
	Národní autority	EuroVoc	AGROVOC
politika	PSH8309	131558	6062
informatika	PSH6548	100223	3864
přírodní vědy	PSH11969	141614	4318
trh práce	PSH1264	142776	28713

Tabla 2. Ejemplos de términos idénticos de tres recursos obtenidos por el método basado en el léxico.

El mapeo basado en conceptos se aplica cuando se compara la terminología de un tesoro especializado y de vocabularios controlados. Se realizan como métodos gráficos comunes o métodos de similaridad de vector.

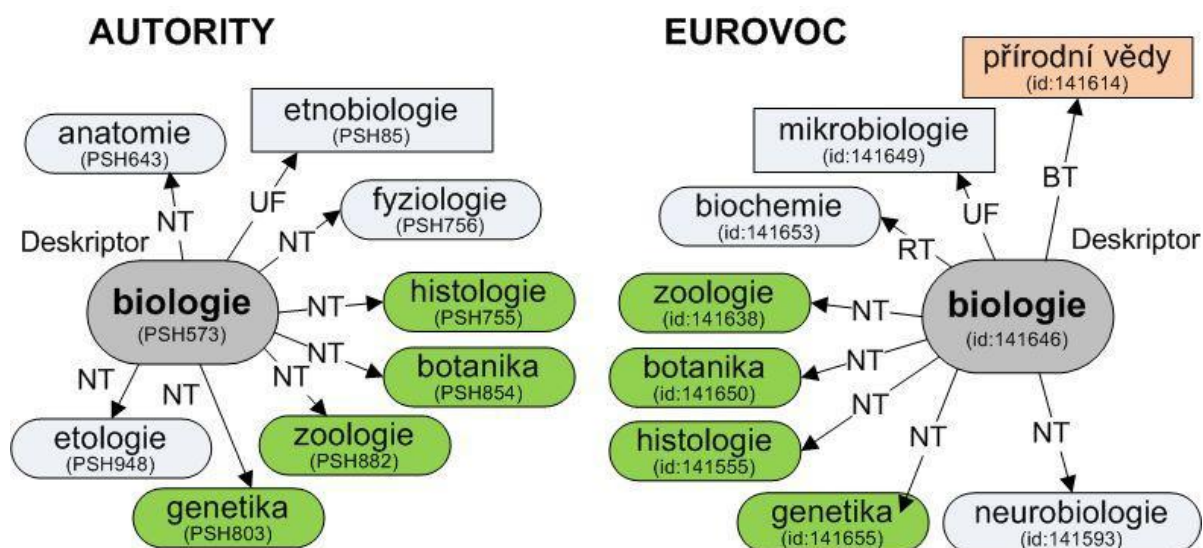


Figura 3 Ejemplo de métodos de gráficos comunes – subgráficos de concepto de “biología” de recursos terminológicos: Autoridades Nacionales y EUROVOC.

