

## **Облегчение доступа к содержанию культурного наследия в Чехии: Национальные авторитетные файлы и проект INTERMI**

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

*Translated by:* Světlana Knollová, National Library of the Czech Republic, Prague, Czech Republic

### **Мария Баликова**

Национальная библиотека Чешской Республики,  
Прага, Чехия  
E-mail address: Marie.Balikova@nkp.cz

### **Мирослав Кунт**

Национальный архив, Прага, Чехия  
E-mail address: Miroslav.Kunt@nacr.cz

### **Яна Шубова**

ООО “Cosmotron Bohemia“, Годонин, Чехия  
E-mail address: subova@cosmotron.cz

### **Надежда Андрейчикова**

ООО “Cosmotron Bohemia“, Годонин, Чехия  
E-mail address: andrejcikova@cosmotron.cz



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

---

### **Аннотация:**

*Реферат посвящен проекту “Интероперабельность в учреждениях – хранителях исторической памяти (INTERMI) в Чехии”, который реализует Национальная библиотека Чешской Республики и Национальный архив в рамках программы НАКИ, финансируемой Министерством культуры Чешской Республики. Целью проекта является создание научной и технологической инфраструктуры для поддержки обработки, кооперации и вторичного использования содержания культурного наследия в форме метаданных информационных объектов, охраняемых в учреждениях – хранителях исторической памяти; это означает*

*построить концептуальную модель базы знаний, которая будет отвечать критериям и требованиям пользователей всех таких учреждений в Чехии. Проект базируется на новой парадигме, которая фокусируется на обработке данных, расширенных за счет необходимой семантической информации, а также с помощью построения комплексных связей между ними. Одновременно в проекте INTERMI анализируется роль национальных авторитетных файлов.*

**Ключевые слова:** учреждения – хранители исторической памяти, национальные авторитетные файлы, культурное наследие, базы знаний, концептуальные модели, энтитет, событийно-ориентированный подход.

---

## **Предисловие и контекст**

Библиотеки, архивы, галереи и музеи как институты сохранения исторической памяти собирают, организуют и делают доступными источники и объекты информации национального, культурного, индустриального и природного наследия. Составной частью культурного наследия является также документальное и архивное наследие.

Сегодня, в период массовой оцифровки главная цель деятельности институтов сохранения исторической памяти заключается в поддержке эффективного выбора, агрегации, использования и вторичного использования оцифрованного контента.

В проекте INTERMI эта главная цель базируется на спецификации требований пользователей, качестве метаданных, связанных с объектами информации (включая однозначную идентификацию объектов информации на уровне метаданных), на событийно-ориентированном подходе, на применении семантических интероперабельных технологий, то есть на концептуальной модели и избранной структуре данных, на инструментах для формальной репрезентации, соответствующих семантическому вебу (напр. XML, RDF, OWL) и на применении принципа взаимосвязанных открытых данных (LOD).

### **Спецификация читательских запросов и их ожидания, что касается проекта INTERMI**

Пользователи учреждений - хранителей исторической памяти в Чехии (широкая публика, каталогизаторы, индексаторы, информационные специалисты, кураторы, архивисты и специалисты из всех отраслей) поддерживают возникновение и развитие инфраструктуры для базы знаний культурного наследия, ратуют за доступ к культурному наследию в сети в машиночитаемой форме понятной и для человека, то есть за распространение культурного контента на основе принципов семантического веба и взаимосвязанных открытых данных. Они отдают себе отчет в том, что это предполагает развитие и применение высококачественных стандартов, правил и схем метаданных, соответствующих семантическому вебу. Что касается энтитетов, пользователям нужна более подробная информация в существующих авторитетных описаниях, тем не менее, они отдают себе отчет в том, что необходимо перейти от актуальных авторитетных файлов к данным, которые имеют ценность для сети, от терминов к понятиям.

### **Информационные объекты и их идентификация на уровне метаданных**

В проекте INTERMI информационные объекты представляют собой элементы реального мира, каковыми, например, являются люди, учреждения, трехмерные

объекты (предметы искусства, технические объекты неживой природы), разные виды деятельности, события, театральные и другие представления, места и т.д., которые отображают тему экспонатов коллекций учреждений сохранения исторической памяти или предмет их деятельности. В настоящее время идентификация таких объектов основана на именных и предметных авторитетных записях.

### **Роль национальных авторитетных файлов – сделать доступным цифровой контент в Чехии**

Национальные авторитетные файлы содержат авторитетный файл имен лиц (639 472 описания), авторитетный файл коллективного автора (140 002), авторитетный файл унифицированного заглавия (2 454), авторитетный файл тематических понятий (36 954), авторитетный файл географических наименований (27 170) и авторитетный файл, отражающий жанр и форму (1 783).

Национальные авторитетные файлы с самого начала развивались с помощью корпоративной каталогизации и индексации библиотечных материалов.

Авторитетные описания имен лиц создавались в соответствии с правилами AACR2<sup>1</sup>, в то время как авторитетные файлы предметных рубрик первоначально разрабатывались по правилам LCSH. Позже, в связи с появлением интернета, были введены некоторые новые принципы: посткоординация, авторизация изолированных лексических единиц (не целый ряд предметных рубрик), ограниченное использование рядов предметных рубрик в библиографических описаниях. Главной целью авторитетных описаний является установление авторизованных форм имен/названий, их вариантов и запись этих форм; они вводятся с помощью формата MARC.

