

## Az Interneten elérhető információforrások katalogizálása\*

*Az Interneten elérhető hatalmas mennyiségű információforrás szerverek ezerein oszlik meg. Ezek a dokumentumok változó minőségűek, könnyen manipulálhatók, elérési helyük gyakran változhat; hogy csak néhány jellemzőt említsünk. Hogyan lehet hozzáférést biztosítani a hagyományos dokumentumoktól olyannyira eltérő tulajdonságokkal rendelkező hálózati információforrásokhoz? A tanulmány néhány jelenlegi digitális könyvtári projektet megvizsgálva veszi számba az Internet rendszerezésére irányuló legújabb törekvéseket.*

Az Internet állandóan „építés alatt” van. A róla publikált cikkek vagy könyvek gyakran már kiadásuk időpontjában elavultnak számítanak. A legkurresebb és legrelevánsabb információ többnyire csak az Interneten található meg. Nagyon sok írás – éppúgy, mint a jelen tanulmány is – ezért az Interneten elérhető forrásokra épül, még akkor is, ha ezek a források általában tünékeny természetűek, vagy szerzőik eredete olykor kétséges. Valamekkora minőségi biztosítékot jelent azonban az, ha olyan elektronikus dokumentumokat használunk fel, amelyek valamely szakmai szervezet jóvoltából érhetőek el, és a dokumentumokat még akkor is el tudjuk érni valahol, ha az Interneten való elérési címük, azaz URL-jük (Uniform Resource Locator) időközben megváltozik [1]. Sajnos azonban ez nem mindig lehetséges, mert nagyon sok kiváló dokumentum nem tartalmaz megfelelő információt a származására vagy közreadójára vonatkozóan.

A weboldalak és az Interneten elérhető egyéb források exponenciális növekedésével az Interneten, és annak legközvetlenebb felületén, a World Wide Weben, egyre nehezebb megtalálni az éppen szükséges információt. A robotokkal felszerelt keresőoldalak, mint például az Infoseek vagy az AltaVista, folyamatosan fejlesztik ugyan keresési lehetőségeiket, de a gyakorlatlan keresők számára továbbra is sokszor csalódásokat okoznak. A tárgy szerinti felosztású, kategóriaorientált listák, mint például a Yahoo!, könnyű kereshetőséget kínálnak, de az általuk lefedett területek mérete és kiterjedtsége meglehetősen korlátozott. Függetlenül tehát attól, hogy milyen keresőrendszert használnak, még a gyakorlott felhasználók is gyakran elgondolkoznak, hogy vajon létezne-e jobb megoldás a keresésre [2].

A jelenleg létező több millió Internet-forrás szerverek ezerein oszlik el. „Ez az anyag alig rendszerezett, minősége és stabilitása változó, nehéz

értelmezni, böngészni, keresni, vagy hivatkozni rá” [3]. J. Quittner már 1992-ben felhívta a figyelmet arra, hogy az Internet legnagyobb problémája: olyan sok információt tartalmaz, hogy nemcsak nehéz megtalálni azt az információt, amire szükségünk van, de sokszor még azt sem tudjuk, egyáltalán hogyan, illetve hol kezdjük el a keresést [4]. Ed Krol, az „Autóstoposok kézikönyve az Internethez” szerzője szintén úgy jellemezte az Internetet, hogy az nem más, mint egy könyvtár, ahol a könyvek szanaszét hevernek a padlón, és nincs katalógus. A rendszerezésre való igény tehát már elég korán megvolt a számítógépes szakemberek részéről, és közülük többen is könyvtári analógiákat használtak, viszont nem gondoltak arra, hogy a könyvtárosok által már kialakított rendszereket használjanak az elektronikus dokumentumok rendszerezésére [5].

Lehet-e egyáltalán rendszerezni az Internetet? Ha igen, akkor a válasz a világháló rendezetlenségére vajon a technológiában, azaz a nagyobb, gyorsabb keresőprogramokban, vagy pedig a humán erőforrásokban, azaz a kicsi, de bizonyos szempontok szerint értékelt hipertext kapcsolatokból álló gyűjteményekben rejlik? Elképzelhető, hogy az Internet nagyjából szervezetlen, rendezetlen marad változékonysága, a publikálásban mindenki számára egyenlőséget biztosító jellege miatt, de ugyanakkor egyre nő az olyan projektek száma, amelyek kísérletet tesznek az Internet-források valamiféle rendszerezésére, és ezen törekvésük során vagy új rendszerezési technikák kifejlesztésével, vagy pedig a könyvtárakban már bevált és évtizedek óta használt rendszerek felhasználásával próbálkoznak [2].

\* A tanulmány a Neumann-ház számára készült a Nemzeti Kulturális Alap támogatásával.

