INTERPI

INTERMI
Interoperability in memory institutions in Czechia

Marie Balíková, Jana Šubová, Naděžda Andrejčíková, Miroslav Kunt

URLA2014

Goal of the project

- „Interoperability in memory institutions" (INTERMI)
- 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**

- 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 access to national cultural heritage **accessible on the Web** in **human readable and machine friendly form**

Partners of the project

Funded by

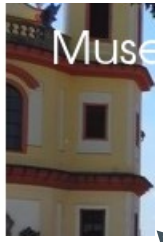
MINISTRY OF CULTURE
CZECH REPUBLIC

National Library of the Czech Republic

Project coordinators

Národní archiv

INTERPI



cosmotron
Cosmotron Slovakia & Bohemia

CITeM
METODICKÉ CENTRUM PRO INFORMAČNÍ TECHNOLOGIE V MUZEJNICTVÍ
MORAVSKÉ ZEMSKÉ MUZEUM

State District Archive
CHRUDIM

Project participants

INTERMI - project's beginning phase

Analysis to identify

- **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 in the Web

Libraries

- well developed **services and library systems**, metadata schemas, and advanced search techniques;
- **authority control** is fully provided for all access points;
- **interoperability** is achieved through the use of standardized access points – national authority files

Weaknesses and constraints of libraries

- **implicit semantics** of metadata schemas based on the MARC format,
- **brief information** in authority records,
- **emphasis on access points** only and **not** on an **expression** of **relationships** among related entities

Exspectations from INTERMI project

- move from authorities to entities
- more information about entities in authority records
- development of an infrustructure of knowledge base of cultural heritage content
- making its content accessible on the Web in human readable and machine friendly form
- sharing cultural heritage the linked open data way

ARCHIVES

- information resources are arranged according **to the provenance principle of archive fonds**
- collections of historical records have been created and maintained for specific purposes such as **administrative, business or legal needs**
- they have been preserved as they **hold long-term archival value**
- **limited access** to information **for general public**

Weaknesses and constraints of archives

- **heterogeneity** of metadata schemas, many **general** and **domain specific data structure standards** developed to meet the needs and criteria of individual user communities,
- variety of **home grown indexes**, registries, inventories

Exspectations from INTERMI project

- use of common metadata schemas, standards, rules
- more information about entities in authority records
- standardized access points
- sharing cultural heritage the linked open data way

MUSEUMS, GALLERIES

- Museums and galleries have rare and unique items (not "self describing") for which traditional bibliographic or archival description is not optimal.
- Description tends to be more subjective
- Limited access to information for general public

Weaknesses and constraints of museums and galleries

- Heterogeneity of metadata schemas, many general and domain specific data structure standards developed to meet the needs and criteria of individual user communities,
- Variety of home grown indexes, registries, inventories, no use of controlled vocabularies, thesauri

Exspectations from INTERMI project

- use of common metadata schemas, standards and rules applied in the domain
- standardized access points
- more information about entities in authority records
- sharing cultural heritage the linked open data way

WITH THE ADVENT OF WORLD WIDE WEB

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
- To **publish memory institutions data as linked data**

Need for sharing and reusing metadata/descriptions, new standards, models…

- Libraries - RDA rules, Conceptual models for bibliographic universum
- Museums and galleries - CCO rules, conceptual model CIDOC CRM
- cooperative projects based on national authorities
- Archives - international standards, new fundamental rules, models.

Moving away from authorities to entities (real world entities)

RESULTS OF INTERMI PROJECT

- INTERMI conceptual model
  - abstract level
  - specific level
- INTERMI web interface for processing of data of entities
  - Searching
  - Creating/modifying
- INTERMI rules
  - INTERMI knowledge model rules
  - INTERMI rules of mapping of controlled vocabularies, indexes used in memory institutions

INTERMI CONCEPTUAL MODEL – ABSTRACT LEVEL

- INTERMI conceptual model – **abstract level**
- inspired by **CIDOC CRM and by principles of object-oriented paradigm**
- based on 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, to enable further expansion of the model,
  - **polymorphism –** element can work as different element depending on their use,
  - **event-based description** – many entity attributes are described by event; we think that the event-centric approach reflects the concept of linked data.

