

Forrás katalógus fejlesztése a NODES projektben

Herdon Miklós¹ – Várallyai László¹ – Szilágyi Róbert¹ – Lengyel Péter¹

¹Debreceni Egyetem, ATC, Agrárgazdasági és Vidékfejlesztési Kar, Gazdasági és Agrárinformatikai Tanszék, 4032 Debrecen, Böszörményi u. 138.

Absztrakt. A Dublin Core-nak az a szerepe, hogy párhuzamosan használják olyan metaadat-szabványokkal, amelyeknek másféle szemantika az alapja. Az egyszerűség egyrészt csökkenti a metaadatok elkészítésének költségeit és elősegíti a műveleti együttműködést. Másrészt viszont az egyszerűség nem fogadja be azt a szemantikus és funkcionális gazdagságot, amelyet az összetett metaadatformátumok támogatnak. A Dublin Core elemkészlet a részletező gazdagságot áldozza fel a mindenre kiterjedő hozzáférhetőségért. Az elemek leírásában minden egyes elemhez, jelentésének egyértelmű meghatározása érdekében, leíró megnevezés tartozik, továbbá egyedi, gép által értelmezhető egyetlen, egybeírt szóból álló név, amelynek célja, hogy az elemek kódolási sémák szerinti leírását egyszerűbbé tegye.

Kulcsszavak. Dublin Core, NODES, Forrás katalógus, elemkészlet.

Bevezetés

A Dublin Core Metaadat Kezdeményezés (The Dublin Core Metadata Initiative, DCMI) az Ohio állambeli Dublinban 1995-ben született könyvtárosok, digitális könyvtárkutatók, tartalomszolgáltatók és jelölőnyelv-szakértők közreműködésével. A szakterületeken túlnyúló keresési lehetőségek iránti érdeklődés azt eredményezte, hogy egyre többen vettek részt egy sor későbbi DCMI munkaértekezleten. A Dublin Core metaadat elemkészlete 15 alapelemből áll. A Dublin Core már több mint húsz nyelven áll rendelkezésre, elfogadta az Európai Szabványosítási Bizottság/Információs Társadalom Szabványosítási Rendszer (European Committee for Standardization/Information Society Standardization System, CEN/ISSS), továbbá fölvtették két Internet RFC dokumentumba (Requests for Comments, RFC). Hivatalosan elismerte a WWW Konzorcium (WWW Consortium) és az ISO 23950. A Dublin Core metaadatok amerikai nemzeti szabványként fogadták el (ANSI/NISO Z39.85), több mint 7 kormány hivatalosan jóváhagyta az elektronikus formában tárolt kormányzati információk keresésének segítésére, tovább á több nemzetek feletti szervezet alkalmazza, mint amilyen az Egészségügyi Világszervezet (World Health Organization, WHO). Könyvtárakban, levéltárakban, oktatási és kormányzati alkalmazásokban számos szakmai közösség specifikus metaadataihoz is felhasználják kiindulási alapként a Dublin Core-t. A Dublin Core-t nem arra szánták, hogy kizorítson bármely más metaadatszabványt. Inkább az a szerepe, hogy párhuzamosan használják, gyakran ugyanabban a forrásleírásban, olyan metaadat-szabványokkal, amelyeknek másféle szemantika az alapja. Az egyszerűség egyrészt csökkenti a metaadatok elkészítésének költségeit és elősegíti a műveleti együttműködést. Másrészt viszont az egyszerűség nem fogadja be azt a szemantikus és funkcionális gazdagságot, amelyet az összetett metaadatformátumok támogatnak. A Dublin Core elemkészlet valójában a részletező gazdagságot áldozza fel a mindenre kiterjedő hozzáférhetőségért.

Elemkészlet

Az elemek alábbi leírásában minden egyes elemhez, jelentésének egyértelmű meghatározása érdekében, leíró megnevezés tartozik, továbbá egyedi, gép által értelmezhető egyetlen, egybeírt szóból álló név, amelynek célja, hogy az elemek kódolási sémák szerinti leírását egyszerűbbé tegye.

Bár bizonyos környezetek, mint amilyen a HTML, érzéketlenek a kis- és nagybetűkre, legjobb megoldásként ajánlatos következetesen ragaszkodni a következőkben megadott elemnevek írásmódjához, hogy elkerüljék a konfliktusokat arra az esetre, amikor metaadatok később kivonatolnak vagy olyan, kis- és nagybetűkre érzékeny környezetbe

konvertálnak, mint az XML (Extensible Markup Language, kiterjeszhető jelölőnyelv).

