

# A Neumann-galaxis

## Az elektronikus kor és a számítógépes adatformátumok

SALGÁNÉ MEDVECZKI Marianna

A Debreceni Egyetem Informatikai Intézete által rendezett 2. Gyires Béla Informatikai Napon 2004. május 14-én elhangzott előadás átdolgozott változata.

### Bevezetés

Azért választottam a fenti címet, mert úgy érzem, jól kifejezi azt a kettősséget, amellyel az elektronikus kor mostani szakaszában, az új kommunikációs formák megjelenésével a Gutenberg-galaxist őrző, de a Neumann-galaxis kihívásaira reagáló könyvtáraknak meg kell birkózniuk.

Az információ- és könyvtártudomány, vagy mondhatjuk könyvtár-informatikának is, a nyugat-európai országokban, valamint a tengerentúlon évtizedes hagyománnyal rendelkező, tudományos eredményeket felmutató szakterületnek számít. Interdiszciplináris kapcsolatban áll más tudományterü-

letekkel is (pl. az informatikával, kommunikációtudománnyal stb.), bár vannak, akik e „multi-szerepét” kétségbe vonják.

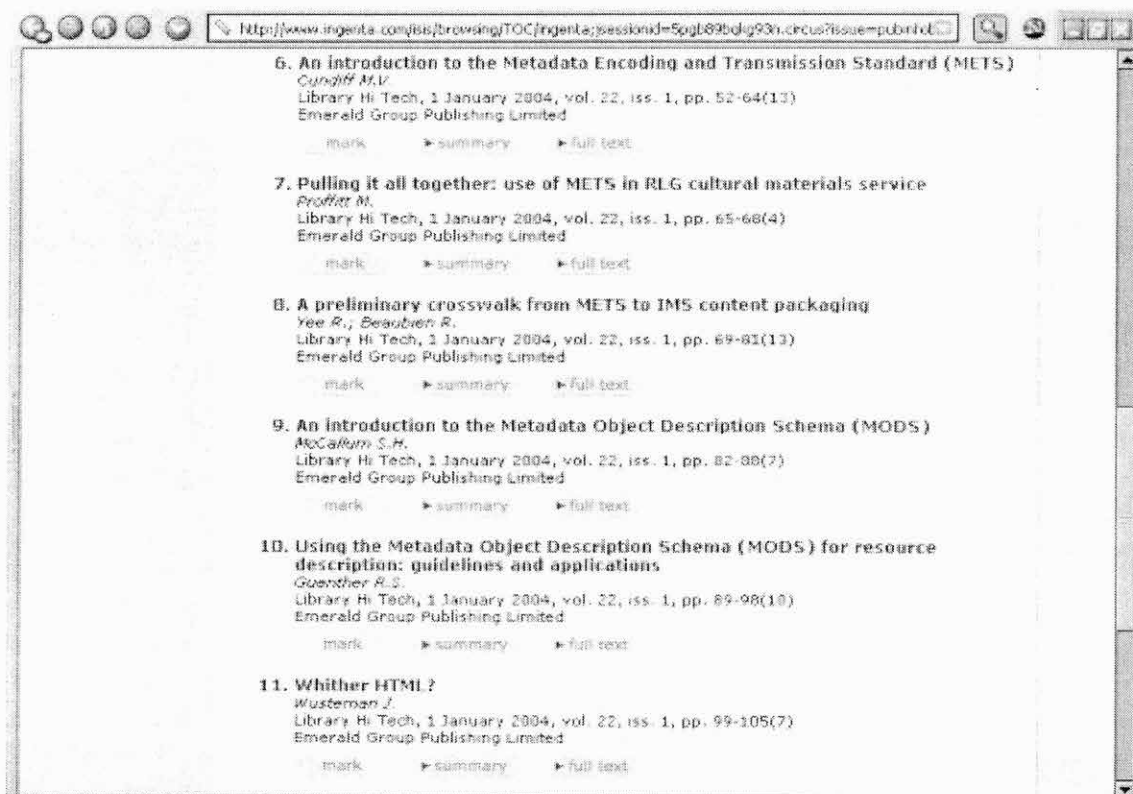
Nálunk alig múlt tíz-tizenöt éve, hogy a nagyobb magyar könyvtárak megvásárolták első integrált számítógépes rendszerüket, s lassan valamennyi könyvtártípusban megkezdődött a munkafolyamatok gépesítése. Szerencsére, a nyugati gyakorlathoz képest fennálló, mintegy 30 éves, kezdeti lemaradásunk, mára már jelentősen csökkent. Remélem, hogy ez a lendület információs társadalommá válásunk küszöbén sem törik meg és „ügyünk” pártfogásra talál minden fórumon. Bízom abban is, hogy

a magyar tudományos élet is be-, ill. elfogadja majd a könyvtár-informatika iránt érdeklődő szakembereket.

A könyvtárak erősödő informatikai irányultságának igazolására elég, ha egy pillantást vetünk, mondjuk az egyik rangos amerikai könyvtár-informatikai folyóirat, a *Library Hi-Tech* tetszőleges számának tartalomjegyzékére. Elég átfutni a tanulmányok címét és már nem is kell mérlegelni: létezik-e a könyvtár-informatika vagy sem?