INTERMI CONCEPTUAL MODEL – SPECIFIC LEVEL

Specific level of INTERMI conceptual model

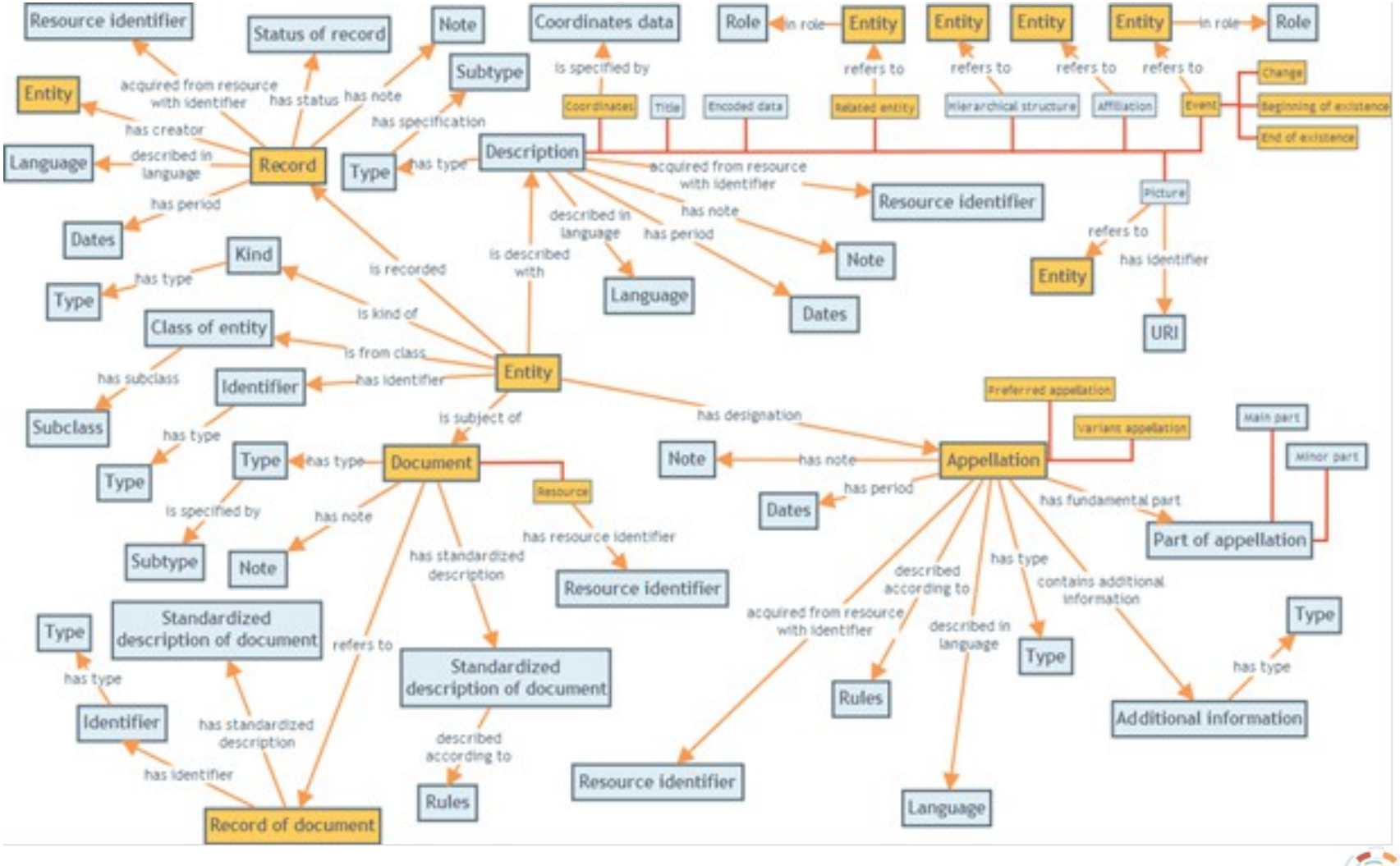
- is realized in 7 classes of entities separated to subclasses with corresponding **attributes, relations and events**.
  - **personal**,
  - **family**,
  - **corporate** body,
  - **geographic** object,
  - **work**,
  - **event**,
  - **general object** entities,

Definition of a class is based on **common** characteristics.

