Knowledge Representation from Classification Schema to Semantic Web (II)

Dr. Silvia-Adriana Tomescu
"Carol I" Central University Library of Bucharest
E-mail: silvia.tomescu@bcub.ro

Dr. Silvia-Adriana Tomescu is head of Research.Methodology Department of "Carol I" Central University Library. Her main academic interests are ranging from e-learning, librarianship, higher education, academic writing to cultural diplomacy and Romanian exile. She published translations and studies in the field of communication and information science.

Abstract

We aim to approach in this essay the technical paths used to represent hibrid documents in online environment using specific standars. The multimedia contents increase and the diverse storage formats need refined instruments for the search and retrieval process. Dedicated applications require a high level of interperability therefore the necesity of standardization. This study tries to argument theoretically the need to set and reflect logically the media objects properties in metadata schema and to illustrate the importance of ontologies and taxonomies in online environment representation.

Keywords: *mediacontent, metadata, taxonomies, ontologies, digital objects*

III. Knowledge Representataion through Media Content

3.1. Knowledge Representation instruments

Knowledge Representation is also a process of Media Content Representation through classification schema, taxonomies, and ontologies, reflected in semantic manner on the web. This process involves the uses of math languages and exchange standards. *Digital objects* are represented through metadata schemas, precisely a set of *semantic descriptions* gathered around a specific object for explaining, locating, making it easier to retrieve in a catalog or collection. Any type of media content needs special information management strategy to improve the creation, filtering and delivery process. This activity is based on both *content* and *context* metadata description and strong interoperability

standards. Metadata description is the mirror of any digital object, proving the basic info about the content. The example given in the figure is a Dublin Core metadata description of a movie in Europeana Library Catalogue. Dublin Core Metadata Element Set (1) is a framework for general, common descriptors. It includes information referring to *title, creator, subject, publisher*, etc. It is not a metadata schema for multimedia but has common elements. The types of the documents from a specific data base or catalogue require a special metadata standard.

Fig. 1. **Metadata description in Europeana of a movie** (Century Journal: Emil Cioran)

The basic elements of this schema refer to: C**ontent**: title, subject, description, language, relation, coverage; **Intellectual Rights**: creator, publisher, contributor, rights; **Instance**: date, time, format, identifier. The **Title**: is name given to the resource; the **Creator**: An entity primarily responsible for making the resource; The **Subject**: The topic of the resource; **Description**: An account of the resource; The **Publisher** is an entity responsible for making the resource available; The **Contributor** is an entity responsible for contributing to the resource.

The language used for Knowledge Representation starts from the idea that a scientific organization and good information depends on well-structured domain or well-structured things. The concepts are organized according to dictionaries and glossaries Knowledge representation languages being the groundwork of the lexicons structure, which organizes the notions according to language rules for an efficient search and

retrieve process; in addition, associates semantic categories in order to refine the search. The management of information in a multimedia environment must take into account:

1. The users and their information and media skills needs.
2. The workflow needs for metadata and information management.
3. The creative media domains: Art, Media, Business.
4. The media objects update.
5. The media documents representation.

The Knowledge Representation instruments are glossaries – concept maps; thesauri – taxonomies; domain maps – ontology; specific representation languages. Ontologies are multifaceted conceptualization of a knowledge domain defined in computer processing programs by entity-relation association. Depending on the area in which they are used [ex. Biology, astronomy] there are specialized or general ontologies. The entity shapes the class of objects; the attributes shape the proprieties of the entities. The members of a class are the *instances* of that *class*. Ontologies are expressed through specialized programming languages as KIF [Knowledge Interchange Format] or through languages based on descriptive logics [OWL; OIL].

