Reprezentarea cunoaşterii de la schemele de clasificare la web-ul semantic (II)

Dr. Silvia-Adriana Tomescu
Biblioteca Centrală Universitară „Carol I”
Email: silvia.tomescu@bcub.ro

Dr. Silvia-Adriana Tomescu coordonează serviciul Cercetare. Metodologie din cadrul Bibliotecii Centrale Universitare „Carol I” din Bucureşti. Principalele sfere de cercetare ştiinţifică sunt: e-learning, biblioteconomie, redactare academică, diplomaţie culturală şi exil românesc. A publicat traduceri şi studii în domeniul comunicării şi al ştiinţei informării.

Rezumat

În acest eseu ne propunem să abordăm modalităţile de reprezentare a documentelor hibride în mediul online cu ajutorul standardelor specifice. Evoluţia conţinuturilor multimedia şi diversificarea suporturilor şi formatelor de stocare implică utilizarea unor instrumente elaborate de căutare, regăsire şi filtrare. Aplicaţiile dedicate au nevoie de un grad sporit de interoperabilitate ceea ce conduce spre nevoia acută de standardizare. Acest studiu încearcă să argumenteze teoretic nevoia de a încadra şi reflecta logic proprietăţile multimedia cu ajutorul standardelor de metadate şi să ilustreze importanţa taxonomiilor şi a ontologiilor în maniera de reflectare online a acestora.

Cuvinte cheie: *conţinut media, metadate, taxonomii, ontologii, obiecte digitale*

III. Reprezentarea cunoașterii prin conținuturi media

3.1. Instrumente de reprezentarea cunoașterii

Reprezentarea cunoașterii este procesul de ilustrare a conținuturilor cu ajutorul schemelor de clasificare, taxonomii, ontologii, structurate logic și semnatic pe web. Acest demers are la bază utilizarea limbajelor matematice și a formatelor de interschimb. Obiectele digitale sunt reprezentate prin arhitecturi de metadate, seturi de descriptori semantici grupați în jurul unui obiect pentru a-l explica, localiza, eficientiza în cataloage sau colecții. Orice tip de conținut media are nevoie de strategii de management al informației specifice pentru a optimiza procesul de creare, filtrare, și partajare, bazate pe metadate de conținut și context și standarde de interoperabilitate. Metadatele exprimă identitatea oricărui obiect digital, furnizând informații primare asupra conținutului.

BDD-A27025 © 2015 Editura Universităţii din Bucureşti
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

Exemplul din figura alăturată descrie un film în catalogul Europeana. Dublin Core Metadata Element Set (1) este cadrul general de descriere al aceatuia și cuprinde informații referitoare la: *titlu, creator, subiect, editor* etc. Nu este o schemă de metadate pentru materiale multimedia, însă are elemente comune cu acestea. Tipuri diferite de documente din baze de date sau cataloage cer standarde specifice de metadate.

