ELSEVIER

The International Congress on Management of Cultural Heritage and Cultural Memory Institutions (URLA 2014)

Interoperability in memory institutions in Czechia: INTERMI project

Marie Balíková[a1], Jana Šubová[b], Miroslav Kunt[b], Nadežda Andrejčíková[b]

aNárodní knihovna České republiky, Klementinum 190, Praha 1 110 00, Czech Republic
bCosmotron Bohemia, s. r. o., Pančava 415/11, Hodonín 695 01, Slovak Republic

Abstract

The paper informs about the project "Interoperability in memory institutions" (INTERMI) realized by the National Library of the Czech Republic and National Archive and funded by the Ministry of Culture of the Czech Republic. The purpose of the project is to create a conceptual model of knowledge which meets the requirements of the users of all memory institutions in Czechia. The project is based on a new paradigm, which is focused on processing of entities enlarged by necessary semantic information and on complex relations among them.

Selection and peer-review under responsibility of Dr. Nevzat Özel and Tolga Çakmak.

Keywords: interoperability; national project; memory institutions

1. Background

   "Interoperability in memory institutions" (INTERMI) DF1P01OVV023 is a five year project (2011-2015) realized by the National Library of the Czech Republic and National Archive within the Programme of applied research and development of national and cultural identity (NAKI) funded by the Ministry of Culture of the Czech Republic.
   INTERMI project represents a common project of memory institutions in Czechia. The cooperating group consisting of libraries, archives, museums and galleries aims to create a uniform and standardized, user-friendly access to national cultural heritage.
   At the beginning of the project, a thorough research and analysis were conducted in all participating institutions with the aim to identify:

---
[1]* Corresponding author. Tel.:+420 221 663 352; fax: +420 221 663 301.
   E-mail address: marie.balikova@nkp.cz

- type of information resources which form part of collections of individual institutions,
- rules, formats, standards used for describing information resources,
- cooperative projects,
- controlled vocabularies, indexes, registers offered to professionals and general public,
- influence of advent of WWW,
- user expectations in accessing information.

Libraries are responsible for information access to written and printed cultural heritage, galleries for collecting and displaying art objects, museums for the research and the availability of three-dimensional objects/artifacts on both object- and collection-level as well; the main goal of archives is to preserve information as evidence of the creator's activities. In the era of mass digitisation, all participating institutions are responsible for creating digital surrogates of information resources to facilitate their access and reproduction. The differentiation of memory institutions according the type of information resources is important because the type of information source influences rules, standards, formats, and current best practice applied by the memory institution.

1.1. Libraries

Since the nineties of the last century, libraries in Czechia have been using international standards such as AACR2, ISBDs, MARC format. Name and Subject Authority files have been created and widely used among libraries in order to ensure standardized access points. A cooperative system for creating and sharing authority records of all types has been created, as well.

1.2. Museums, galleries, archives

The situation in museums, galleries and archives in Czechia was quite different. Unlike libraries that collect information sources of mass production (books), museums tend to collect unique objects. They create metadata/descriptions of high quality both of unique objects, or their sets (collections) as well, but they apply their own rules valid only for one institution, one collection. Therefore their options to share metadata/descriptions/catalogue cards were very limited.

Moreover, each scientific discipline has its own methodological approach and uses its own terminology in describing and accessing their collections. Archives and museums use highly specialized terminology. A wide range of home grown indexes and vocabularies are used by the museums, galleries and archives in Czechia from very specialized to very general, but often their terminology is limited to local (one collection, one institution, one system) practice. Standardized access points don't exist.

With the advent of World Wide Web, a system of interlinked hypertext documents that are accessed via the Internet[1], the possibilities for the presentation of cultural heritage have greatly grown.

Now, memory institutions are digitizing millions of books or collection objects when participating in mass digitization projects. Consequently, a need emerged to set up uniform and standardized rules and conceptual models which could be applied in the domains of memory institutions, so that metadata/descriptions can be shared and reused.

---

[1] http://en.wikipedia.org/wiki/World_Wide_Web

Library community started and still has been developing RDA rules with the aim to make them the standard for memory institutions as well. Conceptual models for bibliographic universum have been developed and harmonised. Museums and galleries apply CCO rules and develop their own conceptual model CIDOC CRM.

In collaboration with libraries, they are participating in cooperative projects based on national authorities.

The archivists translate international standards and based on them create their fundamental rules. All memory institutions are aware of the need to provide access to their collections to the general public.

The public is moving away from institutional and library catalogues, and prefers commercial services. Therefore, memory institutions need to respond to this situation and adapt their published data to this trend: they publish memory institutions data as linked data.