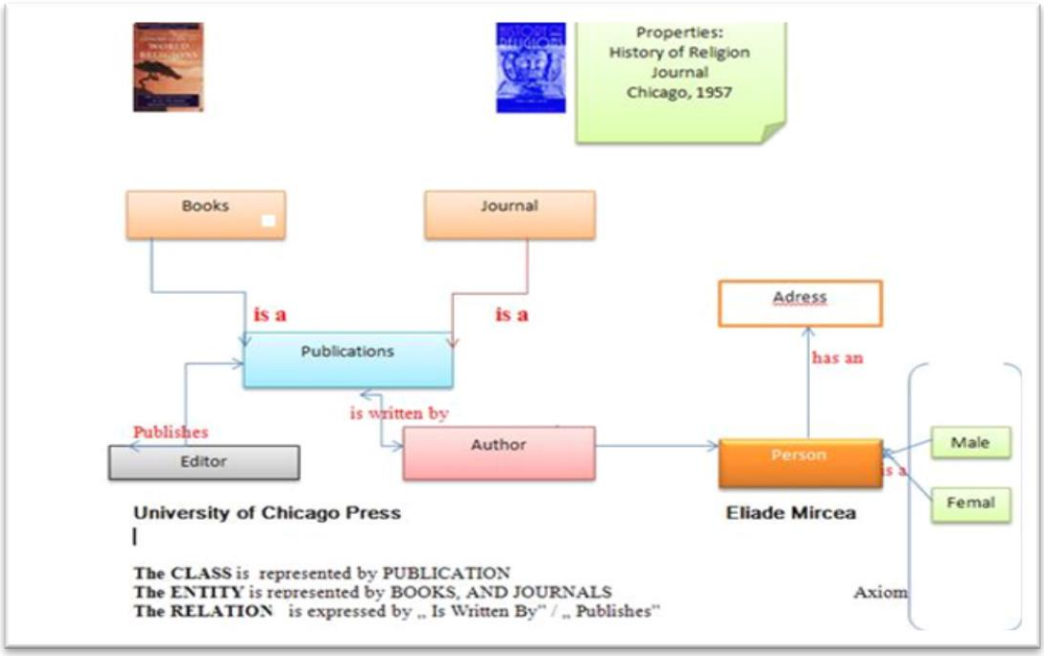

Fig. 2. **The representation of Class –Entity – Relation**

Using ontologies, we make semantic association between things we speak about. Web Ontology Language (2) is made of classes and the instances that shape its structure. It offers the possibility to describe via XML the relations created in order to be processed by the computer. Multimedia domain ontology may be built also using OWL and modulated with elements from MPEG 7 and MPEG 21.(3) The role of ontology from a pragmatic point of view is to provide controlled vocabularies, to create taxonomy for

navigation support, through a hierarchy of terms. This also may be useful in setting the users' expectation settings related to the content.

We consider that, a Digital Object is a representation of the traditional or digital created resource. The metadata are generated to describe the resource in order to help in identification, storage and distribution. The main types of metadata that specialists use belong to three classes: (a) Descriptive – helps in identification and indexing of the object. (b) Structural – offers information about the internal structure of the object. (c) Administrative – includes information about the rights management, hard, soft used for.

The background that World Wide Web Consortium provides for data processing is Resource Description Framework, which through XML language, HTML and URI allows the semantic and syntactic metadata representation [coding, exchange and processing]. RDF is a general language for semantic web representation of the real world. It allows exchange, of structured metadata. It uses XML syntax and the namespace concept. We use RDF to describe the web events, the proprieties of objects, the websites information, and the content of search engines.