## A könyvtárak jövője

A könyvtárak és a könyvtárosság jövőjéről – érthető okokból – igen sok tanulmány jelent meg az utóbbi időben. Az egyik közismert alapműben *Michael Buckland* – magyar fordításban is megjelent könyvében (A könyvtári szolgáltatások újratervezése, 1998) – a könyvtárak fejlődésének három szakaszát írta le: az első szakaszban a könyvtári dokumentumok papír alapúak, a könyvtári segédeszközök, munkafolyamatok is papírhoz kötöttek. A fejlődés következő fokozatát automatizált könyvtárnak nevezte. Ebben a könyvtárban a dokumentumok még döntően papír alapúak, de a könyvtári



1. ábra.

A *Library Hi Tech* 2004-es 1. számának tartalomjegyzéke

munkafolyamatok már automatizáltak. A harmadik szakasz könyvtárát elektronikus könyvtárnak nevezte el. Ennek jellemzője, hogy a dokumentumokat elektronikus formában tárolják, tehát nem papíron vagy egyéb, helyhez kötött médiumon és a munkafolyamatok és a segédeszközök is automatizáltak. Buckland megállapításait a könyvtári szakemberek ma is helytállóknak ítélik.

A papír nélküli könyvtárról szóló negatív jóslatok eddig nem váltak be, és nem valószínű, hogy a jövő könyvtáraiból hiányozni fognak a papír alapú dokumentumok. Ebből következően állítható, hogy nem az elektronikus könyvtárak megteremtése a cél, hanem az olyan integrált könyvtáré, amelyben egyaránt megtalálhatók a hagyományos publikációk és az elektronikus dokumentumok, amelyben a felhasználó hozzájut a fizikai valójában jelenlévő állományhoz éppúgy, mint – a hálózat révén – a világ bármely sarkában vagy éppen hálózaton található elektronikus és nem-elektronikus forrásokhoz. Vagyis, az igazi kihívás – az integrálás.

Az integrált könyvtár fogalmát *Poprády Géza* a könyvtári trendekről készült tanulmányában (Könyvtárosok kézikönyve. 5. 2003.) így határozza meg: [Az integrált könyvtár] „... olyan osztott információs rendszer, amely mind a hagyományos, mind a rendkívül változatos elektronikus dokumentumokat tekintve biztonságos tárolást, a felhasználóknak pedig kényelmes és hatékony hozzáférést biztosít a legfejlettebb digitális technológiákkal és adatátviteli hálózatokkal támogatott könyvtári munkafolyamatok révén.”

## Az információ azonosítása

*Burkart, M.* szerint a könyvtárnak mint a társadalom egyik információs alrendszerének elsődleges feladata az információcsere, információáramlás biztosítása azok között a partnerek között, akik ugyan-

azon téma iránt érdeklődnek, de nem tudnak közvetlen kommunikációs kapcsolatba lépni egymással. (Burkart, 1990).

A könyvtár feladata, hogy a társadalom információkincsét, azaz a közös tudásvagyont megőrizze és gondoskodjon arról, hogy ez a vagyon bekerüljön a kommunikáció csatornájába.

A könyvtárak inputját tehát a dokumentumokban (információforrásokban) megjelenő információ/tudás képezi. Bármilyen eljárást, technológiát alkalmaznak is a könyvtárak a feltárási folyamatok során, első lépésként az információnak azt a mennyiségét/részét kell meghatározni, kiválasztani, amely a folyamatok további szakaszaiban egy egységként jelenik meg.

Az információtudomány szűken értelmezett feladata nem más, mint hogy létrehozza az eredeti szövegek reprezentációját, szurrogátumát. Az információ visszakeresésekor a szurrogátum jelenik meg találatként az adatbázisokban, s ebből lehet eljutni az eredeti szöveghez. A tömörítés mértékétől függően a szurrogátum sokféle lehet. Eredeti dokumentumot képviselhet akár egy kódszám is, bár ennek információtartalma minimális. Másik végletként megvalósulhat a teljes szöveg tárolása, amely kiegészülhet más adatokkal is. Egy tipikus szurrogátum részei – *Horváth Tibor* összefoglalását alkalmazva (Könyvtárosok kézikönyve 2. 2001.) a következők:

- ☞ Azonosító(k);
- ☞ **Bibliográfiai leírás**
- ☞ Szabad szöveges ismertetés, szöveges tömörítés; kivonat referátum, annotáció;
- ☞ generalizáló osztályozási jelzet, kód;
- ☞ individualizáló osztályozási jelzet, kód;
- ☞ felhasználói kódok;
- ☞ alkalmazási kódok.

A könyvtáros a feltárási során a dokumentumokról reprezentációt készít, amelynek egyik legelterjedtebb formája a bibliográfiai leírás, illetve a besoro-

lási adatokkal kiegészített bibliográfiai tétel (számítógépes környezetben: bibliográfiai rekord). A bibliográfiai tétel önmagában is megjelenhet szurrogátumként.

A feltárás célja mindig az, hogy a majdani kommunikációs folyamatban a használó a leggyorsabban és leghatékonyabban találja meg a releváns információt problémája megoldásához, ezért a könyvtáraknak gondoskodniuk kell a használói igényekhez kapcsolódó információk magas színvonalú, korszerű feldolgozásáról.