However it is allowed to use characteristics from one class to describe entity primarily categorized in another class (examples are castles, dams etc.).

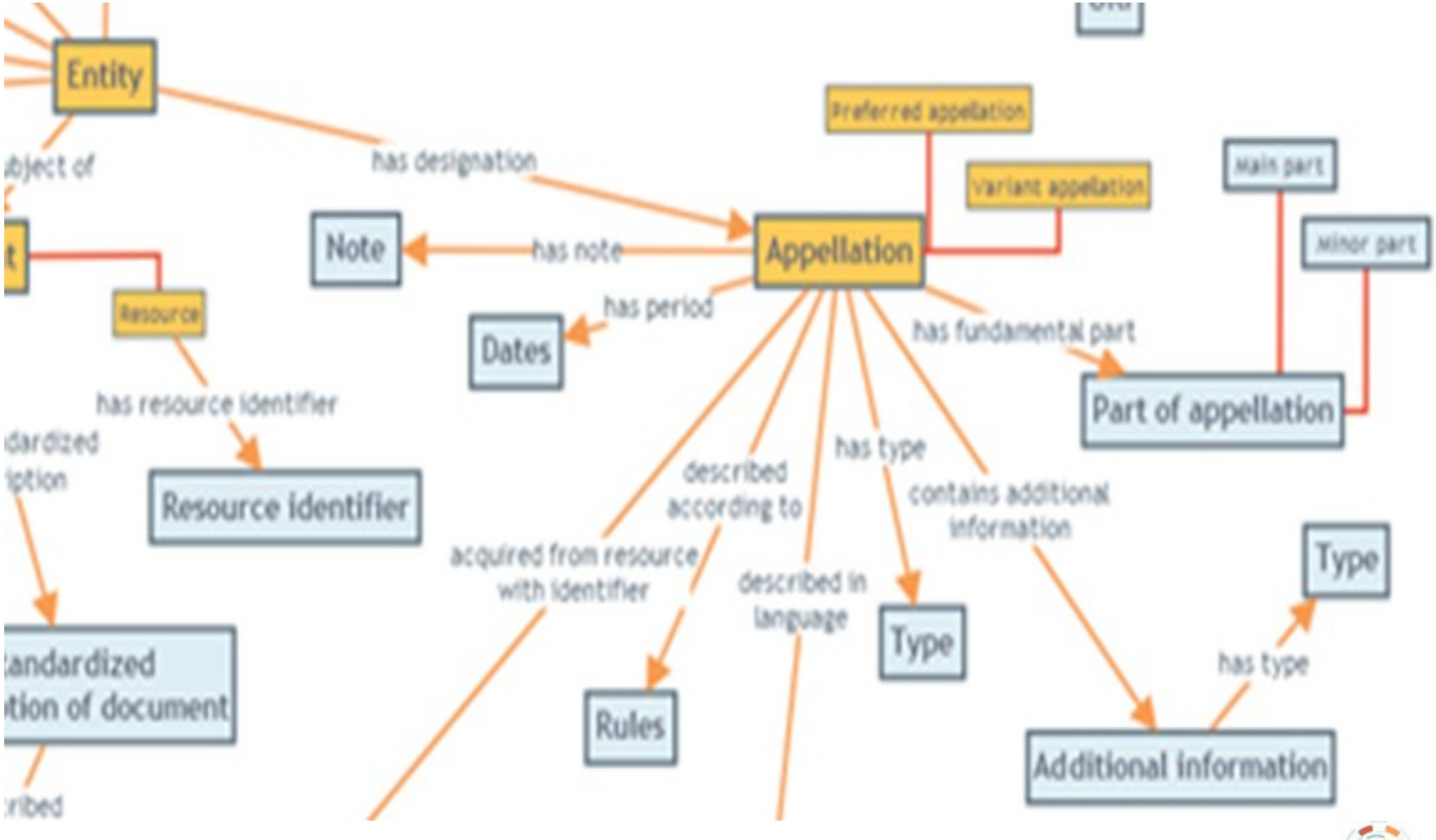
Using this principle we accept a **complexity** of real universe of entities.