### **Сильные стороны традиционных авторитетных файлов**

В традиционных авторитетных файлах установлены традиционные точки доступа; они являются составляющей хорошо разработанной практики и поддерживаются как библиотеками, так и другими учреждениями. Библиотеки в Чехии принимают участие в проекте “Корпоративное создание и использование национальных авторитетных файлов“, в то время как музеи и галереи руководствуются проектами “Музейные авторитетные файлы“ и “Регистр коллекции изящных искусств“. Все эти проекты базируются на национальных авторитетных файлах; “Музейные авторитетные файлы“ и “Регистр коллекции изящных искусств“ разрабатывают только авторитетные файлы имен (точки доступа - индивидуальный и коллективный авторы).

### **Слабые стороны традиционных авторитетных файлов**

#### **Система, основанная на описании**

Традиционные библиотечные авторитетные файлы являются системами, основанными на описании. Такие файлы создаются прежде всего для людей и соответствуют их пониманию; с другой стороны они не предоставлены в машиночитаемом формате, а лишь в машинопонимаемом.

#### **Отсутствие достаточной детализации данных**

AACR2 и формат MARC не обеспечивают достаточную степень детализации введенной информации. Формат MARC не предоставляет нужную детализацию элементов данных, которые бы могли однозначно служить функциям Semantic Web и

---

<sup>1</sup> С мая 2015 г. были введены правила RDA.

LOD. “Описание в MARC не было задумано как блок элементов данных, а как формат для сохранения и изображения текста описаний библиотечного каталога.”<sup>2</sup>

### **Энтитеты одного и того же типа в разных авторитетных файлах**

Другой проблемой является то, что некоторые энтитеты классифицируются в разнообразных/различных авторитетных файлах в соответствии с полями MARC, в которые они вводятся. Некоторые энтитеты, как например, описание события появляются в двух авторитетных файлах: в файле конференций (конференции, семинары, симпозиумы) и в тематическом файле (войны, битвы и т.д.). Конструкции, которые имеют названия, включаются в файл коллективного автора, конструкции без названия классифицированы в файле тематического понятия. Эти энтитеты должны быть классифицированы скорее в соответствии с их атрибутами и способностью выражать взаимосвязь с другими энтитетами, чем в соответствии с полями MARC.

### **Мало информации в авторитетных описаниях**

Авторитетные описания содержат мало сведений о данном энтитете: они концентрируются на избранных формах имен/названий, введенных как главные рубрики, и их вариантах, введенных в качестве ссылок. Здесь нет возможности ввести всю информацию, необходимую для выражения свойств/атрибутов и взаимосвязи между энтитетами в структурированном виде. В лучшем случае можно записать определенную информацию только как незакодированный текст в поле замечаний.

### **Термин, обозначающий тип коллективного автора**

В рамках рубрики коллективного автора нет возможности ввести термин, обозначающий тип коллективного автора (напр.: ltd, plc), тем не менее для архивиста тип коллективного автора является одной из самых важных характеристик, которая отличает подобные энтитеты один от другого.

### **Использование национальных авторитетных файлов в проекте INTERMI**

После глубокого анализа было решено использовать авторитетные файлы в соответствии со следующими принципами:

- Стандартизированные точки доступа, то есть предпочитаемые или вариантные формы энтитетов должны быть в проекте использованы.
- Вся контекстная информация, появляющаяся в авторитетных описаниях, должна войти в систему INTERMI в структурированной форме.
- Необходимо всячески избегать тенденции принимать решения только лишь на основе формата MARC.
- Необходимо создать собственную концептуальную модель и формат структуры данных, которые позволят удовлетворить запросы отдельных групп пользователей.
- Концептуальная модель INTERMI должна быть основана на энтитетах (не на авторитетных файлах), что означает серьезную тенденцию перехода от авторитетных файлов к энтитетам, которые представляют ценность для сетевого пространства.

---

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

## Концептуальная модель INTERMI

Концептуальная модель INTERMI основана на принципах объектно-ориентированной парадигмы, на событийно-ориентированном подходе и способности конвертации в CIDOC CRM.

### Принципы объектно-ориентированной парадигмы

- Детализация – данные делятся на более малые, носящие смысл элементы, которые позволяют лучше определить тип данных и обобщить их таким образом, чтобы была обеспечена их флексибельность.
- Наследование – иерархическая структура элементов предусматривает и содействует наследованию атрибутов или свойств.
- Повторное использование – элементы должны быть описаны так, чтобы их можно было использовать повторно в разных позициях модели и способствовать таким образом ее дальнейшему развитию.
- Полиморфизм – определенные элементы могут в зависимости от их применения функционировать как разные типы элементов

### Событийно-ориентированный подход

События представляют собой специфические действия, которые происходят на определенном месте в определенное время и описаны как информационные объекты. Мы согласны с мнением, что “моделирование события так абстрактно, что может быть использовано для описания предметов культуры и документации научных наблюдений”<sup>3</sup>.

В проекте INTERMI появляются два типа событий: события как “класс энтитетов событий” и события, которые применяются для выражения комплексных атрибутов и взаимосвязей в целях описания других энтитетов, так называемые “дескриптивные энтитеты”. События (классы энтитетов) включают в себя названные временные события (длительные или краткосрочные; одноразовые или повторяющиеся), а также названные энтитеты, которые имеют отношение к человеческой деятельности, к народным обычаям. “Дескриптивные энтитеты” связаны с событиями жизни или истории энтитета (напр. рождение, возникновение, установление, разрушение).