Valamennyi elem szabadon választható és ismételhető. A metaadatelemek tetszőleges sorrendben tüntethetők fel. Ugyanazon elemnek (pl. létrehozó) többszöri előfordulása esetén a szolgáltató fontossági sorrendet határozhat meg, de ennek a sorrendnek a betartását nem minden rendszer szavatolja. A világméretű együttműködés elősegítése érdekében számos elem értékének megadására szabályozott szótár használata ajánlatos. Feltételezhető, hogy egyéb szabályozott szótárakat is létrehoznak az együttműködés érdekében meghatározott helyi alkalmazási területeken.

1. táblázat Elemek

Elemnév	Megnevezés	Tartalom	Magyarázat
Title	cím	az információforrásnak adott név.	Az a megnevezés, amelyen az információforrás általában ismert.
Creator	létrehozó	az információforrás tartalmának létrehozásáért elsősorban felelős entitás.	A létrehozó lehet személy, szervezet vagy szolgáltató.
Subject	tárgy- és kulcsszavak, jelzetek	az információforrás tárgyának megadása.	Szabályozott tárgyszójegyzékek elemeinek vagy rendszerbe foglalt osztályozási jelzetek használata ajánlatos.
Description	leírás	az információforrás tartalmának ismertetése.	Például: a referátum, a tartalomjegyzék, a hivatkozás a tartalom grafikus megjelenítésére vagy a tartalom szabadon megfogalmazott leírása.
Publisher	kiadó	az információforrás nyilvánosságához közvetítéséért felelős entitás	Például: kiadó: személy, testület vagy szolgáltató.
Contributor	közreműködő	az információforrás tartalmához készült hozzájárulás létrehozásáért felelős entitás.	Például: személy, testület vagy szolgáltató.
Date	dátum	az információforrás létezése során előforduló esemény időpontja (dátuma).	Az információforrás létrehozásával vagy hozzáférhetővé tételével kapcsolatos dátum. ISO 8601 szerinti kódolt formátum. Amely (többek között) az ÉÉÉÉ-HH-NN (YYYY-MM-DD) formában tartalmazza a dátumot.
Type	típus	az információforrás tartalmának jellege, vagy fajtája.	Az információforrás típusára vonatkozó, általános kategóriákat, funkciókat, fajtákat vagy a tartalom összetettségének szintjeit leíró kifejezések.
Format	formátum	az információforrás fizikai vagy digitális megjelenési formája.	Az információforrás hordozójának típusát vagy terjedelmét (méret és időtartam) tartalmazza. A formátum a szoftver, a hardver vagy más, az információforrás megjelenítéséhez, vagy működtetéséhez szükséges eszközök megnevezésére is használható.
Identifier	forrásazonosító	az információforrásra való, adott környezetben belüli egyértelmű hivatkozás.	Ajánlatos az információforrás azonosítására szolgáló, szabványos azonosító rendszernek megfelelő jelsorozat használata. Például: URI (Uniform Resource Identifier) URL (Uniform Resource Locator), digitálisobjektum-azonosító DOI (Digital Object Identifier) és a nemzetközi szabványos könyvazonosító szám ISBN (International Standard Book

			Number.
Source	eredeti információforrás	hivatkozás arra az eredeti információforrásra, amelyből a jelen információforrás származik.	Az adott információforrás részben vagy egészben eredeti forrásanyagból származtatható.
Language	nyelv	az információforrás intellektuális tartalmának nyelve.	Ajánlatos az RFC 3066 használata, amely az ISO 639 szabvánnyal együtt tartalmazza a két- vagy hárombetűs nyelvkódokat és a tetszőlegesen alkalmazható alkódokat. Például: .en. vagy .eng. használata az angol nyelvre, .akk. az akkádra, és en-GB. az Egyesült Királyságban használt angolra.
Relation	kapcsolat	hivatkozás az információforrással kapcsolatban lévő másik információforrásra.	Ajánlatos a hivatkozott információforrás azonosítására hivatalos azonosító rendszer megfelelő jelsorozatának/értékének használata.
Coverage	tér-idő vonatkozás	az információforrás tartalma vagy alkalmazási területe térben vagy időben (kiterjedés).	A kiterjedés térbeli hely (helynév vagy földrajzi koordináták), idő (idő megnevezése, dátum vagy időintervallum) vagy hatáskör Ajánlatos szabályozott szótárban (például földrajzi tezaurusban TGN (Thesaurus of Geographic Names) található kifejezést választani, a számszerű értékek, koordináták vagy időintervallumok helyett megnevezett helyeket és időszakokat használni.
Rights	jogok	információk az információforrással kapcsolatos jogokról.	A jogok az információforrások jogkezelésére vonatkozó adatokat vagy az olyan szolgáltatásra vonatkozó hivatkozást tartalmaznak, amelyek erre vonatkozó információkat szolgáltatnak. Ezen információk gyakran magukban foglalják a szellemi tulajdon védelmére vonatkozó jogokat IPR (Intellectual Property Rights), a szerzői vagy kiadói jogfenntartásokat (copyright) és a különböző vagyoni jogokat. Ha ez az elem hiányzik, nem feltételezhető, hogy bármely jog kapcsolódik az információforráshoz.