A könyvtárak automatizálásával együtt a hagyományos bibliográfiai feltárási eszközök újabb, hatékonyabb eszközökkel, megoldásokkal cserélődnek le. Változik a technológia, és változnak a szellemi eszközök (szabványok, szabályzatok, metaadatok stb.) is.

A bibliográfiai adatelemek szerepe a dokumentumok egyértelmű azonosítása, az egymással összefüggő dokumentumok és a hozzáférési pontok megmutatása.

*Hagler, R. (1997)* szerint a bibliográfiai adatok, ill. adatelemek jelentik azokat a „legkisebb építőköveket”, amelyek együttesen azonosítanak egy dokumentumot vagy a dokumentum bibliográfiai rekordját. Ezek a – szabványosan összeállított és használt – építőelemek összefüggő és széles körű struktúrát alkotnak, bármely dokumentum legeredményesebb meghatározásához, és a bibliográfiai adatbázis leghatékonyabb kezeléséhez. (Rác, 2001)

Nemzetközi és nemzeti szabványok egész sora rendelkezik ezekről a kérdésekről.

Elkészültek az ISBD-ajánlások (International Standard Bibliographic Description) elsősorban a bibliográfiai információk nemzetközi felhasználásának és a bibliográfiai adatok cseréjének segítésére, valamint további nemzetközi és nemzeti szabványok, szakmai szabályzatok kerültek kidolgozásra. Az ISBD-hez hasonló szerepet tölt be a besorolási adatok terén, az 1984-ben megjelent GARE

(Guidelines for Authority and References Entries) útmutató, amelyet az IFLA katalogizálási és információtechnológiai bizottsága dolgozott ki, megcélözva egy nemzetközi authority-rendszer (egységesített besorolási adat-állomány) kialakítását. A GARE meghatározza az egységesített besorolási adatok adatelemeit, azok sorrendjét, központozási jeleit, az adathoz tartozó utalókat, kapcsolódó adatokat és megjegyzéseket, azaz egy nemzetközi szinten egységes szabályrendszer felállítására vállalkozik (Rác, 2001).

## **Az információtechnika hatása a bibliográfiai adatfeldolgozásra**

Az információtechnika változásai pl. az internet használata; a bibliográfiai adatokat kezelő könyvtári számítógépes rendszerek elterjedése, a helyi és nemzeti könyvtári hálózatok létrejötte, az elektronikus központi katalógusok megjelenése, az országos közös katalogizálás megszervezése, az információkeresési, dokumentumtovábbítási protokollok kidolgozása, a világszabvánnyá váló Z39.50-en alapuló ügyfélkiszolgáló technológia megteremtése; a digitalizálás térhódítása, s a retrospektív konverziók megindulása – hogy csak néhány fontos tényezőt emeljünk ki – alapvetően átalakították a könyvtárak működését.

A digitalizálás eredményeként folyamatosan bővül a könyvtárakban gyűjtendő és feldolgozandó dokumentumok köre, ami azt is jelenti, hogy megnő a katalogizálást végzők szerepe a világhálón elérhető információforrások feldolgozásában, hiszen a helyi és távoli elérésű elektronikus dokumentumokat ugyanúgy be kell vonni a bibliográfiai számbavétel körébe, mint a hagyományos dokumentumokat.

A retrospektív konverzió révén megvalósulhat a széles körű hozzáférés a korábbi cédulakatalógusokhoz, tehát a korábban hagyományos eszközök-

kel feltárt könyvtári állományrészek sem válnak „holtta” elektronikus korunkban.

Egy mű tetszőleges megjelenési formájának, ill. e megjelenési forma tartalmának és bibliográfiai kapcsolatainak integrált feltáráshoz szükségessé vált a meglévő katalogizálási szabványok és adatsere-formátumok továbbfejlesztése, valamint a katalogizálási szabályok újragondolása.

Az IFLA egységesítési törekvései a katalogizálás területén megkívánták, hogy az eredményes információcseré érdekében ne csak a bibliográfiai tételek tartalmával és formájával foglalkozzanak, hanem a gépi rekordok formátumával is. Az 1981-ben megjelent ISO 2709 szabvány rögzítette a szabványos rekordszerkezetet és a továbbiakban minden MARC formátumot erre a szabványra építettek, megnövelve a könyvtári rendszerek közötti rekordcseré lehetőségét.

A MARC formátumokat az egyes országok katalogizálási szabályzatainak megfelelően alakították ki, így – bár hasonlítanak egymásra – lényeges különbségek is vannak közöttük. Ez okozza, hogy az egyes MARC formátumok között nem valósulhat meg a közvetlen adatcseré, hanem konverziós programokra van szükség.

Az országos jelentőségű MARC-okon kívül egyes szoftverek vagy rendszerek saját formátumokat is kialakítottak, amelyek tovább színesítik a használatban lévő formátumok körét.

A hatékony információcseré érdekében a katalogizálási szabályok nemzetközi egységesítésének az adatsere-formátumokra is ki kellett terjednie. Szükségessé vált egy nemzetközi csereformátum kidolgozása, amely közvetítőként működhet az egyes MARC formátumok között, és segítheti a konverziós programok kialakítását. Ez lett a UNIMARC, melyet az IFLA elsősorban a nemzeti könyvtári rendszerek közötti adatcserére szánta.