## Lehet-e megszerezni az Internetet?

Sok könyvtáros megkérdőjelezi, hogy az Internet jelenlegi állapotában egyáltalán megszerezhető-e, mert úgy tartja, hogy az Internet jelenleg túlságosan instabil az üzleti versengésből származó újabb és újabb technológiák és termékek gyakori megjelenése miatt. P. Caplan néhány évvel ezelőtt még a következő kérdéseket tette fel a MARC és az AACR szabványok online dokumentumokra való kiterjesztése kapcsán: „Érdemes-e a könyvtárosoknak az Interneten talált anyagok katalogizálásával foglalkozni? Ha igen, akkor a MARC a legmegfelelőbb erre a célra? Az olyan hálózati eszközök, mint az Archie vagy a gopher vajon nem teszik-e fölöslegessé a könyvtári katalógusokat?” [6]

E. Jul a könyvtári szakemberek Internet-források katalogizálására vonatkozó reakcióit megvizsgálva három okot jelöl meg, amely miatt a könyvtárosok sokszor idegenkednek az Internet forrásainak katalogizálásától [7]. Az első ok az, hogy sokan még mindig úgy tartják, az Interneten nincs semmi, ami katalogizálásra érdemes lenne. Ezt többnyire az magyarázza, hogy az Interneten nagyon sok az olyan, könnyen elérhető anyag, ami a könyvtárak gyűjtőköre kívül esik. Ha azonban belegondolunk a hagyományos könyvtári gyűjteményszervezésbe, akkor nyilvánvaló, hogy a könyvtárak a nyomtatott anyagokat is szelektálják, tehát az Internet esetében sem kell, illetve lehet teljességre törekedniük. Az Internet-források teljes kizárása viszont olyan értékes dokumentumok elvesztését jelentheti, amelyek adott esetben nemcsak megfelelnek a könyvtári gyűjtőkörének, de elengedhetetlen fontosságúak is.

A második általános feltételezés az Internet forrásait illetően: minden, ami az Interneten megtalálható, holnap már elérhetetlen lesz. Igaz, hogy a megbízható, hosszú távú hozzáférés az Internet-források esetében még megoldatlan probléma, de ez nem jelenti azt, hogy valamennyi Internet-forrás tünékeny lenne. Érthető módon vannak olyan források, amelyek éppen jellegüknél fogva rövid életűek. Sok egyéb forrás viszont a URL változása miatt lesz elérhetetlen, és így a használó számára értéktelen. A fontosabb Internet-források létrehozói azonban megbízhatóságra törekednek, és a fő céljuk az, hogy használóik könnyűszerrel, és hosszú távon hozzá tudjanak férni weboldalaikhoz. A katalogizálóknak tehát elsősorban a stabil, megbízható források katalogizálására kell törekedniük. Természetesen ez a fajta szelekció önmagában nem mentesíti a katalogizálást végző intézményt a források rendszeres revíziója alól. Alternatív megoldást jelenthetnek a weboldalak nyomom követésére az Egységes Forrásnevek (Uniform Resource Name = URN) és az

Állandó Egységes Forráslokalizálók (Uniform Resource Locator = URL). A URN rendszer egy világszerte egyedi, és tárolási helytől független elnevezést támogat, amelynek segítségével kideríthető a forrás tárolási helye. A PURL lényege pedig az, hogy minden egyes aktuális URL-hez egy állandó URL-t rendel hozzá, és ily módon a forrás az aktuális URL megváltozása után is könnyűszerrel felkutatható.

A harmadik gyakori érv az Internet forrásainak katalogizálása ellen, hogy az eredetileg nyomtatott dokumentumokra kifejlesztett szabványok, mint a MARC vagy az AACR2, egyébként sem lennének alkalmasak a hálózati források leírására. Ez a probléma időről időre felmerül az új médiumok megjelenésekor. Sokan szeretnék, ha a MARC sokkal rugalmasabb lenne. Igaz, nem tökéletes, de mégis az egyik legelfogadottabb szabvány a gépek közti kommunikációra és a bibliográfiai adatcserére. A MARC Internet-forrásokra való adaptálásának nagyon sok közvetlen előnye van: jelenleg már sok MARC-alapú rendszer létezik; az osztott katalogizálás előnyeit továbbra is élvezhetik a könyvtárak; lehetséges az Internet-források rekordjainak a helyi katalógusokban való megjelenítése. Hasonló a helyzet az AACR2 szabályokkal. Az AACR2 szintén folyamatos revízióra szorul, hogy alkalmazkodni tudjon a változó technológiákhoz.

Jelenleg a World Wide Web legismertebb keresőoldalai nem könyvtárosok munkái. Rendszerezettségük mindig a keresőoldali létrehozójának „egyéni” elgondolásait tükrözi. A legtöbb keresőszolgáltatás elsősorban kulcsszavakon alapul, és lehetővé teszi a Boole-logika szerinti keresést. A csak kulcsszavakon alapuló Internet-indexelés azonban problematikus lehet. A WAIS (Wide Area Information Server) és a hozzá hasonló eszközök súlyozott relevanciaértékelést alkalmaznak, de még nagyon messze vannak a természetes nyelvi feldolgozás teljes kihasználásától, mivel nem képesek a kontextus értékelésére. Az eddigi tapasztalatok alapján a tárgyi hozzáférés majdnem mindig preferáltabb, mint a kulcsszó szerinti keresés. Ha valakinek keresést kell végeznie, akkor sokkal sikeresebb a tárgyi felosztású fastruktúrák használata során, mint a szóalapú keresési stratégiával, különösen akkor, ha a kereső nem jártas az adott szakterületen, vagy nem ismeri az adott nyelvet.