Событийно-ориентированный подход предполагает, что много атрибутов энтитета можно описать с помощью события; мы думаем, что событийно-ориентированный подход отражает технологию Linked Data .

### Информационные технологии, использованные в проекте INTERMI

Главной целью проекта является создание базы знаний и онтологии информационных объектов, которые существуют в области культурного наследия. Необходимость разрешить технологические вопросы стала очевидной, так что мы должны были справиться с двумя основными проблемами:

---

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

- Каким образом предоставлять и сохранять данные, охваченные концептуальной моделью INTERMI? Это предполагает найти внутреннюю модель для хранилища данных, которая сможет поддерживать открытость INTERMI, быстрый доступ к данным и удобное отображение. Мы решили использовать собственную структуру, основанную на XML, которая поддерживает взаимосвязи между энтитетами. База INTERMI также располагает мощной базой данных объектов.
- Каким образом отображать данные для внешних субъектов (библиотечные системы, применяемые в архивах и музеях системы управления данными, включая сетевые системы)? Концепции Semantic Web и Linked Data представляют собой инструменты, подходящие для INTERMI, так как они предлагают технологии для презентации данных в контексте, а не только для презентации содержания.

**К технологиям Semantic Web (семантической паутины) относятся:**

- URI (uniform resource identifier - унифицированный идентификатор ресурса), используется для ссылок на объекты и абстрактные понятия;
- RDF и RDF Schema представляет собой способ отображения информации с помощью абстрактной модели метаданных. RDF позволяет описывать простые данные с помощью триплета *субъект-отношение-объект*;
- OWL – язык для создания онтологий;
- SPARQL – язык запросов к данным.

В соответствии с концепцией Linked Data и Semantic Web было решено выбрать из различных стандартов те, которые больше всего подходят для презентации данных. Для проекта были рекомендованы следующие стандарты:

- SKOS – для изображения тезаурусов – может быть использован в энтитетах класса “общее понятие”.
- FOAF – онтология для описания людей, видов деятельности и отношений – может быть использована в энтитетах класса “человек/создание”.

Для сохранения преемственности необходимо предоставить данные из базы знаний INTERMI в структуре, которая в настоящее время чаще всего используется в библиотеках – имеется в виду MARC21/XML для авторитетных описаний.

Презентация данных INTERMI в некоторых специфических структурах предполагает выбор данных, которые могут быть использованы в специфической структуре (или стандарте). Для этого необходима более частая трансформация данных из внутренней во внешнюю структуру.

Для внешних систем доступ к данным INTERMI будет обеспечен с помощью сетевых услуг (напр. log-in of user, find record, get record).

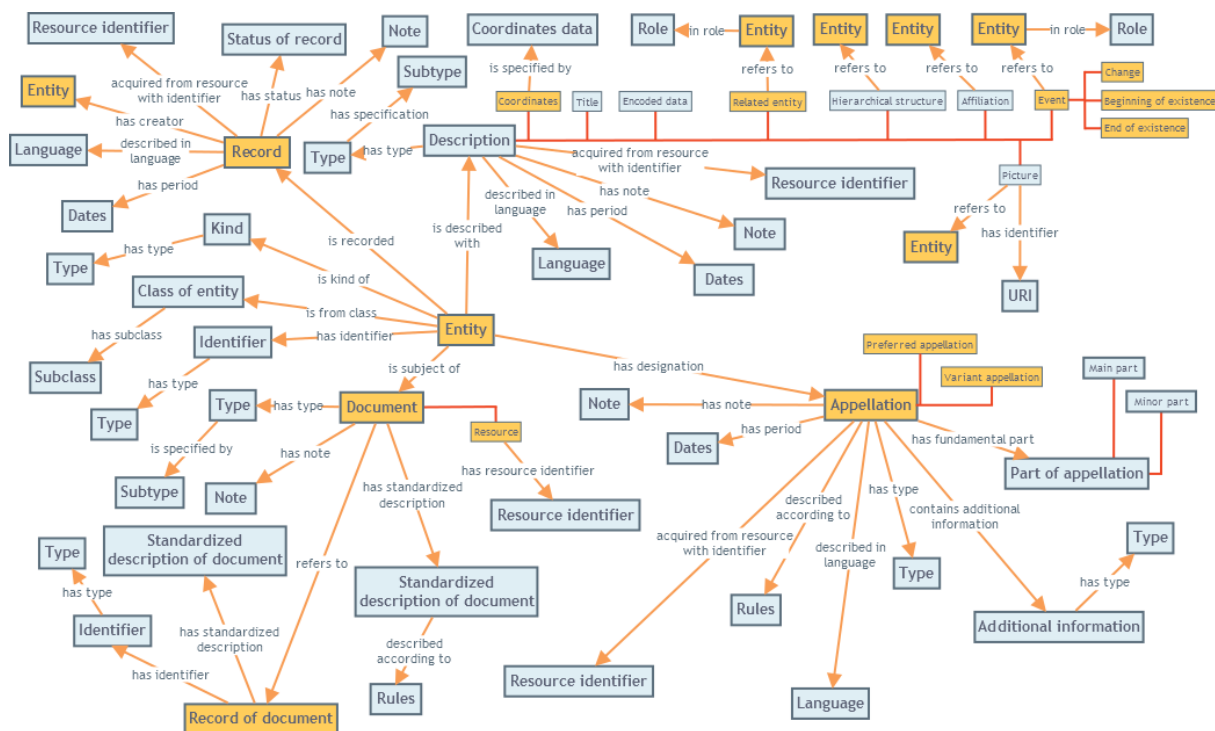


Рисунок 1. Концептуальная модель INTERMI

## Ключевые элементы концептуальной модели INTERMI: энтитеты и отношения

Энтитет является самым важным элементом концептуальной модели INTERMI.