Léteznek más egységesítési törekvések is, amelyek a különböző formátumok összehangolását célozzák. Az UNESCO által 1974-ben közreadott

UNISIST formátum a referáló és indexelő szolgáltatások adatbázisai közötti információcseré céljait szolgálja. A nemzetközi egységesítés érdekében 1978-ban megkezdődött a UNIMARC és UNISIST összehangolása, és kidolgoztak egy közös adatsere-formátumot, a CCF-t (Common Communication Format). Célja az adatkommunikáció a könyvtárak és a tájékoztató szolgáltatások között oly módon, hogy meghatározták a minden rekordban kötelező és a dokumentum azonosításához nélkülözhetetlen adatelemek összességét.

A MARC formátumok egységesítése nemcsak az IFLA keretében folyik. A MARC-harmonizációnak elkeresztelt együttműködési programtól azt várják, hogy megkönnyítse a rekordok átvételén alapuló katalogizálást, valamint a konverzió redukálása, ill. elmaradása következtében egymás rekordjai gyorsabban hozzáférhetőek lesznek, s olcsóbb lesz a katalogizálás. A program keretében az AACR (Anglo-American Cataloguing Rules) katalogizálási szabványra támaszkodó USMARC, CANMARC, UKMARC formátumok összehangolására törekednek.

A nemzeti formátumok egymáshoz közelítését ezen túlmenően számos más projekt is sürgeti.

Az országonként eltérő könyvtári szabványok, illetve az egyes könyvtárakban használt különböző számítógépes könyvtári rendszerek nyomán kialakult nemzeti MARC formátumok közötti konverzió a katalogizálás jelenlegi legnagyobb problémája. Az IFLA 1987-ben indította el a UBCIM (Universal Bibliographic Control and International MARC) programját, melynek célkitűzései (a bibliográfiai számbavételt biztosító szabványok és rendszerek nemzeti szintű fejlesztésének, valamint a bibliográfiai adatok nemzetközi szintű cseréjének koordinálása, és a géppel olvasható bibliográfiai adatok cseréjének szabványosítása, a UNIMARC formátum fejlesztése) illeszkednek a kialakult kedvezőtlen helyzethez (Rácz, 2001).

## Az elektronikus dokumentumok és feldolgozásuk

Az elektronizáció korában megjelent új dokumentumtípusokat, vagyis az elektronikus dokumentumokat (és új információhordozókat) is fel kell tárniuk a könyvtáraknak.

Azokat a formátumbővíteket és -továbbfejlesztéseket, amelyek az elektronikus dokumentumok katalogizálhatósága érdekében történtek, a formátumok többségén az 1990-es évek folyamán végezték el. A következő ábra bemutatja a UNIMARC nemzetközi csereformátum esetén azokat az adatmezőket, amelyek az elektronikus dokumentumok leírásakor specifikus tartalommal bírnak.

A hálózati információk feldolgozására különböző metaadat-rendszerek is létrejöttek. Ezek a rendszerek nem a MARC kiváltására születtek (sőt ezek kö-

zül számosat eredetileg nem is könyvtári alkalmazásra fejlesztettek ki), hanem arra, hogy a weben fellelhető, egymástól jelentősen különböző elektronikus dokumentumokat le lehessen írni, és a visszakeresésüket könnyebbé lehessen tenni.

A metaadat olyan adat, amely az információs objektumokat reprezentáló forrásadatokat jellemzi, viszonyait leírja, elősegíti az objektumok fellelhetőségét és hatékony felhasználását. Néhány ismert metaadat-rendszer:

- ☞ OCLC InterCat;
- ☞ Dublin Core;
- ☞ IAFA (Internet Anonymous FTP Archive) -sablonok;
- ☞ WWW Semantic Header;
- ☞ URC (Uniform Resources Characteristic or Uniform Resources Citation);
- ☞ TEI (Text Encoding Initiative) – fejléc;
- ☞ INDECS (Interoperability of Data in E-Commerce Systems);

Cím	200	Megjegyzések:	
Párhuzamos cím	510	Elektronikus dokumentum típusa	336
Szerzőség	200\$f \$g és/vagy 700, 701, 710, 711	Rendszerkövetelmények	337
Kiadás	205	Címre vonatkozó megjegyzések (pl. a cím forrása)	304
Megjelenés	210	Kiadásra vonatkozó megjegyzések (pl. licenz jogok)	305
Fizikai jellemzők	215, 230	Megjelenésre, terjesztésre vonatkozó megjegyzések (pl. shareware)	306
Kísérő anyagok	215, 307	Sorozatra vonatkozó megjegyzések	308
Sorozat	225, 410	Elérhetőségre vonatkozó megjegyzések	310
Elérési információk	345, 010, 011	Tartalomjegyzések	333
		Elektronikus elhelyezés és hozzáférés	856
Kódolt információk:			
Elektronikus forrás nyelve	101 \$a		
Célközönség	100 /17-19		
Publikálás dátuma	100 /8-16		
Publikálás vagy előállítás helye	102		
Elektronikus forráshoz kapcsolódó kódolt adatok	135		