INTERMI CONCEPTUAL MODEL – ABSTRACT LEVEL



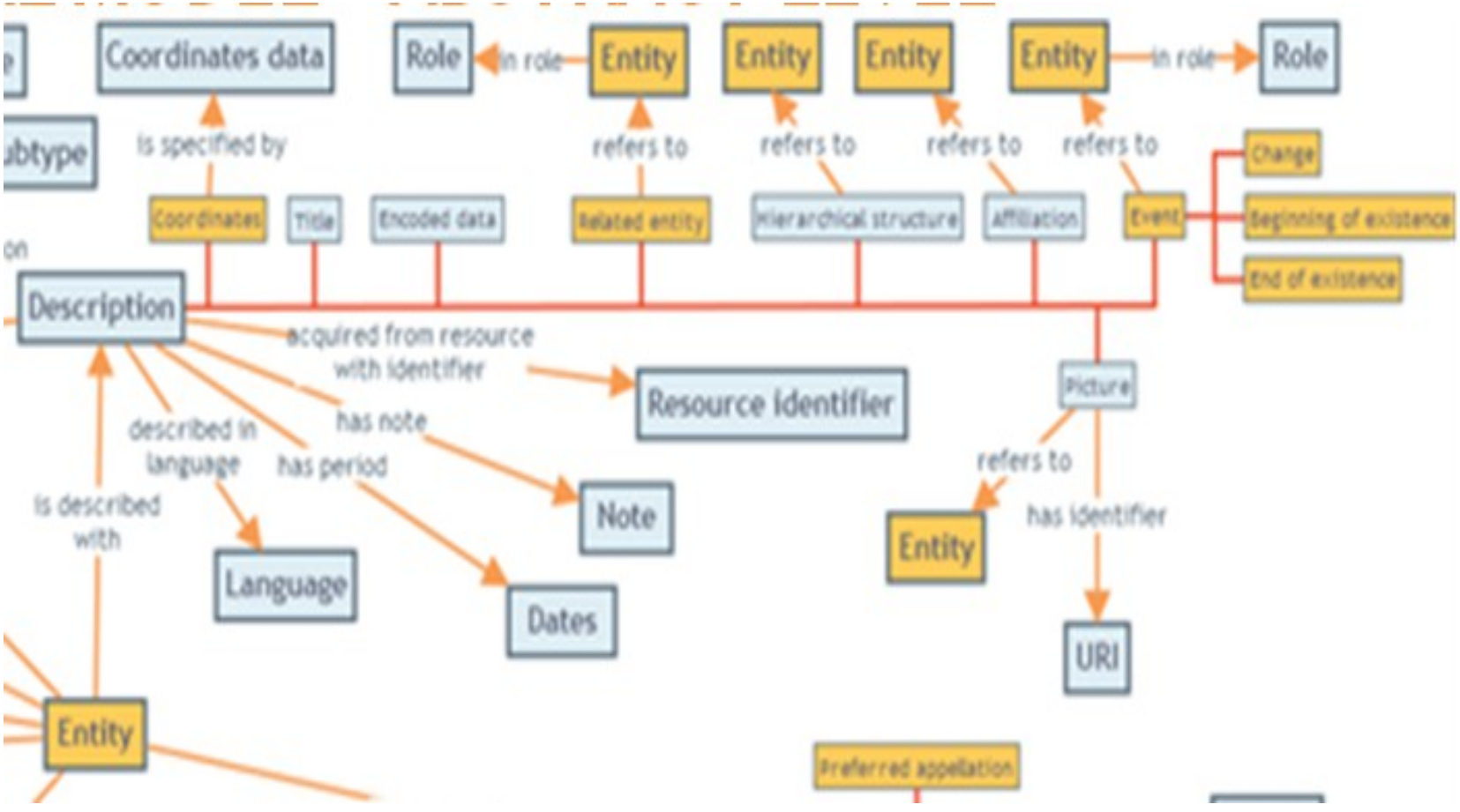
INTERMI CONCEPTUAL MODEL

Attribute of the entity: appelation



INTERMI CONCEPTUAL MODEL

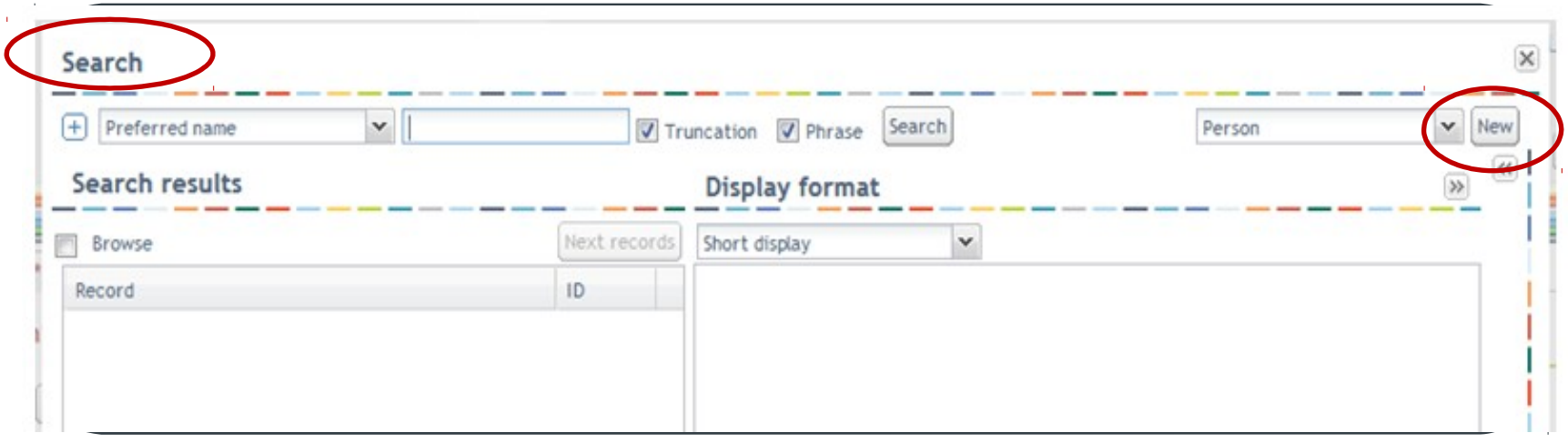
Attribute of the entity: description



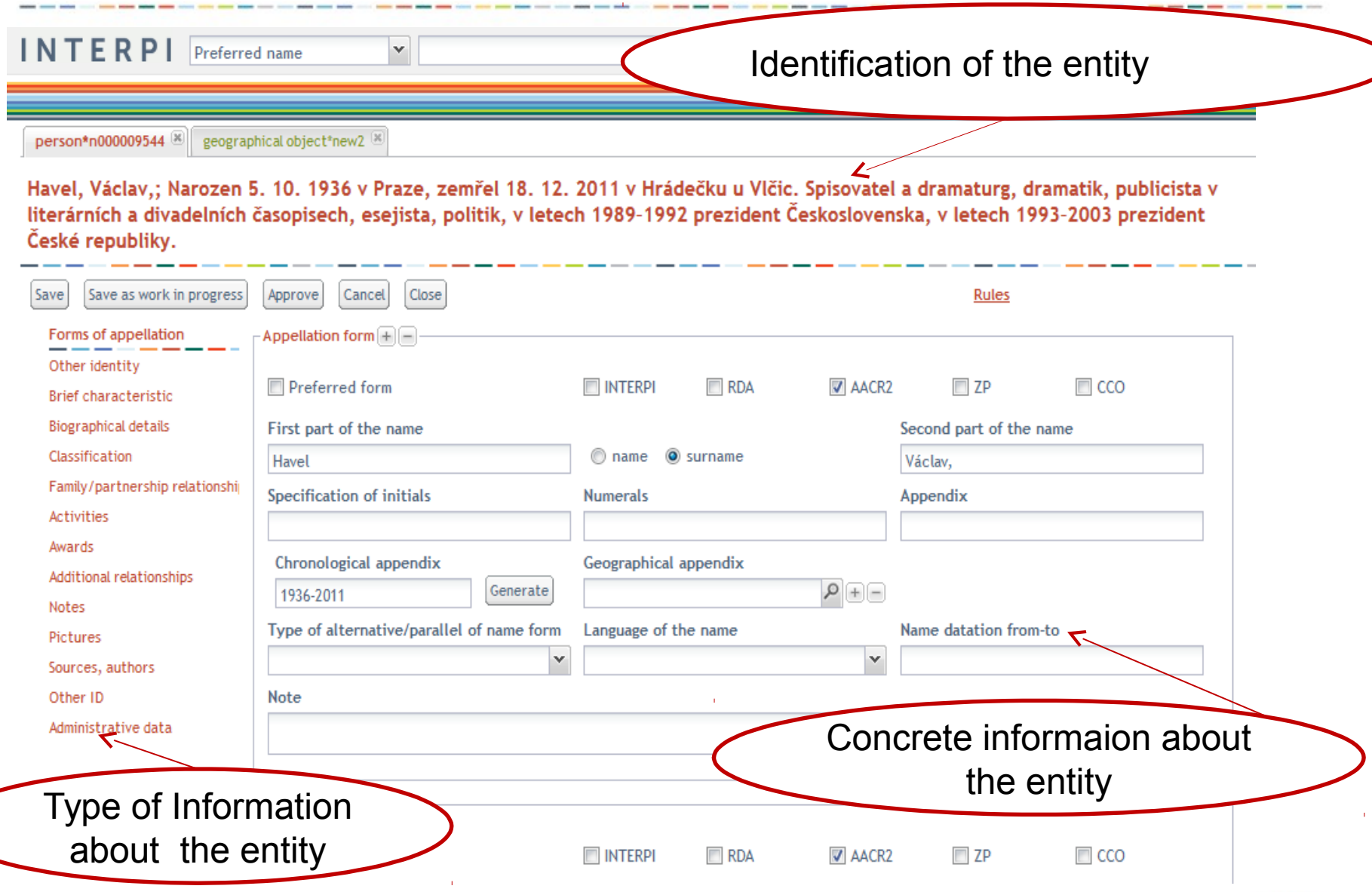
INTERMI WEB INTERFACE FOR PROCESSING: SEARCHING

- Searching part is used as mandatory in the beginning of processing.
- It is necessary to search for a record as the first,
- With this we try to avoid 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.



WEBINTERFACE FOR DATA PROCESSING: CREATE/MODIFY



INTERMI KNOWLEDGE MODEL RULES

- It was not possible to apply **any existing rule in its entirety** (e.g. AACR2, RDA, ISAAR CPF, CCO),
- It was necessary to evaluate main applied rules and to combine **selected principles** on the basis of best practices.
- INTERMI knowledge model rules are based on **RDA principles.**
- For specific areas of characteristics of entities we created specific principles e.g.
  - Identifying corporate bodies we respect the requirement of Archives to preserve terms indicating **the type of a corporate body** and **terms indicating state ownership** of a corporate body – in this case we break RDA rules

INTERMI RULES OF MAPPING OF CONTROLLED VOCABULARIES

- **Mapping of controlled vocabularies** used in memory institutions
- Aim of these INTERMI rules is to **summarize methods** used for mapping various vocabularies to **INTERMI general object entities.**
- 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.

INTERMI PROJECT - EVALUATION

- **Cooperation** among professionals
- Discussion among different communities from memory institutions **to identify their needs** dealing with the **identification** and **description** of entities
- Using **INTERMI entities** in description of collections **will improve quality of user access to information on World Wide Web**;
- INTERMI project will demonstrate ways to present information about entities with applying **technologies of semantic web**.