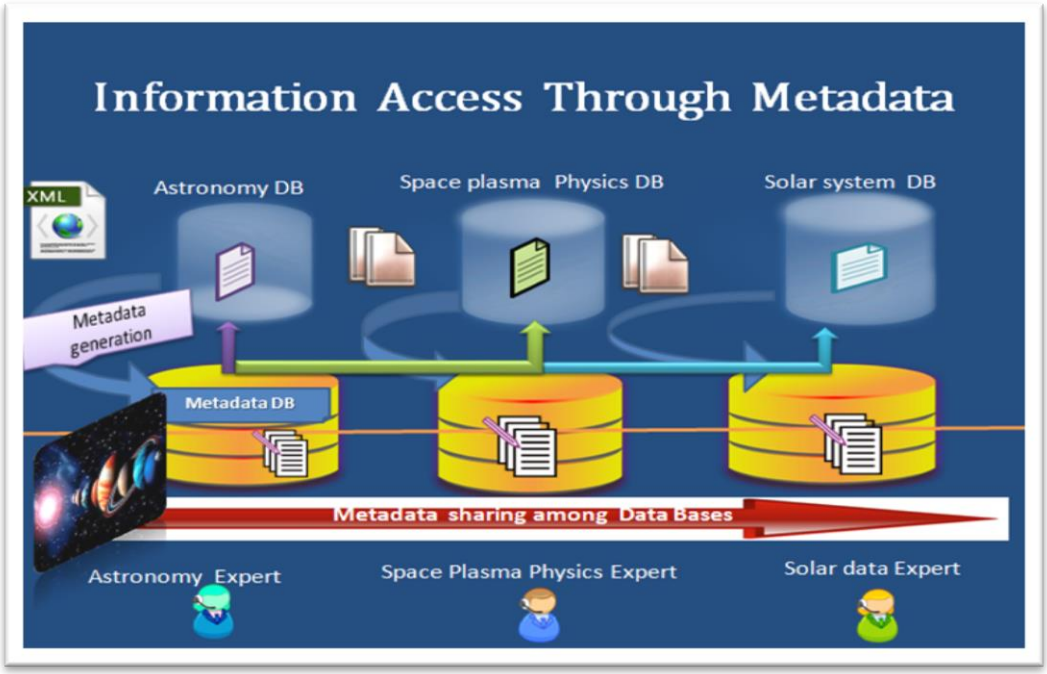

Fig. 3. Information access through metadata

3.2. Metadata standards for multimedia documents

Because a metadata standard is a reference tool with special attributes for the interoperability requirements, consequently, it involves a semantic understanding of media content and its cultural and legal rights. In this semantic environment the controlled vocabulary, the classification schema, and taxonomies play central role. Taxonomy organizes information, that metadata describes. In order to organize the

information, concepts need to be stored as *metadata*. Information is encoded in media content that contains an implicit semantics. In the multimedia environment there are various metadata standards; a strong multimedia schema must take into account some specifications: interoperability, extensibility, modularity and the media format. Once the sources of data collected, based on a strong metadata schema, the media content is easily searched, identified, and delivered to users. The design of a multimedia architecture involves *capturing, annotation, editing, publishing or sharing*. Multimedia material adds value by assigning specific metadata. The most relevant metadata standards, experts team intend to analyze, are: (a) MPEG – Moving Picture Expert Group, that developed standards for digital audio video coded content; (b) SMPTE – Society for Motion Picture and Television Engineering, that produces schema for motion imaging industry ;(c). DCMI-Dublin Core Metadata Initiative, that develops online metadata standards ;(d). IPTC – International Press Telecommunications Council, that maintains metadata for news exchange; (e). PRO-MPEG Forum, develops Tv equipment standards; Tv –Anytime Forum, that develops standards for high volume mass market storage in online platforms.

In the model presented below, we tried to design a model that reflects the *creation, representation, use* and *delivery* of media objects to the specific target – industries [Art & Business]. This strategy involves: (a) the subject domain; (b) media objects; (c) metadata & ontology; (d) representation and delivery standards. **MPEG -7** [Moving Picture Experts Group] is a standard for multimedia semantic description that allows combination of specific audio-visual materials. The audiovisual description as mentioned in ISO/IEC 15938-10 is presented below.