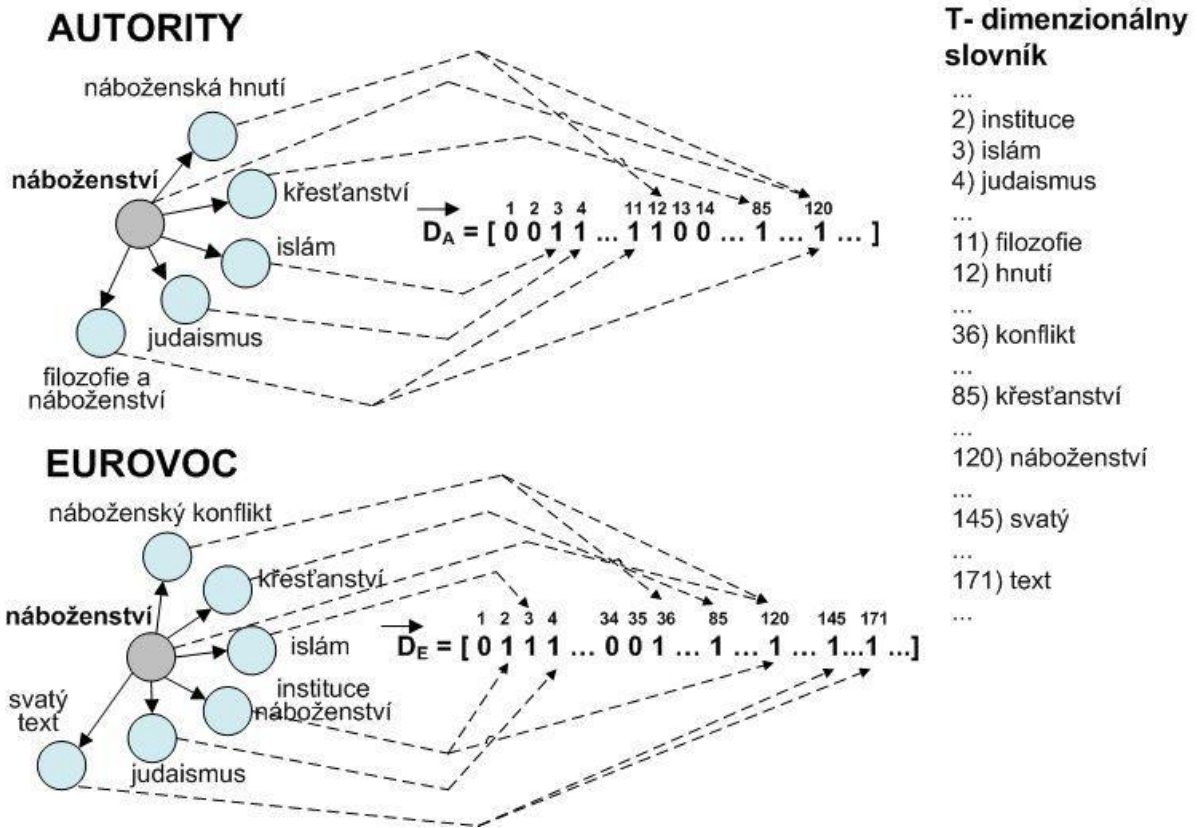


Figura 4. Ejemplo del método de similaridad del vector – T- vector dimensional del concepto de “religión” en fuentes terminológicas: Autoridades Nacionales y Eurovoc.

Cuando usamos el método de mapeo basado en ejemplos -los conceptos equivalentes, generales, específicos, y relacionados están basados en las similitudes léxicas y semánticas entre la fuente y los objetivos del Tesaurus que se buscan.

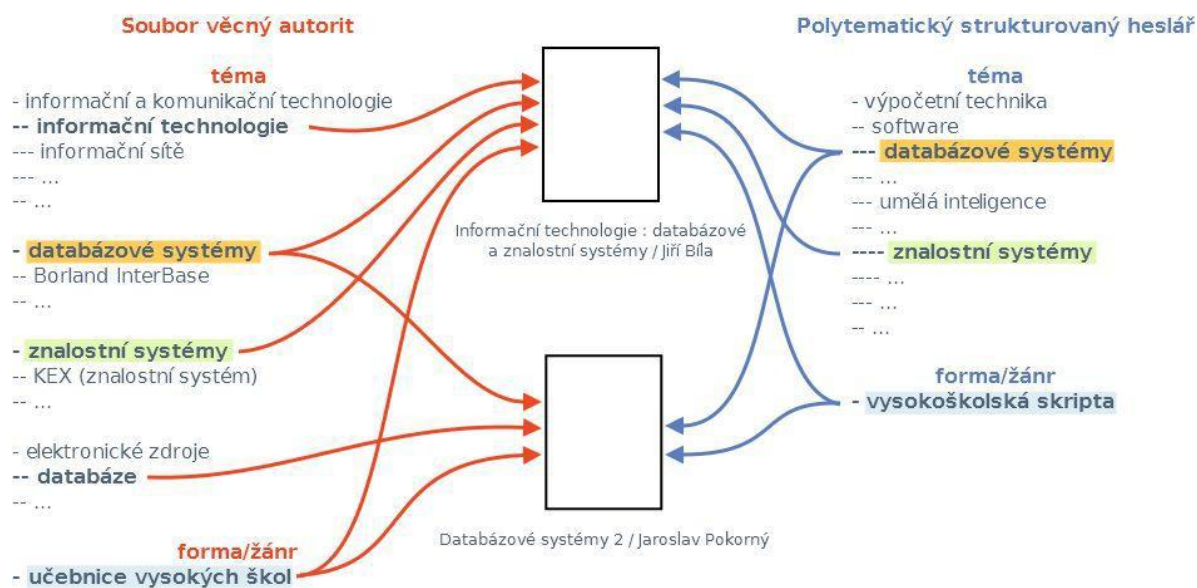


Figura 5. Ejemplo de mapeo basado en los ejemplos.