The analysis shows that a consensus on one particular standardized description and formal representation of the data would be very difficult. We found it more convenient to concentrate on standardization of access points that are both subject of interest of the users, and the basic principle of Linked Data. In effect, this means moving away from authorities to entities (real world entities). This is precisely the subject of our research within INTERMI project.

2. Interoperability in memory institutions - INTERMI project

The aim of INTERMI project is to create a scientific and technological infrastructure to support the processing, sharing and use of the cultural content in the form of metadata about information objects preserved in memory institutions. The information objects represent objects of the real world, e.g. persons, institutions, activities, three-dimensional objects (i.e. artistic and technical objects, objects of inanimate nature), events and performances, artistic and other achievements as well as their textual, visual and sound representations (digital included), which are subject to the activity of memory institutions.

A basic prerequisite for identification of the information objects are both subject and name authorities, in the extent defined by the needs of libraries. The goal of this project is to extend the function of authorities - unified selection of elements/access points - by adding related information objects. In this way, their information and documentation potentiality will increase so that they can not only meet the requirements of different types of memory institutions, but create a much larger basis for semantic interoperability of cultural content for all of them.

One of the main results of the project will be a tool for user-friendly and differentiated acquisition and access of information about these information objects. This will also create a basic prerequisite to access and share the cultural content across memory institutions ensuring semantic interoperability at the conceptual level. All memory institutions as well as the general public are invited to use the proposed system.

The project is complementary to projects such as those on the National digital library, the National digital archive, the establishment of an integrated system of management of collections, and the methodology of tools for processing and providing access to the historical collections in Czechia.

Current partial results of INTERMI project include:

- INTERMI conceptual model on abstract and specific level,

- definition of seven classes of entities (personal, family, corporate body, geographic object, work, event, and general object entities),
- web interface for processing of data of entities,
- proposal of rules for data processing based on both INTERMI conceptual model and data structure, and rules applied in different memory institutions.

2.1. INTERMI conceptual model

INTERMI conceptual model is one of the most important outputs of the first stage of INTERMI. The INTERMI conceptual model can be described in two levels: abstract and specific.

The abstract level of INTERMI conceptual model is inspired by CIDOC CRM and by principles of object-oriented paradigm. It has been decided not to apply existing conceptual models in their entirety, because when we describe an entity, much more information about this entity is required, especially with regard to the properties and relations. It means that we met with insufficient complexity and usability of existing models.

As mentioned above, we concentrate on information objects connected to items in collections, not on items as such. To fulfil requirements defined by memory institutions we create an abstract level of INTERMI conceptual model based on following principles:

- data granularity – data are divided into smaller meaningful elements which allows to define better type of data and to generalize them,
- data inheritance – hierarchical structure of elements requires and supports inheritance of attributes or properties,
- data reuse – elements are designed to be reused in more than one location in the model and to enable further expansion of the model,
- polymorphism – element can work as different element depending on their use,
- event-based description – as many entity attributes as possible are described by event; we think that the event-centric approach reflects the concept of linked data.
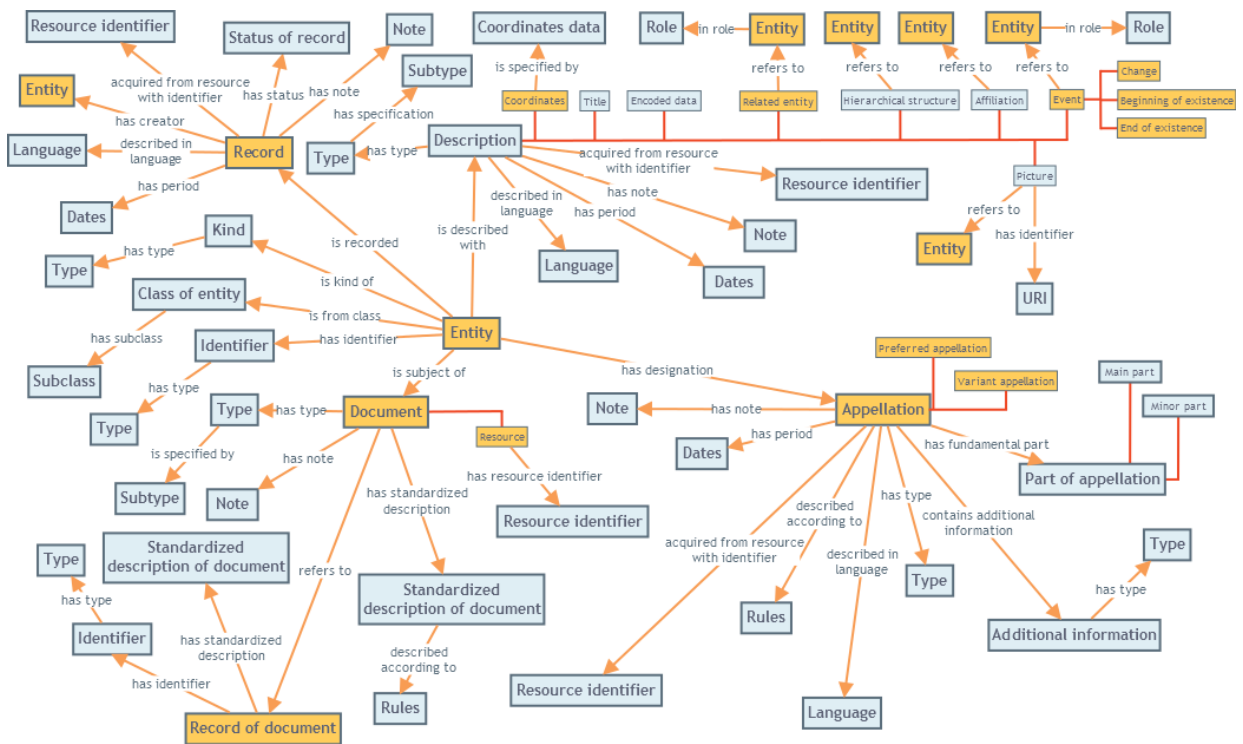
Fig. 1. INTERMI conceptual model – abstract level