### 2. ábra.

A UNIMARC elektronikus dokumentumokra vonatkozó adatmezői

```

001 123456789
100 ## $a19950105d1991####k||y0engy0103####ba
101 0# $aeng
135 ## $adrunu--uuuuu
200 1# $aWho's who in the Internet$bcomputer
file$ebiographies of IAB, IESG, and IRSG members$fG.
Malkin$gNetwork Working Group
210 ## $d1991
230 ## $aComputer data
300 ## $a"August 1991"
304 ## $aTitle from title screen
330 ## $aBibliographical information about members of
the Internet Activities Board (IAB), the Internet
Engineering Steering Group (IESG) of the Internet
Research Task Force (IETF), the Internet Research
Steering Group (IRSG) and the Internet Research Task
Force (IRTF)
336 ## $aText
700 #1 $aMalkin,$bGary Scott
801 #0 $aUS$bDLC$c19960104$gAACR2
856 1# $uftp.sunet.se/pub/Internet-documents-fyi/
fyi9.html
856 4# $uhttp://www.globecom.net/ietf/rfc/rfc13336.
856 4# $uhttp://webrum.uni-annheim.de/rz/nerz/onfig/
1web/1netz/1inte/1fyi/9fyi0.htm
    
```

### 3. ábra

*Egy elektronikus dokumentum UNIMARC  
bibliográfiai rekordja*

- ☞ DOI (Digital Object Identifier);
- ☞ DDI (Data Documentation Initiative);
- ☞ EAD (Encoded Archival Description);
- ☞ GEM (Gateway to Education Materials);
- ☞ GILS (Government Information Locator Service);
- ☞ ONIX (Guideline for Online Information Exchange<sup>®</sup>);
- ☞ VRA (Visual Resources Association).

A digitális állományok tartalmának leírására szolgáló metaadatok használata elengedhetetlen a nagy gyűjtemények pontos és megfelelő megismerésé-

hez. A metaadatok segítenek a szabadszöveges keresés félreérthetőségének kiküszöbölésében és olyan szemantikai szempontok megadásában, amelyek szűkítik és konkretizálják az információkeresést.

Az elektronikus dokumentumok katalógusba történő integrálását illetően jelenleg még két szemléleti irányzat uralkodik. Az egyik értelmezésben a bibliográfiai leírás nemzetközi szabványaival (ISBD) összhangban alakítható ki az elektronikus források ún. metaadatainak beépítése az OPAC-ba, az erre kidolgozott speciális szabványok (pl. Dublin Core) alapján, valamint MARC formátumra konvertáló programok (MARC-SGML) segítségével. Itt a cél az, hogy adatcsere esetén kiderüljön, miről szól a dokumentum.

A másik nézet szerint az internetforrások feltárása a hagyományos katalógustétel struktúrájával való szakítást igényel, mert nem a tétel egyes mezőinek, hanem a tartalomnak a kulcsszavas keresésére épül. Vagyis a lehető legegyszerűbb formára törekszik, hogy a keresés könnyebb legyen.

A metaadat-rendszerek közül a Dublin Core struktúrájáról lehet leginkább elmondani, hogy nem idegen a MARC-tól, megfeleltetésükről konverziós táblák segítségével lehet gondoskodni. Elterjedését elősegíti, hogy adatait az európai szabványosítási szervezet, a CEN (European Committee for Standardization) is jóváhagyta.

A szerteágazó igények miatt ma még nem beszélhetünk egységes, témafüggetlen metaadat-formátumról. A különböző változatok a leírás mélysége szerint csoportosíthatók.

1990-ben Stockholmban szemináriumot rendeztek a bibliográfiai rekordokról, ahol a különböző célú felhasználás szempontjából tárgyalták a bibliográfiai rekord sajátosságait. A szeminárium határozatainak egyike a bibliográfiai rekordok funkcionális követelményeinek kutatását tűzte ki célul. A projekt a bibliográfiai adatok megváltozott környezetéből indul ki. Az elkészült és 1998-ban megjelent FRBR (Functional Requirements for Bibliographic Re-

cords)-tanulmányok kettős célja van: definiálja azokat a funkciókat, amelyeket a bibliográfiai rekord betölt a különböző kiadványtípusokat, a különböző alkalmazásokat és a különböző felhasználói igényeket illetően; ajánlást teygen a rekordok funkcióinak alapszintjére és javasolt minimális elemkészletére.

Ezen túlmenően az FRBR-tanulmány felállítja a bibliográfiai adatok entitás-kapcsolat modelljét. A tanulmány a legszélesebb értelemben tekinti át a bibliográfiai rekord funkcióit. Vagyis egy rekord nemcsak leíró elemeket tartalmaz, hanem a hozzáférési pontokat (név, cím, tárgyszó, stb.), egyéb „szerző” elemeket (osztályozási jelzet stb.) és az annotációt is. Így olyan szerkezetet hoz létre, ami tisztán, egyszerűen és pontosan értelmezi, hogy miről ad információt a bibliográfiai rekord, és milyen felhasználói igényeket kell a rekordnak megvalósítani. A kutatócsoport elemezte a bibliográfiai adatokkal kapcsolatos entitásokat, meghatározta attribútumaikat és kapcsolataikat.