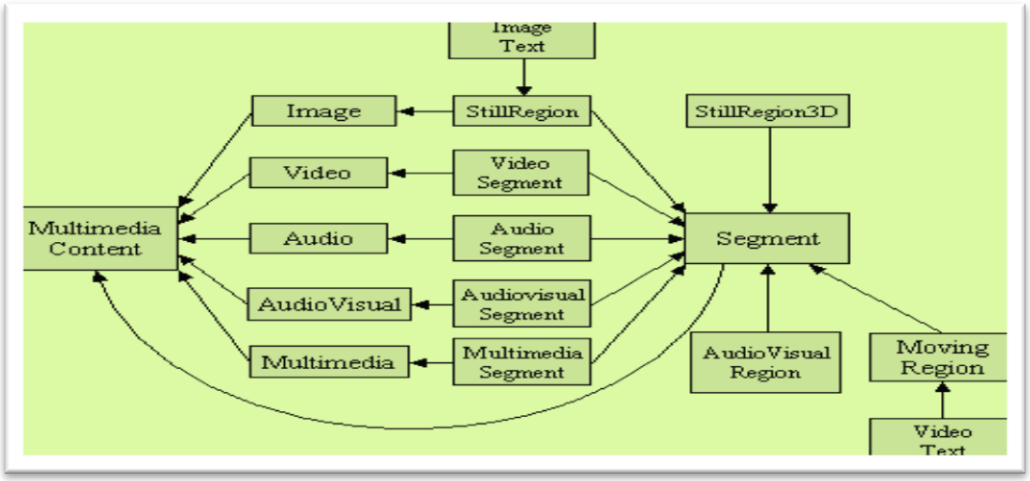

Fig. 4. **Multimedia class hierarchy in MPEG 7**

The MPEG 7 standard is based on descriptors for semantic and syntactic representation as follows: (a) information related to content creation; (b) information related to contents usage; (c) information related to storage; (d) information on temporal components; € information related to event; (f) information related to collection; (g) information related to users-content interaction. It provides instruments for multimedia schema, encoding specification. It comprises more than 450 metadata types using XML

as support language. MPEG 7[multimedia description] and MPEG 21[focus on exchange and distribution of digital items] metadata schema has an important role in music, TV, film and broadcast industries.

CIDOC Conceptual Reference Model (4) is an international standard for metadata formats from cultural domains that describes the core semantics, the essential rules of the documents. It covers 86 classes and 137 proprieties, offering the possibility to describe hundreds metadata formats. It also serves as data migration and exchange format on the semantic web. *Semantic interoperability* in media can be achieved with an "extensible ontology of relationships" and event modeling. The CRM provides a shared description rather than the prescription of a common data structure. CRM instances are encoded RDF XML language. Data falling within the scope of the CRM can be imported, exported, changed into a different system without losing its meaning.

Example: The content „ The Golden hand of the artist" by Zgondoiu Mihai – is stored in MySQL database and provided the bases for 3D content metadata creation. For this CIDOC CRM model was used to provide a mapping of it. D2R Server enhances publishing the MYSQL DB content into RDF for Semantic Web.

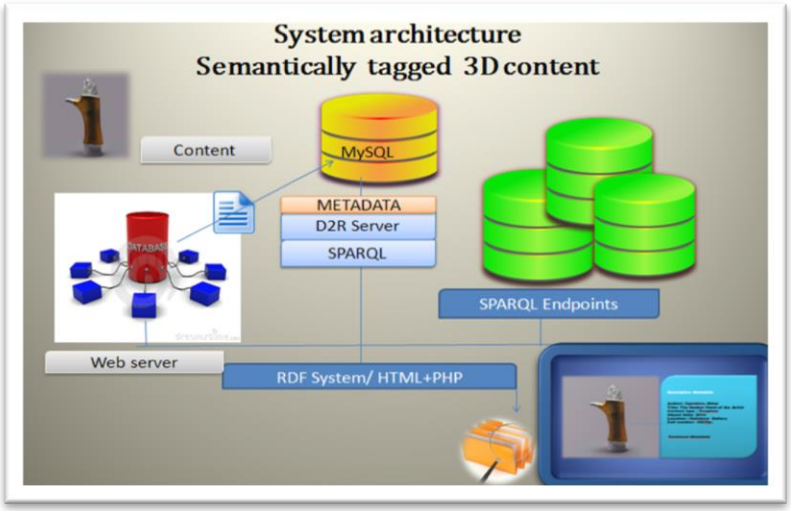

Fig. 5. **Mihai Zgondoiu. The Golden Hand of the Artist** (sculpture) – **3D content representation.**