Specific level of INTERMI conceptual model is realized in 7 classes of entities separated to subclasses with corresponding attributes, relations and events.

INTERMI conceptual model is designed as convertible to CIDOC CRM, and will be implemented from technological point of view in structure based on XML. In this state of project we are considering various standards related to linked data concept and we are trying to find one or more suitable standard for data presentation. Internally – as to data storage – we are using XML based proprietary structure that cover INTERMI conceptual model in its width and ensures system flexibility.

2.2. Seven classes of entities

The INTERPI conceptual model on specific level was defined after intensive interdisciplinary discussion. The main classes of entities are those of:

- personal entities,
- family entities,
- corporate body entities,
- geographic object entities,
- work entities,
- event entities,
- general object entities.

Definition of a class is based on a level of common characteristics. Main goal of definition of classes was to easily determine and describe characteristic of entities belonging to specific class. However it is allowed to use characteristics from one class to describe entity primarily categorized in another class (examples are castles, dams etc.). Using this principle we accept a complexity of real universe of entities.

Table 1. Summary of classes of entities with number of subclasses

Class of entities	Number of subclasses
Person/creature	4
Family	3
Corporate body	12
Geographic object	9
Work	7
Event	3
General concept	7

2.2.1. Person/creature – class of entities

This class of entities includes real human beings, fictional human beings and also named figures such as religion figures, figures from fairy tales, supernatural figures, animals etc.

Regarding to this class we have to solve some crucial problems followed from cooperation between various types of memory institution. The main issue was to define what exactly entity of this class is. Is it person in its complexity – it means with all variation of names and identities? Or, is it entity in a sense of bibliographic entity – it means variation of names that appears in documents connected by topic, or area of interest? As a result we offer two ways how to record information about person with more fictional identities. The first (and recommended) way is to describe person with more fictional identities in one record. It is necessary to choose one preferred appellation that is more used, more known etc. Other appellations are recorded as variant appellations. This method is convenient in the case when person use both fictional and real identity with the same biographical data. The second way supported by INTERMI rules is to describe real and fictional identity separately – in two more records. This is useful in cases of different biographical data used by alternative identity (as fictional curriculum vitae). It is necessary to create relation between all records connected to one entity.

Further problems to be solved by INTERMI rules are those of collective pseudonym, anonymous creator and unknown creator. A group using collective pseudonym is included in this class of entities if the name they use is similar to that of person. Specific entities as anonymous creator and unknown creator are also included in the "person/creature" class of entities.

2.2.2. Family – class of entities

"Family" class of entities is constituted by families and parts (branches) of families. In connection to this class we have to mention different approach applied in archives that concede a concept of families as corporate bodies (as named and organized group of people). In the INTERMI project we decided to create separate class of these entities to describe specific characteristics of families also with complexity of problems of genealogy.

2.2.3. Corporate body – class of entities

"Corporate body" class of entities was one of the most discussed classes regarding to different methods of description applied in archives and libraries. Rules applied in libraries leave out a term describing type of corporate body (e.g.: ltd, plc), however for archivists type of corporate body it is an important attribute that determine separated entity. In this case, INTERMI rules adopt principles from archives and include type of corporate body as one of attributes which determine corporate body entity.