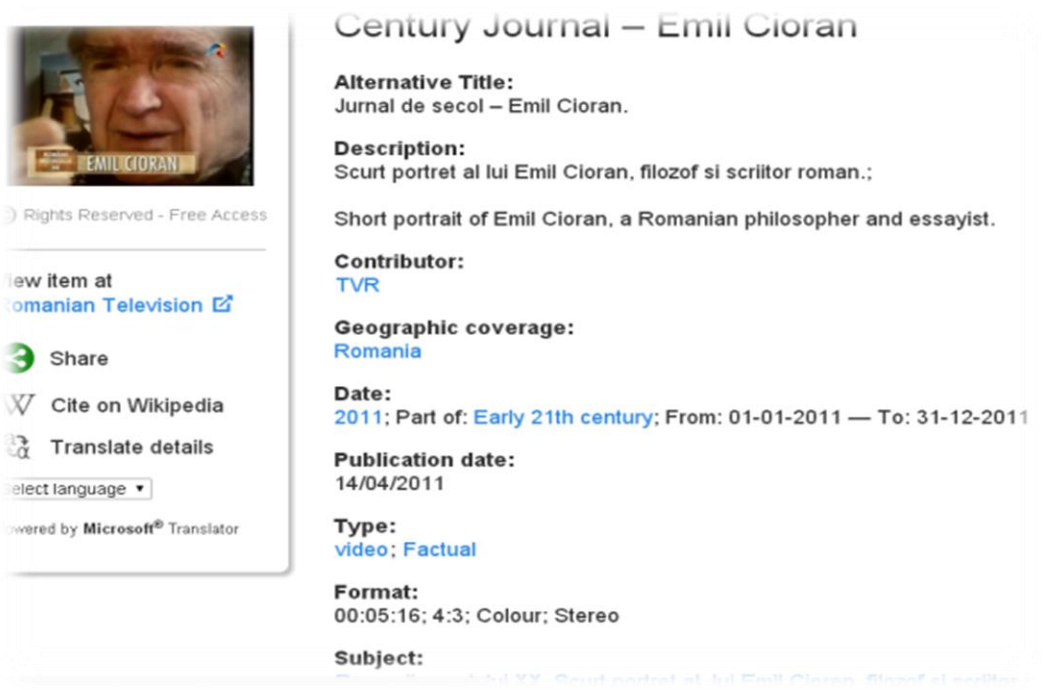


Fig. 1. **Metadate în Europeana** (Material audio-video: Jurnal de secol: Emil Cioran)

Elementele de bază ale acestei scheme se referă la: (a) **Conținut**: titlu, subiect, descriere, limbă, relație; (b) **Drepturi intelectuale**: creator, editor, contributor, drepturi; (c) **Instanță:** dată, timp, format, identificator. *Titlul* este reprezentat de numele dat resursei; *Creatorul* este orice entitate responsabilă de producerea resursei; *Subiectul*: este dat de conținutul/topical resursei; *Editorul* este entitatea responsabilă de partajarea resursei. Noțiunile și conceptele sunt structurate pe principiul dicționarelor și al glosarelor și sunt organizate în acord cu regulile limbajului natural pentru a eficientiza procesul de căutare și regăsire. În plus, le atribuie categorii semantice pentru a rafina căutarea.

Managementul informației într-un mediu multimedia trebuie să ia în considerare:

a) Utilizatorii, nevoile informaționale și abilitățile acestora;
b) Fluxul de lucru pentru managementul informației și al metadatelor;
c) Domeniile (multi)media;
d) Graficul încărcării obiectelor digitale;
e) Produsul final.

Instrumentele de reprezentare a cunoașterii în format digital sunt glosare – hărți conceptuale; tezaure – taxonomii; hărți de domeniu – ontologii; limbaje de reprezentare

BDD-A27025 © 2015 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

specifice. Ontologiile sunt conceptualizări multifațetate ale unui domeniu al cunoașterii definit de asocierea *entitate – relație* în procesarea automatizată. Există ontologii specifice domeniilor: astrologie, biologie, clasificându-se astfel drept ontologii speciale, sau generale. Entitatea modelează clase de obiecte; atributele modelează proprietățile. Membrii unei clase reprezintă instanțele acestei categorii. Ontologiile sunt exprimate prin limbaje de programare ca KIF [Knowledge Interchange Format] sau prin intermediul limbajelor bazate pe logica descriptivă [OWL; OIL].