Az elemzés gyakorlati célja a katalogizálás racionalizálása, elméleti célja pedig a bibliográfiai univerzumnak, céljainak és kapcsolatrendszerének leírása, kiindulási alapot szolgáltatva a további formátum-fejlesztésekhez és kutatásokhoz.

Emellett új lendületet kapott az egységesített besorolási adatok kérdése a FRANAR (Functional Requirements and Numbering of Authority Records) program elindításával. Célja, hogy elősegítse a virtuális, osztott besorolási-adatállományok gyakorlati megvalósulását, valamint kidolgozza a besorolási rekordok funkcionális követelményeit, megvalósítási tanulmányt készít az ISADN (International Standard of Authority Data Number)-ről.

A (félleg) strukturált adatok, strukturált szövegek leírására és számítógépes környezetben történő kezelésében új lehetőségeket teremt az XML (Extensible Markup Language – Kiterjesztett Jelölő-

nyelv), amely a World Wide Web Consortium (W3C) 1998-ban megjelent nyílt szabványa.

Az XML számos olyan előnyös tulajdonsággal rendelkezik, amellyel kivívhatja a könyvtáros szakma figyelmét is, például a formátum-fejlesztések területén. Olyan valós problémákra kínál megoldást, amelyekkel a napjainkban használt könyvtári számítógépes rendszerek adatformátumai jelenleg is küzdenek. Ezen túlmenően az XML könyvtári területen történő alkalmazása megoldást jelenthet az Internet előretöréséből fakadó problémákra, valamint nyíltabb rendszer létrehozását teszi lehetővé a MARC alapú zártabb könyvtári rendszerekhez képest. (Nyílt rendszer az, amely olyan részekből áll, melyek egységes, gyártó független és nemzetközileg elfogadott szabványokon alapuló felületeket nyújtanak. Ennek köszönhetően az egyes részek rugalmasan összekapcsolhatók, valamint segítenek összehangolni a különböző hardver és szoftver termékek használatát.) Az internet globális világában egyre fontosabbá váló követelmény, hogy a könyvtári OPAC-okhoz hasonló közszolgáltatások nyíltak, átjárhatók, más keresők által elérhetők, mélyen linkelhetők legyenek. Ma ennek a követelménynek egy átlagos könyvtári rendszer maradéktalanul nem képes megfelelni.

## **Kutatások a Debreceni Egyetem Könyvtárinformatikai tanszékén: a BDML nyelv**

A Debreceni Egyetem Komputergrafikai és Könyvtárinformatikai tanszékén elindult egy új fejlesztés, amelynek célja a bibliográfiai leírás formalizálása az XML jelölőnyelv segítségével, támaszkodva a munka során a nyelvnek azon fontos jellemvonásaira, amelyek a könyvtári területen történő alkalmazhatóságát erősítik.

Az XML segítségével ugyanis újabb jelölőnyelvek készíthetők. Ebből következően a fejlesztés egy olyan formális szabálygyűjtemény kialakítására irányult, amely XML nyelven megadja a szabványos bibliográfiai leírások szerkezetének sémáját. A készülő nyelv a BDML (Bibliographic Description Markup Language) nevet kapta.

Az XML érdekes lehet mindazok számára, akik a dokumentumaik kialakításában és a kezelésében nagyobb rugalmasságot kívánnak elérni, akik sokoldalúan támogatott rugalmas tárolási vagy általános adatformátumot, valamint fejlett, ám ezzel együtt nem túl drága alkalmazásokat keresnek. Az XML a számítógép számára megkönnyíti az adatok generálását, olvasását és biztosítja, hogy az adatszerkezet egyértelmű legyen.

Az XML nem egy programozási nyelv és használatához vagy megtanulásához nem szükséges programozónak lenni (bár egy „kis” informatikai ismeret sosem árt). A legtöbb számítógépes adatformátumtól eltérően az XML-jelölések az ember számára is érthetőek, mivel semmi más nem tartalmaznak, mint közönséges szöveget.

Az XML biztosítja az információ tervezését, autonómiáját, többszöri felhasználását, kereshetőségét, számítógépes feldolgozhatóságát, információk összekötését, az információ karbantartását és a minőséget, azaz a tudatos bánásmódot.

- |               |   |                                   |
|---------------|---|-----------------------------------|
| ☞ XML         | } | <b>BDML nyelv<br/>fejlesztése</b> |
| ☞ FRBR        |   |                                   |
| ☞ FRANAR      |   |                                   |
| ☞ Dublin Core |   |                                   |
| ☞ ISBD-k      |   |                                   |
| ☞ GARE        |   |                                   |

*A BDML nyelv „pillérei”*

A BDML nyelv egy kísérlet arra vonatkozóan, lehetséges-e az XML segítségével olyan új jelölőnyelvet

kifejleszteni a bibliográfiai leírások összeállítására, amely lehetővé teszi a bibliográfiai leírásban rejlő lehetőségek jobb kihasználását a leírás azonosító funkciójának csorbítása nélkül.

A ma létező metaadat-rendszerek többségében az elemek mindegyike szabadon választható és ismételhető, ezek sorrendje nem meghatározott, azonban az általános érthetőség miatt mégis kötött szó-készlet alkalmazása ajánlott vagy szükséges. Az egyes elemeket különböző minősítőkkal lehet rendszerint finomítani, jelentésüket szűkíteni.