A fenti táblázat alapján elkészítettük azt a weblapot, amelyen keresztül lehetőség van az általunk összegyűjtött adatok felvitelére. Valamennyi projekt résztvevő számára nyitott ez az oldal, ahol angol nyelven nyílik lehetőség a kívánt adatok felvitelére. Az űrlap URL címe a következő: <http://odin.agr.unideb.hu/nodes/content/dc-list.php>

A következő ábra a felviteli űrlapot mutatja be:

Content	
Title	<input type="text"/>
Creator	<input type="text"/>
Subject	HOW TO USE A COMPUTER
Description	<input type="text"/> HOW TO USE A COMPUTER WHAT IS INTERNET USING AN ELECTRONIC MAIL SYSTEM TO COMMUNICATE CREATING - DESIGNING AND MANAGING A DOCUMENT CREATING - DESIGNING AND MANAGING A DOCUMENT A WEB DOCUMENT MANAGEMENT OF INFORMATION *COMMUNICATE* WITH OFF-LINE and ON-LINE MULTIMEDIA
Publisher	INTERNET BASED SERVICES FOR CITIZENS / FOR ALL (ADVANCED)
Contributor	LAW, ETHICS AND CITIZENSHIP
Date	AG / FARMING PROFESSIONAL USAGES OF COMPUTER AND INTERNET
Type	DESIGNING AND MANAGING *ADVANCED* WEB DOCUMENTS THE CONCEPTS OF E-COMMERCE AND E-ADMINISTRATION
Format	<input type="text"/>
Identifier	<input type="text"/>
Source	<input type="text"/>
Language	english
Relation	<input type="text"/>
Coverage	<input type="text"/>
Rights	<input type="text"/>
<input type="button" value="Save"/>	

1. ábra.
Beviteli űrlap

A "Subject" részben a projekt résztvevők által korábban elfogadott témakörök szerepelnek, melyet legördülő listából választhatunk ki.

Ugyancsak legördülő menüből választhatjuk ki a nyelvet, melyek közül a projekt nyelvként is használt "angol" az alapértelmezett.

Természetesen az adatbázisban azt is rögzíteni kívánjuk, hogy mely személy és melyik országból töltötte fel a hivatkozást és lehetőséget kívánunk biztosítani arra is, hogy a feltöltött anyagot, a projekt résztvevői nyelvére le lehessen fordítani. Ezeket természetesen szintén az adatbázisban tároljuk, hogy a későbbi lekérdezések során az adott nyelvi alapú lekérdezésre is legyen lehetősége a megfelelő jogosultsággal rendelkező felhasználónak.

Ugyanezen a lapon két linket találhatunk, az egyik angol nyelvű és a Dublin Core hivatalos szabványáról és elemeiről szól, a másik magyar nyelvű, a szabvány hivatalos fordítása.

A felvitt adatokat lehetőségünk van megtekinteni és ugyanezen a lapon a "New content" linkre kattintva az 1. ábrán látható adatfelviteli lapra jutunk vissza.

A fejlesztés jelenlegi fázisában kerül kidolgozásra a lekérdező űrlap, ahol megfelelő jogosultsági szintek alapján fogalmazhatjuk az adatbázissal kapcsolatos lekérdezéseinket felhasználóbarát módon.