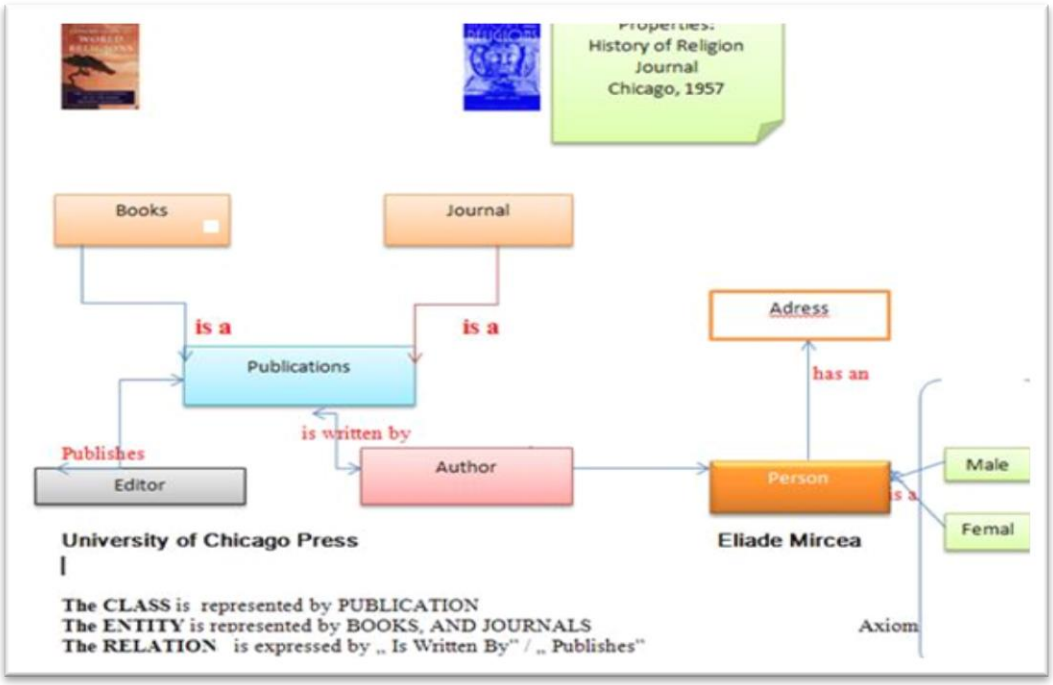


Fig. 2. **Modelul Clasă-Entitate-Relație**

Cu ajutorul ontologiilor realizăm asocieri semantice ale lucrurilor despre care vorbim; web ontologiile (2) sunt structurate în clase și instanțele care le definesc structura. Acestea oferă posiblitatea de a descrie relațiile cu ajutorul limbajului XML, pentru a fi procesate de computer. Ontologiile multimedia pot fi construite folosind OWL și modelate cu elemente din standardele MPEG 7 și MPEG 21.(3) Rolul ontologiei din punct de vedere pragmatic este acela de a furniza vocabulare controlate, a crea taxonomii pentru eficiență în navigare, prin ierarhii de termeni. Acest lucru poate fi util în analizarea așteptărilor utilizatorilor în legătură cu conținutul. Un obiect digital este o reprezentare a resursei tradiționale sau digitale create. Metadatele sunt generate pentru a descrie resursa pentru a ajuta la identificarea, stocarea și distribuirea acesteia. Principalele tipuri de metadate pe care specialiștii folosesc trei clase sunt: (a) descriptive – ajută la identificarea și indexarea obiectului; (b) structurale – oferă informații despre structura internă a obiectului; (c)administrative – includ informații despre gestionarea drepturilor hard, soft utilizate. Cadrul de lucru pe care consoţiul WWW îl agrează pentru procesarea datelor este RDF Resource Description Framework; acesta permite reprezentarea semantică şi

BDD-A27025 © 2015 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

sintactică a metdatelor, prin intermediul limbajelor XML, HTML şi URI Universal Resource Locator [codare, schimb şi procesare]. RDF este limbajul general pentru reprezentarea semantică a lumii reale în rețeaua WWW, permite schimbul metadatelor structurate, utilizează sintaxe XML şi formatul namespace. Cu ajutorul RDF descriem evenimente, informaţii, conţinutul pentru o bună reprezentare şi eficienţă în căutare.