### El papel de las autoridades de materia

Los archivos de autoridades de materia han sido elegidos como una base para el mapeo/armonización de recursos terminológicos usados en la memoria de las instituciones. El archivo de autoridades de materia representa un conjunto de términos generales cuyo contenido cubre todos los campos tanto a un nivel general como específico. En lo que se refiere al archivo de autoridad de materias Checo, nuestra intención es crear una herramienta de mapeo y armonización, lo que significa transformarla en una ontología “general”. Lo que supone

- Definir el significado de términos de materia exactamente añadiendo calificadores, notaciones de CDU y notas de alcance, para que los términos de materia sean capaces de representar conceptos aislados bien definidos.
- Expresar estructura jerárquica que sea completa y consistente
- Para usar identificadores únicos permanentes que deben ser independientes del lenguaje para evitar la necesidad de cambiar ID en conjunción con el cambio de forma preferido aplicado en la descripción del concepto.
- Para expresar archivos de términos de materia en una forma comprensible por la máquina dentro del marco de la Web Semántica., cuando usar RDF aplicando SKOS; capacitar a los datos para ser unidos y fusionados con otros datos RDF por aplicaciones de Web semántica.

## **Conclusión**

### **Importancia del Proyecto INTERMI**

El proyecto INTERMI provee un espacio para la creación y preservación de datos sobre entidades utilizados en las instituciones de la memoria en Chequia. Representa una herramienta amigable para el acceso y reparto del contenido del Patrimonio Cultural a través de la memoria de las instituciones asegurando la interoperabilidad semántica a un nivel conceptual. El objetivo del proyecto es crear un modelo conceptual y ontología para hacer disponible el Patrimonio Cultural nacional Checo en la web. El proyecto se construye en proyectos y objetivos de investigación anteriores, e.g. Archivos de autoridad nacional y autoridades de museo, pero además de esto, el objetivo para una construcción exhaustiva de una infraestructura para edificar el modelo basado en conocimiento del contenido del patrimonio cultural y para la apertura de su uso en la forma de un trabajo operativo piloto. El proyecto es complementario a los proyectos como la Biblioteca Digital Nacional, el archivo digital nacional.

### **Impacto de la colaboración entre las instituciones de la memoria.**

Debe señalarse que el proyecto tiene un impacto sobre la discusión entre diferentes comunidades de la memoria de las instituciones. Expertos han tenido la oportunidad para reunirse juntos y expresar sus necesidades trabajando con la descripción e identificación de entidades de sus colecciones. Al final del proyecto INTERMI predecimos una colaboración efectiva entre los profesionales que debería tener un impacto práctico en su práctica diaria cuando describan recursos de información usando entidades INTERMI. Para la realización de este objetivo, se ofrece un conjunto de servicios Web INTERMI para su aplicación en el sistema local de información.

### **Impacto del acceso al usuario a la información.**

Usando las entidades INTERMI en la descripción de las colecciones de las instituciones de la memoria mejorará la calidad del acceso del usuario al patrimonio cultural nacional publicado en la World Wide Web; al mismo tiempo el proyecto INTERMI demostrará formas de presentar información sobre entidades con la aplicación de las tecnologías de la Web semántica.

## **References**

Ansgar Scherp , Thomas Franz , Carsten Saathoff , Steffen Staab, A core ontology on events for representing occurrences in the real world, *Multimedia Tools and Applications*, v.58 n.2, p.293-331, May 2012

American Library Association et al. *Anglo-American Cataloguing Rules*. 2nd Edition. Edited by Michael Gorman and Paul W. Winkler. Chicago: American Library Association, 1978.

Bountouri, L., Papatheodorou, C., & Gergatsoulis, M. (2010, January). Modelling the public sector information through CIDOC conceptual reference model. In *On the Move to Meaningful Internet Systems: OTM 2010 Workshops* (pp. 404-413). Springer Berlin Heidelberg.

Constantia Kakali, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Integrating Dublin Core Metadata for Cultural Heritage



Collections Using Ontologies. In: Proceedings of the International Conference on Dublin Core and metadata Applications (DC-2007), pp. 128-139. 27-31 August, 2007. Singapore.

Coyle, Karen. "Metadata mix and match." *Information Standards Quarterly* 21, no. 1 (2009): 9-11.  
Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & van de Sompel, H. (2010, August). The europeana data model (EDM). In *World Library and Information Congress: 76th IFLA general conference and assembly* (pp. 10-15).

Gaitanou, P., & Gergatsoulis, M. (2011). A semantic mapping of VRA Core 4.0 to the CIDOC Conceptual Reference Model. In *Metadata and Semantic Research* (pp. 387-399). Springer Berlin Heidelberg.

Gergatsoulis, M., Bountouri, L., Gaitanou, P., & Papatheodorou, C. (2010). Mapping cultural metadata schemas to CIDOC conceptual reference model. In *Artificial Intelligence: Theories, Models and Applications* (pp. 321-326). Springer Berlin Heidelberg.