A BDML elsősorban a magyar katalógizálási szabványokra, szabályzatokra, ill. előírásokra támaszkodik. Illeszkedve a bibliográfiai leírás definíciójában szereplő „meghatározott szabályok szerinti egységes szerkezetben, formában és sorrendben leírt adatok összessége” gondolathoz, valamint a nemzetközi érthetőség érdekében, az ISBD szabályzatokban elfogadott adatszoportokat, adatelemeket megjelenítő, kötött sorrendű elem-készletet deklarál a BDML nyelv.

A bibliográfiai leírás fogalmából következik, hogy magának a leírásnak formálisan értelmezhetőnek, az adatelemeknek formálisan felismerhetőnek kell lenniük. Ha a könyvtáros szakember rápillant a leírás adott pontjára, máris tudja, hogy ott például egy kiadó megnevezéséről van szó, hiszen az adatelem környezetéből, ill. a megelőző egyezményes jelből ez egyértelműen kiderül. A számítógépek nem ilyen „okosak”: pontosan meg kell mondunk a számukra, mi mit jelent, az hol található és mi a teendő vele. Az XML olyan új nyelv, amelynek kifejezett célja az, hogy az adatok önmagukért beszéljenek.

A BDML rendelkezik az elemek sorrendjéről, amely a bibliográfiai adatformátumoknál egy kritikus szempont. Az elemekhez szöveggörnyezettől függően sorrendi szabályozást ad (az elemek XML-ben két logikai művelettel kapcsolhatók össze, a sorozat, illetve a választás összekapcsolóval) és megad előfordulási szabályokat (az elemek kötelezősége)

gére és ismételhetőségére vonatkozóan) az ún. mennyiségi jelzők alkalmazásával. Mivel XML-ben lehetőség van az elemek tetszőleges mélységű egymásba ágyazására, nagyfokú rugalmasság érhető el. Ennek köszönhetően a bibliográfiai leírás adatelemeinek, vagy néhány összetartozó adatelemének ismételhetősége, opcionálitása, feltételektől függő kötelezősége (például a bibliográfiai leírás 1. adatszoportjában a különböző párhuzamosan meglévő adatelemek kezelése, akár – könyvek esetében – szerzői vagy egyes gyűjtemények esetén is) BDML-formátumban nem jelent problémát, rugalmasan kezelhetők a különböző előfordulás-variációk.

## A BDML nyelv jellemzői

- ☉ Rugalmas, tetszőlegesen bővíthető, platform-független metanyelv, ill. adatformátum (az adatstruktúra tartalmi és nem formai szempontok alapján történő leírása, amelyet elektronikus úton tárolunk olyan formában, hogy egyaránt érthető legyen a számítógépes programok és az emberek számára is);
  - ☉ Lehetővé teszi, hogy a metaadatok elkülönüljenek a megjelenítési formátumoktól;
  - ☉ „Önleíró” adatelemek (a megcélzott közönség és a használt média követelményeihez igazodóan használható, alakítható és formázható);
  - ☉ Rugalmasan illeszkedik a feldolgozandó dokumentumhoz (az XML segítségével definiált nyelvek szükség szerint bővíthetők, ugyanis ezekben nincs előre definiált elemlista, saját elemeket tetszőlegesen lehet deklarálni);
  - ☉ A különböző nyelveken írt szövegek kevert használata (a UNICODE karakterkódolási-rendszer megengedi a diakritikus és egyéb speciális jelek, valamint a nem latin betűk alkalmazását, normál szöveggént való kezelését; XML nyelvet értelmezni képes szoftver e karakterkészletek bármely kombinációját kezelni tudja, így az XML
- az adatcserét nemcsak a különböző számítógépes rendszerek, hanem az országok és kultúrák között is lehetővé teszi);
- ☉ Bibliográfiai rekordok hipertext rendszerű felfűzése;
  - ☉ Költséghatékony eszköz (az XML nyelv bárki által szabadon hozzáférhető, nyílt szabvány, nincs tulajdonosa, és egyetlen gazdasági érdek sincs rá kizárólagos befolyással).
- Az XML kódok önmagukban nem tartalmaznak utasítást arra vonatkozóan, hogyan jelenjenek meg az adatok a képernyőn vagy nyomtatásban, azaz a kódok a tartalom jellegére koncentrálnak függetlenül a megjelenítő közegtől. A kódolt tartalomhoz tetszőleges számú ún. stíluslap készíthető, s az ezekben a stíluslapokban szereplő szabályokat alkalmazva az anyag a különböző eszközöknek megfelelően automatikusan formázható újra és újra. A bibliográfiai leírások formalizálásánál jól kihasználható az XML legfőbb jellemzője: a tartalom, a megjelenítés és a kapcsolások kiválasztása, amelyek így külön-külön optimálisan kezelhetők. A felhasználást tekintve ez azt jelenti, hogy a bibliográfiai leírások különböző megjelenítési (output) formátumai külön-külön egy-egy stíluslap elkészítésével kényelmesen megoldhatók (figyelembe véve azt is, hogy az elemek – pl. a leírás adatelemeinek, adatszoportjainak stb. – sorrendje XML-ben egyszerűen előírható).
- A gyengén látó emberek számára is előnyös ez a megközelítés. A stíluslap, ill. a megfelelő szoftver segítségével átalakíthatják az XML-t Braille írássá vagy beszéddé.
- A BDML DTD megalkotásával és használatával számos előnyhöz juthatunk. A programozók készíthetnek adatkezelő és feldolgozó szűrőket anélkül, hogy a programokat nem várt adatok kezelésére fel kellene készíteni. Ugyanilyen nagy biztonsággal készíthetők stíluslapok is, és a BDML minden eleméhez stílusszabályok rendelhetők. XML-kompatibi-