Felhasználók

MEK Dublin Core generátor

A Magyar Elektronikus Könyvtár (MEK) szolgáltatásaként lehetőség van Dublin Core metaadatok készítésére (<http://mek.oszk.hu/dc/>). A kapott adatok bemásolhatók a HTML fájlok fejlécébe, vagy elmenthető XML formátumban a dokumentum mellé. A DC metaadatokat egyre több keresőrendszerben használják, Magyarországon például a Nemzeti Digitális Adattárban és a <http://nda.sztaki.hu/keresoben>. A generátort a Magyar Elektronikus Könyvtárért Egyesület készítette és egyelőre csak a Dublin Core alapját jelentő 15-féle adatelem leírását teszi lehetővé. Ajánlott minél több mezőt kitölteni, de csak a cím és az

azonosítóaz, ami lényegében kötelező. Az adatelemeket minősítő opciókat nem kötelező használni, de ahol van értelmük, ott a pontosítás kedvéért ajánlott valamelyiket beállítani.
<http://mek.oszk.hu/dc/szabvany/134715.pdf> A szabvány hivatalos magyar fordítása

Nemzeti Digitális Adattár <http://www.nda.hu>

A Nemzeti Digitális Adattár (NDA) a tudásalapú társadalom, a digitális kor archívuma, mely kulturális örökségünk digitalizált értékeihez biztosít széles körű hozzáférést. Az NDA-t az Informatikai és Hírközlési Minisztérium hozta létre 2003-ban annak érdekében, hogy a nemzeti digitális adatvagyonhoz való hozzáférés biztosításával támogassa a nemzeti kultúra digitális reprezentációját, az információs társadalom fejlődését.

Az NDA az Interneten elérhető magyar nyelvű és magyar vonatkozású kulturális tartalmakat (pl. virtuális könyvtárak, múzeumok, archívumok stb. digitális dokumentumait) katalogizálja, s egy közös keresőfelületen közvetlenül elérhetővé teszi azokat.

A Nemzeti Digitális Adattár tehát:

- 1. az Interneten elérhető elektronikus tartalmak elektronikus katalógusa
- 2. pontosabban: az Interneten elérhető magyar nyelvű és magyar vonatkozású kulturális tartalmak elektronikus katalógusa, azaz a digitális dokumentumokról készült leíró adatok (katalóguscédulák, az ún. meta-adatok) gyűjteménye

Az NDA-keresővel jelenleg 43 partnerintézmény 56 különböző adatbázisa, összesen közel 200.000 dokumentuma érhető el. Az NDA-ban elérhető kulturális online tartalmak az alábbi tematikus csoportokba rendezhetők:

- virtuális könyvtárak (szépirodalom, szakirodalom, elektronikus folyóiratok)
- virtuális múzeumok, gyűjtemények (képzőművészeti és egyéb közgyűjtemények)
- adattárak (társadalomtudományi, történelmi, filmművészeti, zenei, közművelődési)
- műsorarchívumok: (TV- és rádióarchívumok, koncertfelvételek)
- Az NDA-hoz önkéntesen, tartalmi megkötés nélkül csatlakozhatnak a közgyűjtemények és egyéb szervezetek, intézmények, vagyis az adatgazdák.
- Az NDA-hoz csatlakozó adatgazdák az általuk szolgáltatott digitális tartalomról az adott dokumentumtípusnak (szöveg, kép, hang, videó) megfelelő leírást (katalógus-cédulát) készítenek. Az NDA ezeket a leíró adatokat gyűjti össze és tárolja az NDA adatbázisában. A felhasználók az NDA-kereső segítségével ebben az adatbázisban kereshetnek, s a találati listában megjelenő linkekre kattintva közvetlenül elérhetik az adatgazdák által szolgáltatott tartalmat.
- Az elektronikus tartalmat maguk az adatgazdák szolgáltatják, a tartalom frissítését, bővítését az NDA folyamatosan és automatikusan követi. Az adattartalmat és adatminőséget tehát az adatgazdák biztosítják, ezekért felelősséget az NDA nem vállal.
- Az NDA-hoz az adatgazdák folyamatosan csatlakozhatnak, így az elérhető adatok mennyisége folyamatosan bővül.

Nemzeti Audiovizuális Archívum <http://nava.hu/index.php>

A NAVA - teljes nevén Nemzeti Audiovizuális Archívum - a magyar nemzeti műsorszolgáltatói köteles példány archívum, amely jellegénél fogva audiovizuális tartalmakat gyűjt. Gyűjtőkörébe tartoznak az országos földfelszíni terjesztésű televíziók és rádiók magyar gyártású vagy magyar vonatkozású műsorai. A NAVA ezenkívül befogad helyi műsorszolgáltatói vagy bármilyen egyéb audiovizuális tartalmú archívumokat feldolgozási vagy tárolási célból, ezzel is segítve az audiovizuális tartalmak, mint a kulturális örökség részeinek megőrzését.