Энтитеты представляют собой идентифицированные объекты реального мира. Энтитет это понятие (индивидуальное или обобщающее), которое обладает идентификатором, типом, который указывает на категорию энтитета и характеризуется рядом свойств, позволяющих описать любой энтитет этой модели. Одной из самых трудных задач проекта было определить классы энтитетов и их особенности. В ходе интенсивной междисциплинарной дискуссии было решено, что концептуальная модель INTERMI будет реализована на базе 7 классов энтитетов: лицо, род, коллектив, географический объект, произведение, событие и общее понятие. Каждый класс делится на подклассы, т.е. на значительные по смыслу подгруппы энтитетов, которые необходимо точно формулировать, так как они имеют для базы знаний INTERMI большое значение.

Определение класса базируется на общих особенностях энтитетов. Главной целью дефиниции классов было обеспечить простую идентификацию и описание характеристик энтитетов, относящихся к одному классу. Тем не менее, разрешено использовать характеристики одного класса для описания энтитетов, первоначально классифицированных в другом классе (напр. замки, дамбы и т.д.).

<b>Класс энтитетов</b>	<b>Количество</b>	<b>Примеры подклассов энтитетов</b>
<b>Лицо/создание</b>	<b>4</b>	Реальные лица, названные лица в священных текстах, фиктивные и легендарные лица, реальные или фиктивные энтитеты нечеловеческого происхождения
<b>Семья/род</b>	<b>3</b>	Семьи, родовые ветви, фиктивные семьи
<b>Коллектив</b>	<b>12</b>	Административно-территориальные единицы, политические партии и движения, ассоциации, фирмы...
<b>Географический объект</b>	<b>9</b>	Государства, страны, исторические регионы, геоморфологические образования
<b>События</b>	<b>3</b>	Организованные акции и события (конференции, симпозиумы), знаменательные даты, войны, битвы
<b>Произведение</b>	<b>7</b>	Литературные произведения, произведения искусства, широко известные документы (законы, конституции), структуры, названные строения (крепости, храмы...), программы, проекты, гранты
<b>Общее понятие</b>	<b>7</b>	Предметы и их физические части, категории и группы неназванных лиц и коллективов, абстрактные энтитеты, материалы, технологии

Таблицы 1. Классы энтитетов и примеры подклассов энтитетов

## **Краткая характеристика классов энтитетов, принятых в проекте INTERMI**

### **Класс энтитетов: Лицо/создание**

В этот класс энтитетов включаются реальные лица и названные персонажи, появляющиеся в священных текстах, фиктивные и легендарные лица, а также реальные и фиктивные энтитеты нечеловеческого происхождения.

### **Лица, выступающие в нескольких биографических и других ипостасях**

В проекте INTERMI существуют два способа описания сложного персонажа, выступающего в нескольких ипостасях. Первый способ, предпочтительный с точки зрения правил проекта INTERMI, предполагает описание такого персонажа в одной записи с одной выбранной формой обозначения (более известной, чаще используемой и т.д.), в то время как другие формы имени записаны как вариантные. Другой способ, упомянутый в правилах INTERMI, предлагает описывать персонажа, выступающего как реальная и фиктивная личность, отдельно в двух записях. Похоже, что это выгодно в случае, когда у лица с альтернативной личностью существуют другие биографические данные (как фиктивная биография). В таком случае необходимо установить связи между всеми записями, относящимися к одному энтитету.

### **Коллективный псевдоним, аноним и неизвестный автор**

Группа, которая использует коллективный псевдоним, включается в этот класс энтитетов в том случае, если псевдоним можно считать личным именем. Такие специфические энтитеты как анонимный автор и неизвестный автор также рассматриваются в классе “лицо/создание”.



### **Класс энтитетов: Семья/род**

Этот класс включает в себя семьи, роды и ветви родов. В проекте было решено создать отдельно класс энтитетов, описывающих специфические характеристики родов в их комплексности. В этой связи, необходимо заметить, что в архивах применяется другой подход к описанию упомянутых выше энтитетов – там они считаются названными и организованными группами людей, то есть коллективами.

### **Класс энтитетов: Коллектив**

Энтитет коллектива вызывает наибольшее количество дискуссий из-за различных способов описания в архивах и в библиотеках. Правила INTERMI руководствуются практикой, распространенной в архивах, то есть указывается тип коллектива как один из атрибутов для определения этого энтитета. Указанный принцип не отвечает правилам RDA.

### **Класс энтитетов: Географический объект**

Географический объект как энтитет определяется строго определенным комплексом географических координат. Этапы развития географического объекта обычно описываются в одной записи с вариантными формами названий, которые сопровождаются уточнением, в какой отрезок времени эти названия были актуальны. Если нужно, INTERMI позволяет использовать самостоятельную запись для особо важных этапов развития энтитета. В этом случае необходимо установить связь между отдельными записями.