lis szerkesztő használatával a feldolgozó könyvtárosok irányíthatók, és egységes szabályoknak megfelelő BDML-dokumentumok (bibliográfiai leírások) készíthetők (bár az XML megengedi a felhasználó által definiált újabb elemek létrehozását, a szabályozott környezetben használt szerkesztőprogramokban ez a lehetőség letiltható).

## Összegzés

A dokumentumok és források feldolgozásához szabályok, szabványok, kézikönyvek, dokumentációk stb. tekintélyes fegyvertára áll a rendelkezésünkre. A komplex, elektronikus környezet, természetesen, jól felkészült szakembereket kíván, akik képesek eligazodni és lépést tartani a gyorsan változó körülményekkel és elvárásokkal.

Mindebből következik, hogy a dokumentumok, információs források körültekintő és alapos feldolgozása meglehetősen költséges feladat. Ez a költség a számítástechnika és a telekommunikációs vívmányok könyvtári alkalmazásával, a munkamegosztást célzó különféle könyvtári programokkal (közös, osztott katalogizálás), már létező bibliográfiai rekordok letöltésével és helyi igényeknek megfelelő módosításával ésszerűbbé tehető, és nem mellékes az sem, hogy ezeknek a rekordoknak a színvonala és minősége ideális esetben növelhető. Napjainkban azonban még sok időt és energiát követel a különböző könyvtári adatformátumok közötti összhang megteremtése, az egyik adatformátumban szereplő adatok egy másik adatformátumba való konvertálása. Ezen túlmenően emberi közbeavatkozás szükséges annak eldöntésére is, hogy vajon történt-e adatvesztés az átalakítás során, illetve a rekordban szereplő adatok különféle javításaira. S ha még a weben is szeretnénk hozzáférhetővé tenni a katalógusunkat (lehetőleg minél kevesebb munkával)?!

Az XML-ben való gondolkodás megoldást jelenthet mindezekre nehézségekre, hiszen fő célja az internetes információcsere támogatása.

## Irodalom

BERKE Barnabásné: Párizs 1977 – Koppenhága 1998. A nemzeti bibliográfiai ajánlások megújítása. In: Könyv, könyvtár, könyvtáros, 9. évf. 2000. 8. sz. p. 11–20.

BODA, Károly I.: Use of Hypertext in Information Science: Concepts, Systems, Models and Applications. (Ph.D. dissertation.). University of Debrecen, 2000.

BRADLEY, Neil: Az XML-kézikönyv. Bicske, Szak. 2000.

BUCKLAND Michael.: Redesigning library services. A manifesto. Chicago, ALA, 1992. Fordításban: A könyvtári szolgáltatások újratervezése. [ford. Murányi Péter, Pálvölgyi Mihály] Bp. OSZK, 1998. 102 p.

BURKART, M.: Dokumentationssprachen. In: Buder, M. – Rehfeld, W. – Seeger, T. (hrsg.): Grundlügen der praktischen Information und Dokumentation. München, Saur, 1990.

Extensible Markup Language (XML) 1.0. 2. ed. - <http://www.w3.org/TR/REC-xml>

HAGLER, R.: The bibliographic record and information technology. 3. ed. Chicago-London, ALA-Ottawa, CLA, 1997.

HORVÁTH Tibor: A feldolgozás egységei és a feltárás célja. In: Könyvtárosok kézikönyve. 2. Feltárás és visszakeresés. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2001. 15–34. p.

IFLA study on functional requirements for bibliographic records. In: International Cataloguing and Bibliographic Control, vol. 24. 1995. no. 3., p. 43–50.

Mandatory data elements for internationally shared resource authority records. Report of the IFLA UBCIM Working Group on Minimal Level Authority Records and the ISADN. 1998.

MILLER, Dick R.: XML and MARC : a choice or replacement? Chicago, 2000, ALA., 12 p. - <http://elane.stanford.edu/laneauth/ALACHicago2000.html>

POPRÁDY Géza: Könyvtári trendek. In: Könyvtárosok kézikönyve. 5. Segédletek. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2003. 13–62. p.

RÁCZ Ágnes: A kiadványok bibliográfiai számbavétele,; leíró katalogizálás. In: Könyvtárosok kézikönyve. 2. Feltárás és visszakeresés. Szerk. Horváth Tibor, Papp István. Bp. Osiris, 2001. 187–296. p.

TENNANT, Roy: MARC must die. In: Library Journal, 2002, 10. - <http://libraryjournal.reviewsnews.com>

VAJDA Erik: A besorolási adatok egységesítése. Fogalmak tisztázása és terminológiai javaslat. In: Tudományos és Műszaki Tájékoztatás, 37 évf. 1990. 5. sz., p. 186–190.