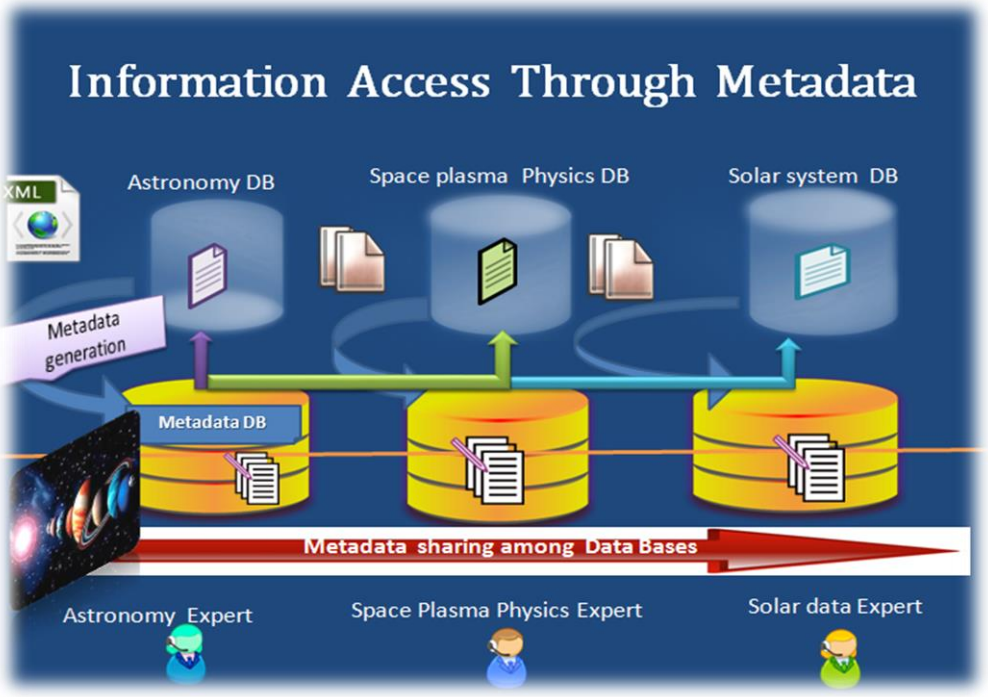


Fig. 3. **Accesul la informaţii cu ajutorul metadatelor**

3.2. Standarde de metadate pentru materiale multimedia

Deoarece un standard de metadate este un instrument de referinţe cu atribute specifice pentru interoperabilitate, presupune o înţelegere semantică a conţinuturilor media şi a drepturilor legale care decurg. În acest mediu semantic, vocabularele controlate , schemele de clasificare şi taxonomiile au un rol fundamental. Taxonomiile organizează informaţia descrisă de metadate. Informaţia este codificată sub conţinutul media, care poartă o semantică implicită. Există numeroase standard pentru sfera multimedia; o schema puternică de metadata specifică multimedia trebuie să ţină cont de: *interoperabilitate, extensibilitate, modularitate şi formatul media*. Odată structurate sursele cu ajutorul unei scheme solide, conţinutul media poate fi căutat, identificat şi furnizat către utilizatori. Designul arhitecturii multimedia implică *captură, adnotare, editare, publicare*, respective *partajare*. Materialele multimedia capătă valoare numai prin adaugarea unor metadate specifice. Cele mai importante standarde de metadate pe care experţii în domeniu le utilizează sunt: (a) MPEG – Moving Picture Expert Group,

BDD-A27025 © 2015 Editura Universităţii din Bucureşti
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

care a dezvoltat standardul pentru codarea conţinuturilor digitale audio-video; (b) SMTPE – Society for Motion Picture and Television Engineering, care elaborează schema pentru industria de filme de animaţie; (c) DCMI – Dublin Core Metadata initiative, care elaborează standardul pentru conţinuturi accesibile online; (d) IPTC – International Press Telecommunications Council, care întreţine metadatele de interschimb pentru ştiri; (e) Forumul PRO-MPEG care dezvoltă standarde pentru echipamentele TV; (f) Forumul TV Anytime care dezvoltă standarde de date stocate pe platforme online. În modelul mai jos, am încercat să prezentăm modalităţi de reprezentare, utilizare, partajare a obiectelor media spre un public specializat[Artă şi Industrie]. Startegia a implicat (a) subiectul; (b) obiectul media; (c) metadate şi ontologii; (d) reprezentare şi standardizare; MPEG 7 Moving Picture Expert Group este un standard de descriere semantică ce permite combinarea elementelor audio-video. Descrierea în manieră standardizată, conform ISO/IEC 15938-10, este prezentată mai jos:

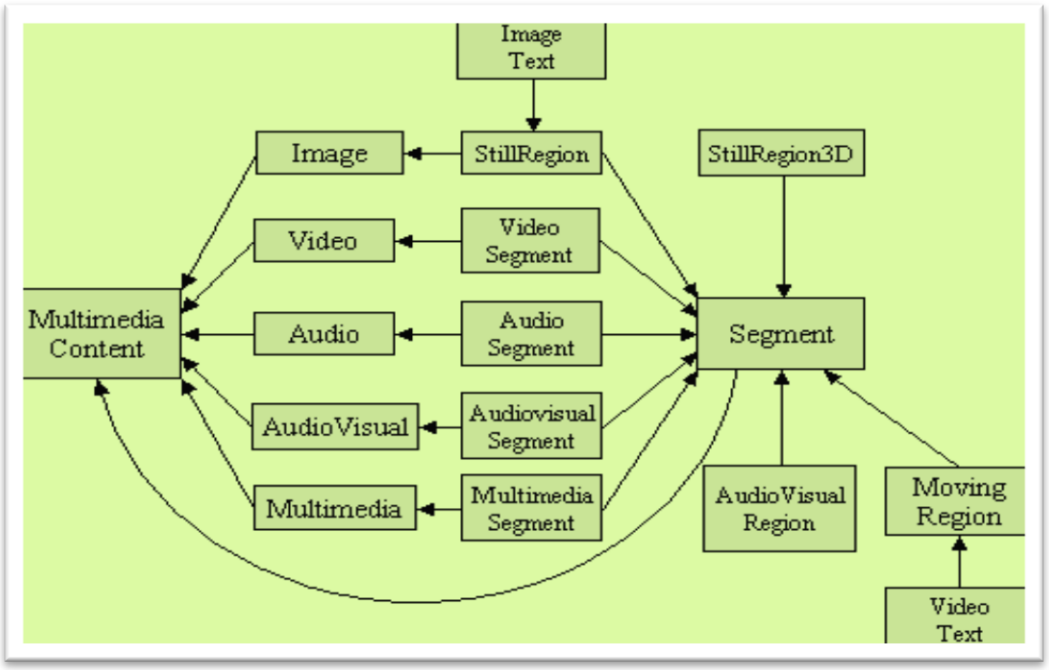


Fig. 4. **Structură ierarhică în MPEG 7**

Această schemă este bazată pe descriptori de reprezentare semantică şi sintactică: (a) informaţii referitoare la crearea conţinutului; (b) informaţii referitoare la utilizarea conţinutului; (c) informaţii referioare la stocare; (d) informaţii asupra coordonatelor temporale; (e) informaţii referioare la eveniment; (f) informaţii referitoare la colecţii; (g) informaţii referitoare la interacţiunea utilizator-conţinuturi. Standardul include specificităţi de codare şi peste 450 tipuri de metadate bazate pe XML ca limbaj de lucru. Cele două standarde MPEG 7 [descriere multimedia] şi MPEG 21 [schimb şi distribuţie de obiecte digitale] au o mare importanţă în televiziune, industriile media. CIDOC Conceptual reference Model (4) este un standard internațional de metadate cu

BDD-A27025 © 2015 Editura Universităţii din Bucureşti
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

aplicabilitate în sfera culturală, care descrie semantica şi regulile de bază ale documentelor. Acoperă 86 de clase şi 137 proprietăţi, oferind posibilitatea de a descrie sute de formate de metadate. Este folosit ca format de schimb şi migrare în webul semantic. Interoperabilitatea semantică în media poate fi obţinută cu ajutorul unei „ontologii de relaţionări extinse". CRM furnizează o descriere partajată şi mai puţin prescrierea unei structrui commune de date. În cadrul CRM instanţele sunt codificate cu ajutorul limbajelor RDF şi XML. Datele pot fi importate, exportate, schimbate fără a le altera semnificaţia.

Exemplificare: Conţinutul „Mâna de aur a artistului", autor Mihai Zgondoiu este stocată într-un server MySQL şi constituie bază pentru crearea unui conţinut 3D de metadate. Pentru mapare a fost utilizat modelul CIDOC-CRM. Prin serverul D2R se permite publicarea conţinutului MYSQL DB în RDF pentru a fi operabil în webul semantic.

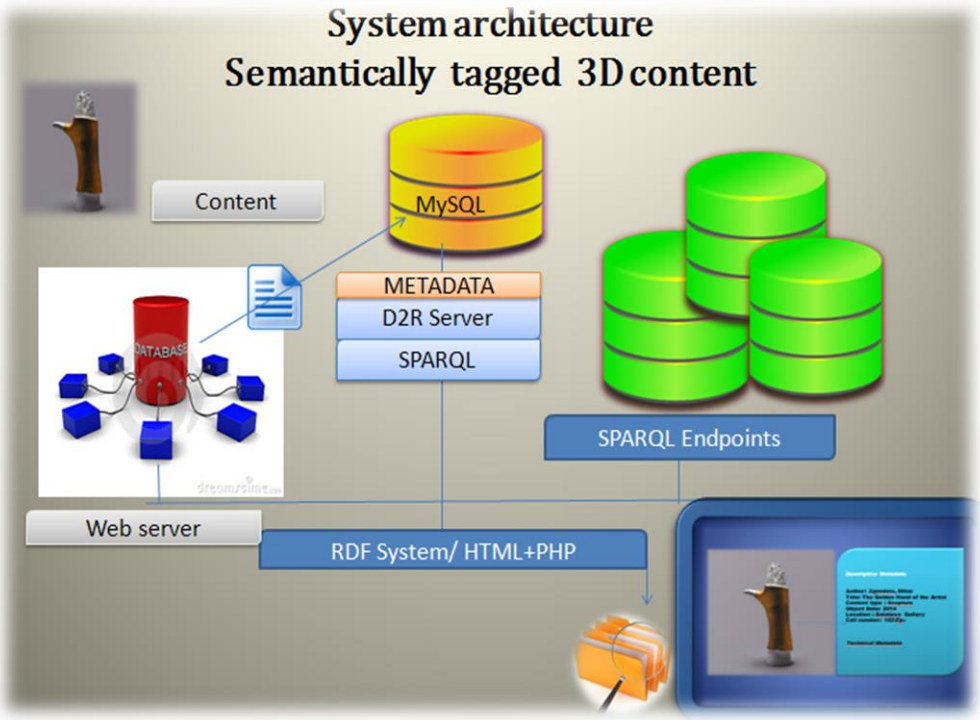


Fig. 5. **Mihai Zgondoiu. The Golden Hand of the Artist (sculptură) – Reprezentarea conţinutului 3D**