### **Класс энтитетов: Произведение**

Этот класс самый сложный среди энтитетов проекта INTERMI. Понятие “произведение” трактуется отдельными хранителями исторической памяти по-разному. Материал *Functional Requirements for Bibliographic Records/Функциональные требования к библиографическим записям* (сокращенно FRBR) определяет “произведение” как “некое интеллектуальное или художественное творение. Произведение является абстрактным энтитетом, не существует материальный объект, который бы соотносился с произведением. Мы воспринимаем произведение сквозь призму индивидуальной реализации и восприятия произведения, тем не менее произведение существует только в общности содержания в рамках разных способов выражения произведения”.<sup>4</sup>

В соответствии с *Cataloguing Cultural Objects* “произведение является определенным интеллектуальным или художественным творением, касающимся прежде всего предметов, созданных людьми, включая строительные объекты, художественные произведения и культурные артефакты... Произведения включают в себя архитектурные объекты, ландшафтную архитектуру, другие строительные объекты, картины, скульптуры, монументальную живопись, рисунки, печатные издания, фотографии, мебель, керамику, орудия труда, одежду, ткани, другие декоративные и утилитарные предметы, а также какой-либо другой предмет из тысяч типов художественных творений или других памятников культуры. Включены также перформанс, инсталляция и энвайронмент. *Наоборот сюда не относятся литературные произведения, музыка, театр, языковое творчество, кулинарное*

---

<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)

*искусство, наука, религия, философия и другие творения нематериальной культуры*<sup>5</sup>.

В связи с тем, что речь идет о двух различных подходах, очень трудно дать определение этому классу энтитетов. Трактовка этого понятия со стороны FRBR встречается с непониманием в музеях и галереях, у которых свое мнение на этот счет (они полагают, что “художественное произведение“ это продукт художественной творческой энергии).

Для идентификации этого класса энтитетов мы определили основные характеристики “произведения“ следующим образом: “Произведение“ является результатом направленной человеческой деятельности и не представляет процесс. Используя данный высокий уровень абстракции описания, мы смогли идентифицировать класс энтитетов “произведение“ способом, приемлемым для всех учреждений – хранителей исторической памяти. Описание энтитета “произведение“ концентрируется на его актуальном состоянии. Каждое произведение комплексно описывается в самостоятельной записи. В случае необходимости можно ввести отдельные записи для значительных исторических периодов или частей произведений (разумеется, создавая связи между записями). Класс энтитетов “произведение“ состоит как из энтитетов, которые представляют собой материальные объекты, так и из энтитетов, которые не существуют в материальной форме. Этот класс включает в себя художественные и литературные произведения, юридические документы (законы и т.д.), нормативы, изделия, товарные знаки, игры, гранты, проекты, строения и т.д.

### **Класс энтитетов: Событие**

Класс энтитетов “событие” включает в себя названные временные события (длительные и кратковременные; имевшее место один раз и повторяющиеся), а также названные энтитеты, которые связаны с такой человеческой деятельностью как культура и народные обычаи. В соответствии с правилами INTERMI каждое повторение или интерпретация события описывается как самостоятельный энтитет. Тем не менее можно создать и сводную (общую) запись, которая будет содержать информацию о полной истории этого события. Такая запись должна быть связана со всеми записями повторений и интерпретаций.

### **Класс энтитетов: Общее понятие**

В этот класс входят общие понятия для специфических энтитетов - таких как категории людей, вещи, животные, растения, а также общие понятия для абстрактных энтитетов - таких как свойства, научные области, художественные направления и т.д.

### **Взаимоотношения между энтитетами в концептуальной модели INTERMI**

Взаимоотношение является выражением связи между двумя или несколькими энтитетами, причем все проявления этих энтитетов должны соответствовать этому взаимоотношению. Часто осмысленное взаимоотношение существует между двумя разными типами энтитетов: напр. творец – произведение, семья – член семьи, человек – географический объект (место рождения). В концептуальной модели INTERMI существует два разных вида взаимоотношений: простой и сложный.

---

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

Для создания простой взаимосвязи в проекте INTERMI нужно установить связь с другим энтитетом (напр. связь между двумя людьми) с возможностью характеризовать роль взаимосвязанных энтитетов (напр. брат/сестра, жена). Можно также присоединить добавочную информацию о взаимоотношении, напр. даты и примечания.

Некоторые взаимосвязи являются сложными: для того, чтобы выразить целый комплекс взаимосвязей в его полноте нужно определить и использовать одно “описанное” событие в жизни или истории энтитета (напр. рождение, возникновение, создание, уничтожение) и целый ряд взаимоотношений, которые с ним связаны.

### **Запись энтитета в проекте INTERMI - структура данных, касающихся энтитета**

В проекте INTERMI энтитет описывается/документируется в записи, содержащей ряд связанных между собой элементов данных, которые сохраняются и обрабатываются вместе. Типичная запись энтитета INTERMI включает в себя следующие сведения:

<b>Идентификатор</b>	<b>INTERMI ID</b> – уникальный, обязательный элемент; допускается несколько идентификаторов, напр. ID национальных авторитетных файлов, ID музейных авторитетов, ID VIAF
<b>Класс/подкласс</b>	Обязательный элемент
<b>Язык</b>	Язык, на котором записываются примечания и дополнительная информация – чешский язык
<b>Правила</b>	Применяются в описании атрибутов и информации, касающейся энтитета, напр. правила RDA, AACR2, CCO, Основные правила (архивы), правила INTERMI
<b>Датировка</b>	Простая в форме текста и комплексная (исчерпывающая информация), которая позволяет отображение и поиск
<b>Замечания</b>	Публичные и непубличные
<b>Краткая характеристика, описание, история</b>	Краткая характеристика: максимально два предложения; описание: максимально 12000 знаков
<b>Обозначение</b>	Предпочтительные и вариантные формы, каждая форма идентифицирована правилами; можно использовать максимально одну предпочтительную форму для одних правил
<b>Части обозначения</b>	Главная часть, другая часть обозначения, дополнительная часть обозначения (тематическая, географическая, хронологическая информация)
<b>Классификация</b>	Характеристика энтитета, выраженная с помощью тематического термина, категории
<b>Отношения и события, использованные в описании энтитета</b>	Простое отношение выражено с помощью связи с другим энтитетом (с названием роли соответствующего энтитета, датировкой отношения и с примечанием); комплексная взаимосвязь выражается с помощью событий, связанных с несколькими типами энтитетов в разных ролях, с разной датировкой и примечаниями. Различаем три типа “описательных” энтитетов: начало, конец существования энтитета, изменение энтитета