Harper, C. A. (2006). Encoding library of congress subject headings in SKOS: authority control for the semantic web, Proceedings of the 2006 international conference on Dublin 15

Core and Metadata Applications: metadata for knowledge and learning, 89-94. Manzanillo, Colima, Mexico: Dublin Core Metadata Initiative.  
IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: Final Report*. München: K.G. Saur, 1998.

ISO 25964-1:2011 = International Standards Organization (ISO). Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval. 1st ed. Geneva: ISO, 2011 (ISO 25964-1- 2011-08-05)

Joint Steering Committee for Development of RDA. *RDA: Resource Description and Access*. Chicago: American Library Association, 2010.

Kramer, R., R. Nikolai, and C. Habeck (1997) "Thesaurus federations: loosely integrated thesauri for document retrieval in networks based on Internet technologies". In *International Journal on Digital Libraries* (1), 122-131

Landry, P. (2000) "The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD)". *Classification and Indexing Workshop, 66th IFLA Council and General Conference, Meeting No. 181* <http://www.ifla.org/IV/ifla66/papers/165-181e.pdf>

Lourdi, Irene, Christos Papatheodorou, and Martin Doerr. "Semantic integration of collection description." *D-Lib Magazine* 15.7/8 (2009): 1082-9873.

MERIMEE, "THESAURUS ARCHITECTURE" for the indexing of complexes, buildings and built works described in the national database "Merimee" about the French Heritage <http://www.culture.gouv.fr/documentation/thesarch/pres.htm>

Mitchell, E. T. (2013). Chapter 1: Metadata Developments in Libraries and Other Cultural Heritage Institutions. *Library Technology Reports*, 49(5), 5-10.

Soergel, D. (1995) *The Art and Architecture Thesaurus (AAT): A critical appraisal*. Visual Resources, X, pp. 369-400

Svensson, Lars G. "Are Current Bibliographic Models Suitable for Integration with the Web." *Information Standards Quarterly* 25 (2013): 7-12.

Thomas R. Gruber (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Nicola Guarino and Roberto Poli (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation. Thomais Stasinopoulou, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Ontology-based Metadata Integration in the Cultural Heritage Domain. In: Proceedings of the 10th International Conference on Asian Digital Libraries, pp.165-175. Hanoi, Vietnam, December 10-13, 2007.

Skinner, J. (2014). Metadata in Archival and Cultural Heritage Settings: A Review of the Literature. *Journal of Library Metadata*, 14(1), 52-68.

Core and Metadata Applications: metadata for knowledge and learning, 89-94. Manzanillo, Colima, Mexico: Dublin Core Metadata Initiative.

IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: Final Report*. München: K.G. Saur, 1998.

ISO 25964-1:2011 = International Standards Organization (ISO). Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval. 1st ed. Geneva: ISO, 2011 (ISO 25964-1- 2011-08-05)

Joint Steering Committee for Development of RDA. *RDA: Resource Description and Access*. Chicago: American Library Association, 2010.

Kramer, R., R. Nikolai, and C. Habeck (1997) "Thesaurus federations: loosely integrated thesauri for document retrieval in networks based on Internet technologies". In *International Journal on Digital Libraries* (1), 122-131

Landry, P. (2000) "The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD)". *Classification and Indexing Workshop, 66th IFLA Council and General Conference, Meeting No. 181* <http://www.ifla.org/IV/ifla66/papers/165-181e.pdf>

Lourdi, Irene, Christos Papatheodorou, and Martin Doerr. "Semantic integration of collection description." *D-Lib Magazine* 15.7/8 (2009): 1082-9873.

MERIMEE, "THESAURUS ARCHITECTURE" for the indexing of complexes, buildings and built works described in the national database "Merimee" about the French Heritage <http://www.culture.gouv.fr/documentation/thesarch/pres.htm>

Mitchell, E. T. (2013). Chapter 1: Metadata Developments in Libraries and Other Cultural Heritage Institutions. *Library Technology Reports*, 49(5), 5-10.

Soergel, D. (1995) *The Art and Architecture Thesaurus (AAT): A critical appraisal*. Visual Resources, X, pp. 369-400

Svensson, Lars G. "Are Current Bibliographic Models Suitable for Integration with the Web." *Information Standards Quarterly* 25 (2013): 7-12.

Thomas R. Gruber (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Nicola Guarino and Roberto Poli (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation. Thomais Stasinopoulou, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Ontology-based Metadata Integration in the Cultural Heritage Domain. In: Proceedings of the 10th International Conference on Asian Digital Libraries, pp.165-175. Hanoi, Vietnam, December 10-13, 2007.

Skinner, J. (2014). Metadata in Archival and Cultural Heritage Settings: A Review of the Literature. *Journal of Library Metadata*, 14(1), 52-68.