A NAVA olyan az elektronikus műsorok számára, mint az Országos Széchényi Könyvtár a nyomtatott kiadványok vagy a Magyar Nemzeti Filmarchívum a magyar filmek számára. Az archívum 2006. január 1-én kezdte meg munkáját.

A NAVA működését a **2004. évi CXXXVII. törvény** szabályozza, a szerzői jogokhoz kapcsolódó feladatait a **117/2004. (IV. 28.) Kormány rendelet**, az audiovizuális műsorszámok felújításának, valamint szolgáltatásának műszaki, minőségi és egyéb követelményeit pedig a **52/2007. (V.17.) GKM-OKM** együttes rendelete határozza meg.

A NAVA különlegessége, hogy gyűjteményéhez on-line hozzáférést biztosít a törvény szabta kereteken belül - archívumának adatbázisa szabadon kereshető, a benne található műsorok teljes terjedelmükben pedig az úgynevezett **NAVA-pontokon** (amelyek könyvtárakban, iskolákban stb. elérhető terminálokat jelent) tekinthetők meg.

A NAVA szolgáltatásainak igénybevétele ingyenes.

Koncepció

A NAVA a magyarországi műsorszolgáltatók által sugárzott műsorszámokat:

- rendszeresen és üzemszerűen gyűjti és tárolja,
- úgynevezett tartalmi leíróadatokkal (meta-adatokkal) ellátja (feldolgozza),
- a nyilvánosság számára oktatási és kutatási célra hozzáférhetővé teszi.

A NAVA ezt a feladatát digitális, informatikai környezetben látja el, minek következtében a rögzített műsorszámok

- digitális feldolgozásra kerülnek,
- digitálisan tárolódnak,
- hozzáférhetővé tételük on-line digitális formában történik.

A NAVA működése során létrejövő gyűjtemény közgyűjtemény és állami tulajdont képez. A gyűjteményre - ha ezt más jogszabály másképp nem szabályozza - a könyvtári közgyűjteményekre vonatkozó jogszabályokat kell alkalmazni.

A NAVA szakmai koncepciója több induló archívum - többek között a görögországi Hellenic National Audiovisual Archive - számára is kiindulópont volt.

Műszaki rendszer

Az archívum katalógusa az Internet felhasználók számára korlátozás nélkül elérhető, a tartalom NAVA végpontok zárt hálózatán keresztül tekinthető meg. A médiafolyamok tekintetében a végpontokon rendelkezésre álló műszaki feltételeket, valamint a NAVA anyagi lehetőségeit figyelembe véve a végpontok számát az első fázisban korlátozzuk. A zárt hálózat (úgynevezett NAVA-pontok) tagjai:

- nyilvános szolgáltatásokat nyújtó könyvtárak,
- iskolai oktatás célját szolgáló intézmények,
- muzeális intézmények,
- levéltárak,
- kép- és hangarchívumok.

A NAVA teljes egészében digitális rendszer, a média tárolása is adatformátumban történik. A formátumok kiválasztásánál alapelvek, hogy az elérhető minőség kövesse a mindenkori műsorszórás által biztosított sugárzási minőséget, illetve tegye lehetővé az internet szolgáltatók által széles körben biztosított sebességen történő szolgáltatást. Nem elsődleges cél az újrafeldolgozható minőségben történő tárolás.

A NAVA tárolási minősége nyílt szabványú:

- MPEG-2, MP@ML (720x576, 25 fps), CBR, 8Mbps.

A NAVA szolgáltatási minősége (ingyenes lejátszó):

- RealMedia, 384x288, 25fps, 512 kbps, RV10 + RA8.

A NAVA szolgáltató felülete - ezért célszerűen belső felépítése is - teljes egészében követi a nemzetközi (lehetőség szerint nyílt) szabványokat, ajánlásokat, kezdeményezéseket, így lehetővé teszi a más archívumokhoz való egyszerű kapcsolódást. A legfontosabb kapcsolódási pontok:

- Motion Pictures Expert Group (MPEG),
- Dublin Core Metadata Initiative (DCMI),
- European Broadcasting Union (EBU),
- Open Archives Initiative (OAI).

Irodalom

Dublin Core <http://dublincore.org/>

Magyar Elektronikus Könyvtár Dublin Core metaadat készítő <http://mek.oszk.hu/dc/>

Nemzeti Digitális Adattár <http://nda.sztaki.hu/>