<b>Визуальная презентация энтитета</b>	Напр. рисунок
<b>Источники информации, касающейся энтитета</b>	Источник информации приводится к целой записи или к специфической особенности; используется формализованное описание

Таблица 2. Запись энтитета в проекте INTERMI

### **Правила концептуальной модели INTERMI**

Правила концептуальной модели INTERMI представляют собой инструкции и указания, применяемые для формулирования данных, типичных для учреждений – хранителей исторической памяти в Чехии. Целью не было создание специфических индивидуальных правил именно для проекта INTERMI. В то же время мы хотели удержать оправдавшие себя методы и уникальность подхода хранителей исторической памяти там, где это было желательно. Мы внимательно изучили правила, которыми руководствуются учреждения – хранители исторической памяти (AACR2, ISAAR CPF, CCO, RDA) и пришли к выводу, что невозможно применить эти правила во всей их полноте. Тогда мы решили использовать правила RDA во всех случаях, где это возможно. Тем не менее для описания специфических областей энтитета, напр. описания наименований, описания датировки или дефиниции правил для мапирования контролируемых словарей, используемых в учреждениях – хранителях исторической памяти, были сформулированы специфические инструкции. Эти специфические инструкции соответствуют напр. правилам „Cataloguing Cultural Objects” (напр. описание датировки) или „Основным правилам для архивов в Чехии“ (напр. описание предпочитаемых и вариантных форм коллектива).

### **Правила INTERMI для обозначения энтитета**

Разные группы пользователей учреждений – хранителей исторической памяти в Чехии привыкли работать с различными предпочитаемыми формами для обозначения одного и того же энтитета (напр. энтитета общего понятия или энтитета коллектива). Эта идея является ключевой для правил INTERMI: допускается создание нескольких предпочитаемых названий энтитета в одной записи. Тем не менее необходимо у каждого такого названия привести правила, в соответствии с которыми оно было создано. Этот принцип позволит нам предложить на рассмотрение несколько точек зрения на данные энтитета, исходя из правил, которыми руководствуются отдельные учреждения (что также поможет в процессе мапирования тезауруса). Мы думаем, что принцип нескольких предпочитаемых форм наименования в одной записи будет в настоящее время целесообразным решением, однако надеемся, что на практике сообщества хранителей исторической памяти будут использовать только одну предпочитаемую форму, так как поймут, что это проще и практичнее.

### **Правила INTERMI для датировки в описании энтитета**

Для ввода датировки в описание энтитета мы решили применить принципы CCO. Эти принципы базируются на вводе данных в двух формах – доступ для пользователя и доступ для компьютера. Для пользователя датировка записывается в текстовой форме (напр. приблизительно г. 1894), для машинной обработки данные записываются в

формате ISO 8601, где первая приведенная дата означает потенциальную начальную дату, а вторая потенциальную заключительную дату (напр. 1890–1900).

### **Правила INTERMI для мапирования контролируемых словарей, реестров и т.д., которые используются в учреждениях – хранителях исторической памяти**

Целью этих правил INTERMI является суммаризация методов, используемых для мапирования различных словарей, которые содержат энтитеты общих понятий. Мы предполагаем, что эти правила станут необходимыми и для учреждений, которые создают собственные специфические контролируемые словари, отвечающие требованиям Semantic Web (семантической паутины).

Процесс мапирования/гармонизации зависит от:

- выбора нормализованной системы терминов, которые будут служить как основа для сравнения и оценки терминов из других терминологических источников,
- выбора инструментов для формальной репрезентации терминологических ресурсов (SKOS),
- описания и характеристики (формальной и семантической) терминов, которые используются в терминологических источниках,
- выбора методов, используемых в процессе мапирования/гармонизации, таких как метод, основанный на лексическом сравнении, на понятийном аппарате и на мапировании на базе инстанций.

Смыслом лексического мапирования является установление степени подобности текстовых рядов.

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

Рисунок 2. Пример идентичных терминов из трех источников, полученный с помощью метода, основанного на лексическом сравнении

Метод сравнения на концептуальном уровне используется в случае, когда нужно сравнить терминологию специализированных тезаурусов и контролируемых словарей. Это реализуется как метод согласованных графиков или метод векторного сходства.