3.3. Reprezentarea cunoaşterii în mediul social creative media

Cunoaşterea domeniului este structurată pe două tipuri de ontologii: *conţinut de creative media* şi *utilizatori de creative media*. În spaţiul multimedia utilizatorii interacţionează, comunică, colaborează şi schimbă conţinuturi ceea ce oferă posibilitatea de a analiza interacţiunile complexe dintre ele. În egală măsură, permite ghidarea utilizatorilor pentru a obţine rezultatul cel mai adecvat în căutare. Reprezentarea

BDD-A27025 © 2015 Editura Universităţii din Bucureşti
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)

comportamentului utilizatorilor este legată de: (a) obiecte preferate; (b) etichetele folosite pentru adnotări; (c) descrierile realizate de utilizatori; (d) comportamentul grupului sau clasei de utilizatori.

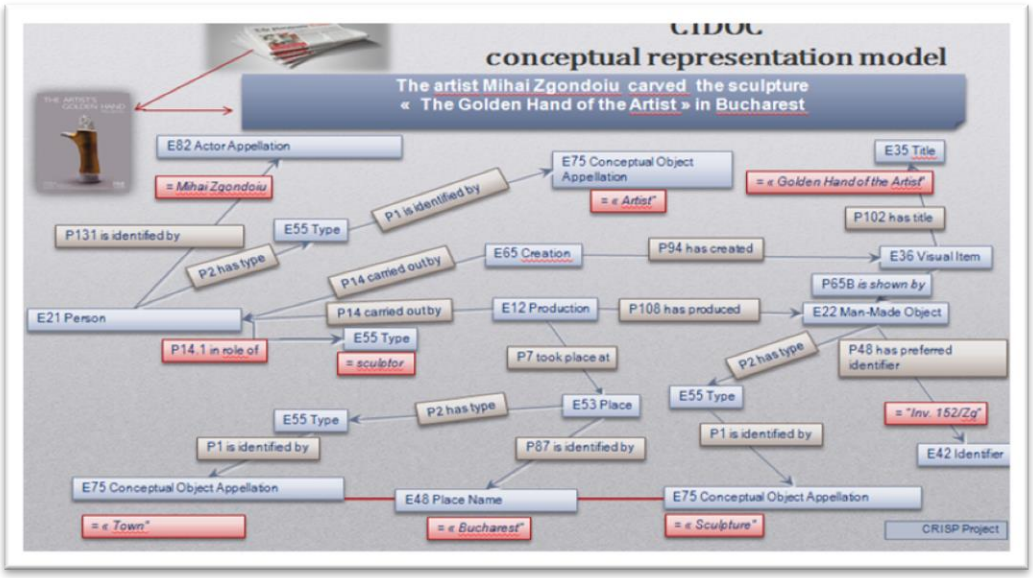


Fig. 6. **Reprezentarea unei ştiri de presă** [The artist Mihai Zgondoiu carved the sculpture The Golden Hand of the Artist] **utilizând CRM**

Structurarea metadatelor este fundamentală pentru orice activitate care presupune optimizarea reprezentării şi disponibilizarea acesteia prin circulaţie, respectiv prin publicarea conţinului creative media online. Pentru a asigura standardizarea descrierilor este nevoie de un efort intelectual și pragmatic deosebit.

Referințe bibliografice

(1) *Dublin Core Metadata Initiative*. [Online]. [Accesat la 12 martie 2015]. Disponibil la: http://dublincore.org/documents/dces/

(2) *OWL Web Ontology Language Overview.* [Online]. [Accesat la 12 martie 2015]. Disponibil la: https://www.w3.org/TR/owl-features/

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

(4) *CIDOC CRM Conceptual Model.* [Online]. [Accesat la 12 martie 2015]. Disponibil la: http://www.cidoc-crm.org/official_release_cidoc.html

BDD-A27025 © 2015 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 23:10:01 UTC)