3.3 Knowledge Representation in the Social Creative Media Community

The domain knowledge is structured on two types of ontologies *creative media content* and *creative media users*. In the multimedia environment the users interact communicate, collaborate and exchange contents. This behaviour offers the possibility to analyse the complex interractions between them. It also allows the creation of complex ontology in order to observe the user-media object relations. This kind of structure allows the users also to be guided through the best suited media objects for his search. The

Knowledge sistem related to users behavoiur is reflected in: 1. *Favourit* creative media objects; 2. *The tags* used for annotation; 3. *The description* made by users; 4. *The behavoiur* in a class/member group.

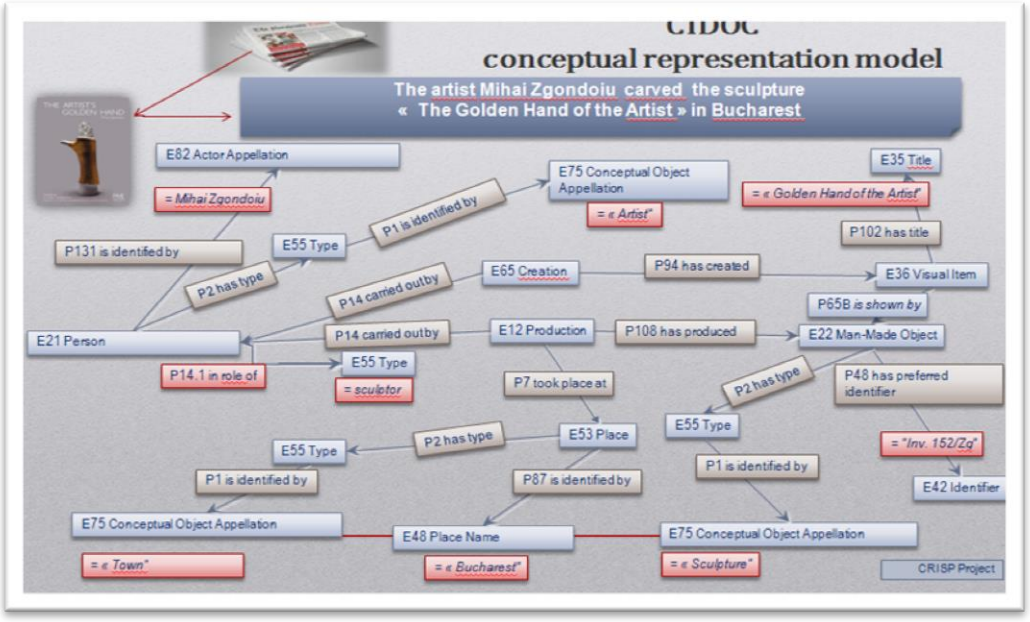

Fig. 6. **News representation** [The artist Mihai Zgondoiu carved the sculpture The Golden Hand of the Artist] **using CRM**

The metadata representation is the groundwork for an activity that includes the improvement of *markup representations* of information and its use in sharing and electronic publishing. In order to ensure the standardization of description a considerable amount of intellectual and practical work is mandatory.

References

(1) *Dublin Core Metadata Initiative.* [Online]. [Accessed at 12 March 2015]. Available at: http://dublincore.org/documents/dces/

(2) *OWL Web Ontology Language Overview.* [Online]. [Accessed at 12 March 2015]. Available at: https://www.w3.org/TR/owl-features/

(3) *ISO/IEC 21000-2:2003. Information technology -- Multimedia framework (MPEG-21) -- Part 2: Digital ItemDeclaration.* [Online]. [Accessed at 12 March 2015]. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=35366

(4) *CIDOC CRM Conceptual Model.* [Online]. [Accessed at 12 March 2015]. Available at: http://www.cidoc-crm.org/official_release_cidoc.html