Other discussion focused on differences between corporate body, works and geographical object regarding buildings (e.g. castles) and cities. It is necessary to distinguish entities as bodies of administration (belonging to "corporate body" class of entities) and entities – delimited and named part of nature – created by humans as a place for living (belonging to "geographic object" class of entities) and entities – buildings (belonging to "work" class of entities).

2.2.4. Geographic object – class of entities

In INTERMI project we define geographic object as entity determined with stable set of geographic coordinates. The development of geographic object is usually described in one record with variant appellations specified by dates when appellations were used. If necessary, INTERMI allows application of separated record principle for every significant stage of entity development (with corresponding relations among records).

2.2.5. Work – class of entities

It should be noted that INTERMI project is the first attempt of a "work" concept application in Czechia. The concept of "work" appeared in FRBR conceptual model; its abstract specification causes problems in its application in library catalogues. In museums and galleries, the FRBR concept of "work" met with misunderstanding because of their different meaning of the concept (as "art work" – product of art creative power).

Definition of work that belongs to "work" class of entities required to pay attention to abstract meaning of the concept of work. "Work" arose from intentional human activity and the description is concentrated to its current state. One work in its complexity is described in one record. If necessary, it is possible to create separated records on significant historical stage or part of work (with relations among records).

"Work" class of entities consists of entities that have some material representation, but also of entities which do not exist in representation any more. This class include art works, general documents, standard documents, products, trademarks, games, grants, projects, buildings etc.

2.2.6. Event – class of entities

"Event" class of entities includes a named temporary event (long-term or short-term; one-time or repeated) and also named entities which are related to human activity such as culture, folk habits.

According to INTERMI rules, every repetition or interpretation of event is described as separated entity. It is possible to create a collective (cover) record that consists of information related to complex history of events. This record is related to all repetitions/interpretations.

2.2.7. General concept – class of entities

This class includes general concepts for specific entities such as categories of people, things, animals, plants and also general concepts for abstract entities such as characteristics, scientific areas, art styles, etc.

2.3. Webinterface for data processing

Webinterface for data processing is another important result of INTERMI project. The aim of INTERMI webinterface is to provide simple, intuitive tool for creating record of entity with accordance to INTERMI rules. From technological point of view it is created using Sencha Ext JS tool and designed with CSS according to INTERMI design manual.

Webinterface is available only for registered users and it uses HTTPS as protocol for secure communication. Users' permissions can be set on various levels:

- level of class of entities – it means that user can modify and create only records of entities which belong to specified class of entities,
- level of group of records – either records of one institution, or records of a network of institutions - it means that user can modify records which have been created by users belonging to the same group and create only records for this specific group,
- level of transaction type – transactions mean creating, modifying, accrediting, deleting (see figure 2) – it is possible to set transactions which are permitted to each user – in this way users are divided in two groups, supervisors and cataloguers.
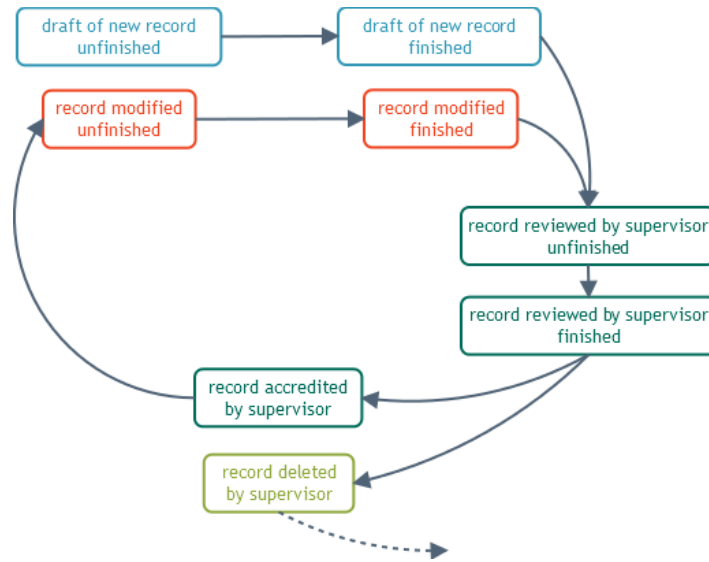
Fig. 2. INTERMI webinterface – life cycle of record

INTERMI webinterface has two main parts – searching and creating/modifying. The searching part is designed as popup window and it is used as mandatory in the beginning of processing. We stress in INTERMI rules and also in webinterface manual that it is necessary to search for a record as the first, with the aim of avoiding multiplicity of entities. According to this important principle we decided untraditionally to put dropdown menu and button for creating new record only to the searching window (see figure 3).