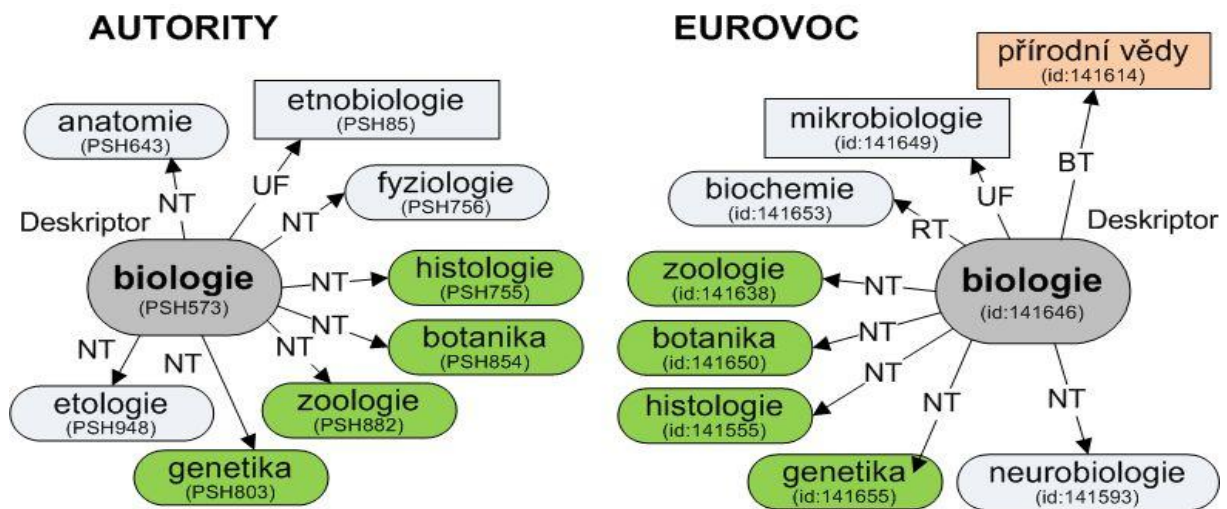


Рисунок 3. Пример применения метода согласованных графиков – подграфик понятия “биология” из терминологических ресурсов: “Национальные авторитетные файлы” и “EUROVOC”

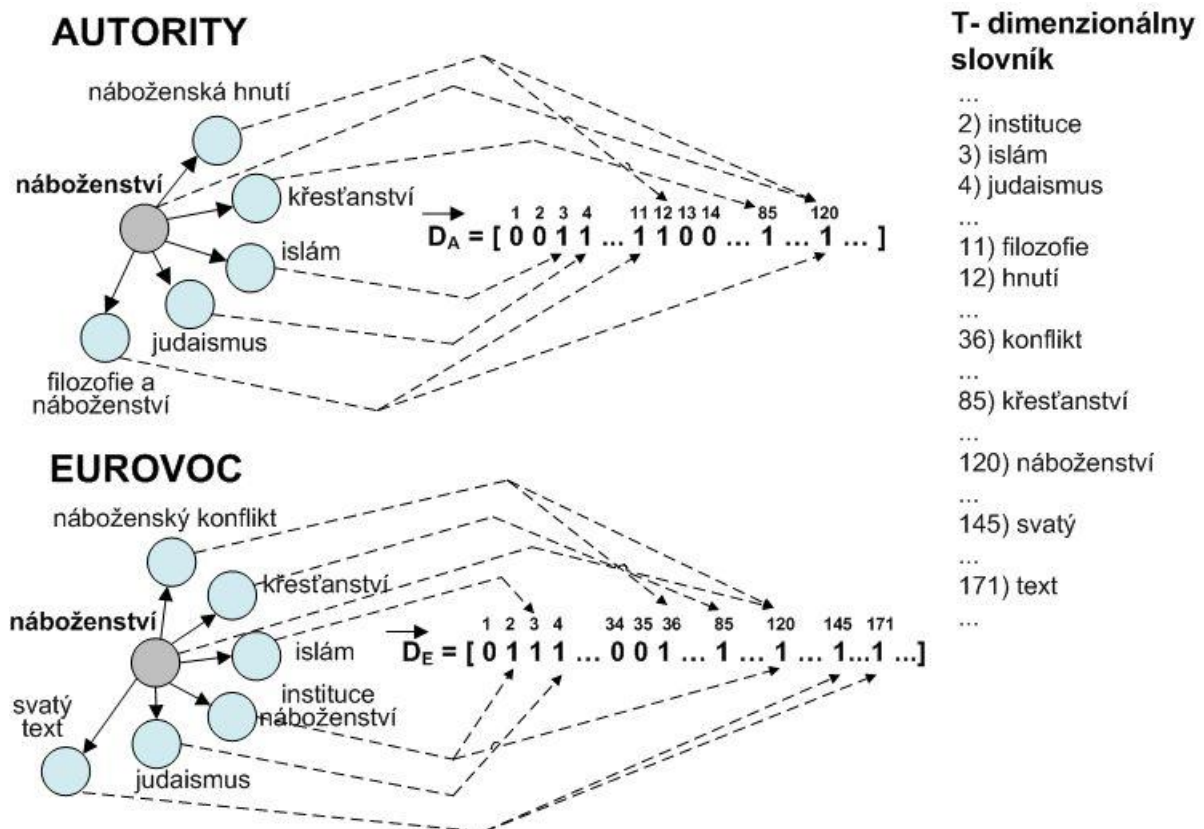


Рисунок 4. Пример применения метода векторного сходства - Т-мерный вектор понятия “религия” в терминологических источниках “Национальные авторитетные файлы” и “EUROVOC”

Благодаря методу сравнения на базе инстанций, на основе лексикального и семантического сходства можно осуществлять поиск эквивалентных, вышестоящих, нижестоящих и соотносительных понятий.

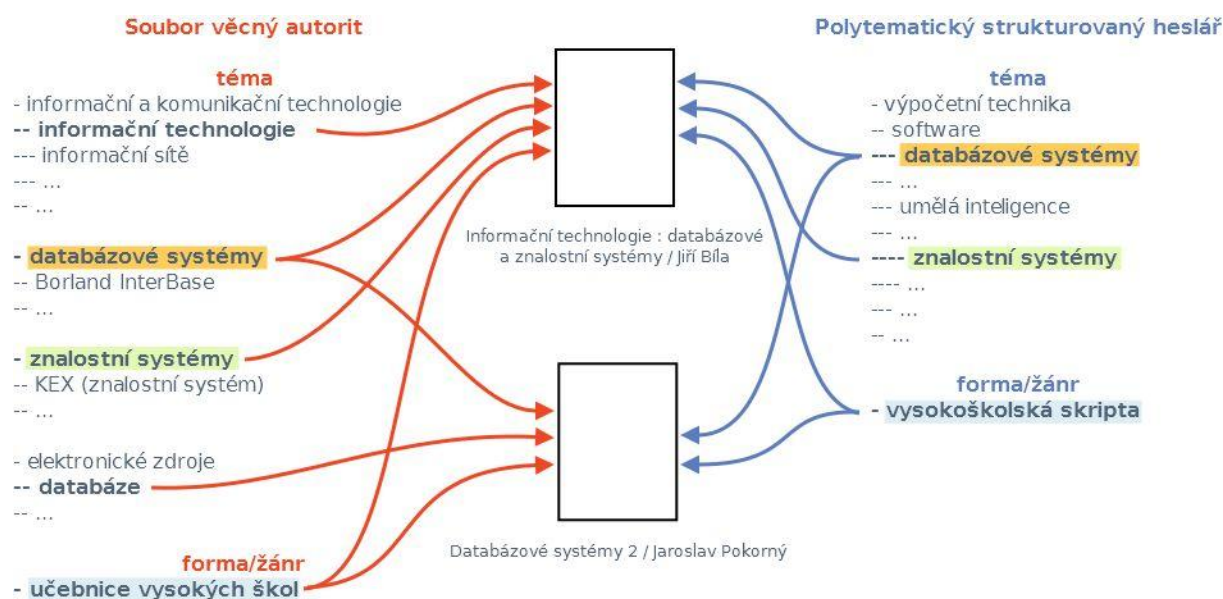


Рисунок 5. Пример применения метода сравнения на базе инстанций

### Роль тематических авторитетных файлов

Тематический авторитетный файл был выбран как основа для мапирования/гармонизации терминологических источников, которые используются в учреждениях – хранителях исторической памяти. Тематические авторитетные файлы представляют собой универсальный источник терминов, которые покрывают все области человеческого знания как на общем уровне, так и на конкретном. Что касается чешских авторитетных файлов, мы намереваемся создать на их основе инструмент, который можно использовать для эффективного мапирования/гармонизации. Это означает их трансформацию в “общую” онтологию. Все это предполагает

- точно определить значение тематических терминов с помощью дополнения квалификатора, определителя УДК и вспомогательных примечаний так, чтобы тематические термины могли лучше представлять понятия,
- изобразить иерархическую структуру так, чтобы она была полной и консистентной,
- использовать постоянный уникальный квалификатор, который не должен зависеть от языка, чтобы не возникла необходимость модифицировать ID в связи с изменением предпочтительной формы, использованной в описании понятия энтитета,
- изобразить тематический авторитетный файл в машинопонимаемой форме в рамках Semantic Web с использованием аппликации RDF - SKOS. Это позволит установить связь и интегрировать между собой данные различных аппликаций RDF в среде Semantic Web.

### Заключение

#### Значение проекта INTERMI

Проект INTERMI является местом для создания и хранения данных энтитетов, которые используются в учреждениях – хранителях исторической памяти Чехии. Он



представляет собой простой в обращении инструмент для доступа и передачи содержания культурного наследия между всеми хранителями исторической памяти, обеспечивая таким образом семантическую интероперабельность на концептуальном уровне.

Цель проекта – создать концептуальную модель и онтологию для того, чтобы чешское культурное наследие могло быть доступно в сети. Проект строится на основе уже существующих исследовательских решений и проектов, среди которых напр. “Национальные авторитетные файлы” и “Музейные авторитетные файлы”, но в отличие от них он стремится создать комплексную инфраструктуру для построения модели знаний для обработки содержания культурного наследия и обеспечения его доступа в качестве прототипа. Этот проект дополняет такие проекты как “Национальная электронная библиотека” и “Национальный электронный архив”.

### **Как влияет проект на сотрудничество между учреждениями – хранителями исторической памяти.**

Необходимо подчеркнуть, что проект оказывает положительное влияние на сотрудничество между разными экспертными группами в этих учреждениях. У экспертов есть возможность встречаться и договариваться о своих требованиях к идентификации и описанию энтитетов, находящихся в их коллекциях. Мы предполагаем, что в конце проекта INTERMI специалисты между собой будут эффективно сотрудничать, что окажет практическое влияние на их каждодневную работу при описании информационных ресурсов с помощью энтитетов INTERMI. Для достижения этой цели в проекте будет предусмотрен ряд сетевых услуг для работы в локальных информационных системах.

### **Как содействует проект доступу пользователей к информации**

Использование энтитетов INTERMI в ходе описания коллекций учреждений – хранителей исторической памяти повлияет также на уровень доступа пользователей к национальному культурному наследию, находящемуся в сети. В то же время проект INTERMI демонстрирует пути предоставления информации об энтитетах с помощью применения технологий Semantic Web.

## **Литература**

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 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.