Fig. 3. INTERMI webinterface – search window

The second part of interface – creating/modifying – contains tabs with records selected for creating or modifying, the heading with the identification of entity, and form located in the main part of the screen where it is possible to fill in information about the entity. Simple and functional design using tabs with groups of information (on the left side) and all possibilities of web-based forms (as drop down menus, radio buttons, popup windows etc.) enables to create description of entity easily despite wide range of information. The form for each class of entities is designed in different color, that on the one hand is useful for clear arrangement of interface, and on the other hand colorfulness of interface indicates interoperability among various cultural heritage institutions.

Fig. 4. INTERMI webinterface – creating/modifying

2.4. INTERMI rules

According to previous results of INTERMI project it was necessary to create INTERMI rules for creating of knowledge model (data processing). It was not possible to apply any existing rule in completeness (e.g. AACR2, RDA, ISAAR CPF, CCO), because of new approach to entities and coverage of INTERMI project. On the other hand wide coverage of INTERMI project appears as advantage, because we had opportunity to evaluate present rules and to combine selected principles on the basis of best practices.

The aim of INTERMI knowledge model rules was to summarize main principles used for entity description. INTERMI knowledge model rules are based on RDA principles. For specific areas of characteristics of entities we created specific principles. To describe, in this study, the most interesting parts of INTERMI knowledge model rules, according to INTERMI conceptual models, we chose:

- entity appellations – preferred and various,
- events and relations,
- dates.

One of the basic problems which we had to solve in INTERMI project was connected to different needs of memory institutions in creating appellation of entity. We found out that it was not possible to agree on one preferred appellation form in some cases (e.g. general object entity, corporate body entity) and various communities of memory institutions were afraid of losing their uniqueness. INTERMI solution described in INTERMI knowledge model rules allows to create more than one preferred appellation of entity – however it is necessary to assign each appellation to those rules according which it was created. This principle enables to provide multiple views on data of entity according to rules used in each memory institution. This is also useful in thesaurus mapping. We hope that the principle of multiple preferred appellation of an entity is main forthcoming solution, but in practical application the communities of memory institutions will be able to use one preferred appellation – because they will find it easy and reasonable.

As mentioned above, INTERMI conceptual model is based on events, therefore INTERMI knowledge model rules contains instructions how to create and use events and relations in description of an entity. It is significant that in INTERMI we entirely create relations among entities on the basis of unique and permanent identifier of entity. This is first step to publish INTERMI data with technologies of Linked Data. To create relation in INTERMI knowledge model rules means to create a link to another entity (e.g. a person to person link) with possibility to specify role of linked entity (e.g. sibling, wife). It is possible to enter additional information about relation – as dates and note. Create event in INTERMI knowledge model rules means to create a set of relations that are connected to one event in the life or history of entity (e.g. birth, establishment, creation, destruction). In one event various relations can appear and it can be described by dates and note, too.

According to using dates in entity description we decided to adopt principles of CCO. These principles are based on entering dates in 2 forms – for user access and for computer access. Uncertain dates are for user access described in text form (e.g. ca 1894), for computer access with a pair of dates in format of ISO 8601, where the first delimits potential initial date, the second potential closing date (e.g. 1890–1900).

Further INTERMI rules are connected to mapping of controlled vocabularies used in memory institutions. The aim of these INTERMI rules is to summarize methods used for mapping various vocabularies to INTERMI general object entities. As the subsidiary result – INTERMI controlled vocabularies mapping rules are important as instructions for institutions that would like to use INTERMI as space for creating specific controlled vocabulary in a way that fit to semantic web. These rules solve problems how to keep specific terms used by specialized communities and how to incorporate them into INTERMI database.

3. Expected use of INTERMI project results – importance of project

In this phase it should be noted that the project has an impact on discussion among different communities from memory institutions. Experts had opportunity to meet together and express their needs dealing with the identification and description of entities from their collections.

With current results (webinterface, rules etc.) INTERMI project provides a space for preservation of data about entities in the form that information systems so far used in memory institutions were not able to offer.

At the end of INTERMI project we foresee an effective cooperation among professionals that will lead to creation of INTERMI knowledge base which would have a practical impact on their everyday practice when describing information resources using INTERMI entities. For realization of this goal, INTERMI will provide a set of web services for application in local information system.

Using INTERMI entities in description of collections will improve quality of user access to information on World Wide Web; at the same time INTERMI project will demonstrate ways to present information about entities with applying technologies of semantic web.