A könyvtárosszakmában az elektronikus dokumentumok nagyszámú megjelenésével az utóbbi években a „falak nélküli könyvtár” fogalma teljesen ismertté és elfogadottá vált. A hagyományos könyvtárak filozófiájában jelentős változás ment végbe: a hangsúly a dokumentumok fizikai birtoklásáról az információ elérhetővé tétele felé tolódott. A CATRIONA (Cataloging and Retrieval of Information Over Networks Applications) projekt [8]

résztevői szerint ahhoz, hogy a falak nélküli könyvtár realitássá váljon, először azt a problémát kell megoldani, hogy a felhasználó hogyan tudja az igényeinek megfelelő forrásokat és szolgáltatásokat megtalálni. Jelenleg úgy tűnik, hogy a megoldás az új szoftverek és a már elismert könyvtári módszerek és gyakorlatok kombinációjában rejlik. Az *Argus Clearinghouse* [9] hangsúlyozza: ahhoz, hogy az Internet sokkal hasznosabb információs környezetté váljon, a humán erőfeszítéseket kombinálni kell a kereső és böngésző eljárásokkal. Csak szoftveres megoldás nem elegendő. Az automatizált rendszerek számára a nyelv és a fogalmak túlságosan kétértelműek ahhoz, hogy megfelelően azonosítani és értékelni tudják őket. Jelenleg úgy tűnik, hogy a mesterséges intelligencia a közeljövőben még nem lesz képes megfelelni ennek az elvárásnak.

Ezt a nézetet láthatóan nem osztja az átlagos Internet-felhasználók és -szolgáltatók többsége. S. G. Steinberg a *Wired* magazinban azt írja, hogy bár a könyvtártudomány tűnik a legnyilvánvalóbb tudományterületnek, ahova az Internet rendszerezésének problémájával fordulhatunk, az mégsem tud segíteni, hiszen a szakma is elégedetlen a jelenlegi osztályozási rendszerekkel [10]. Arra a következtetésre jut, hogy az osztályozási rendszerek csak a polcon való elrendezésre alkalmasak, és alkalmatlanok az elektronikus dokumentumok rendszerezésére.

A Yahoo!, az egyik legelső és legsikeresebb web-keresőrendszer, abból a meggyőződésből kiindulva, hogy a meglévő osztályozási rendszerek nem tudtak lépést tartani a tudomány fejlődésével, megpróbált kitalálni egy új osztályozási rendszert. Ugyanakkor a Yahoo! katalogizálói arról panaszkodnak, hogy az egyes dokumentumok esetében milyen nehéz megtalálni azt az egy elérési pontot, amelynek alapján azt valamelyik kategóriához kapcsolják. Minden egyes URL számára egy rekord készül, és a rekord csak egy helyre kerül be a Yahoo! hierarchikus struktúrájába. Ha a Yahoo! a fazettás osztályozás adta lehetőségeket használná, akkor nem kerülne ilyen nehéz helyzetbe. Az elektronikus világban az osztályozásnak nem kell a polcok számára kifejlesztett lineáris elrendezést követnie. Azaz, az osztályozásban a Yahoo! inkább visszalépett, mintsem előre [1].

## Digitális könyvtárak

A digitális könyvtár kifejezésben a könyvtár szó ugyan tökéletesen tükrözi a projektek jellegét, ám a projektek nagyon kevés hagyományos könyvtári szaktudást tükröznek. A digitális könyvtárak felhasználják ugyan a könyvtárosok gyűjteményépítő tapasztalatait, amikor a saját felhasználói közös-

ségük igényeinek megfelelően válogatnak mind a helyileg elérhető szövegfájlok, képek, video- vagy hangfájlok, mind pedig az Interneten elérhető dokumentumok milliói közül, a digitális könyvtárakban levő források azonban rendszerint nem a könyvtárosok elvei szerint katalogizáltak, sem a fizikai, sem a tartalmi leírást illetően [1].

A digitális könyvtári projektek nagyon sok támogatást kaptak már eddig is olyan szoftverek kifejlesztésére, amelyek ezeknek a könyvtáraknak a kiépítését támogatják. A szakirodalom szintén bővelkedik az új, innovatív visszakeresésről és szövegelemző alkalmazásokról szóló cikkekben. A hagyományos osztályozást és a bibliográfiai leírást azonban csak ritkán említik. Ezekről a technikákról általában úgy szólnak, hogy túlságosan sokba kerülne az időigényes humán munka miatt, vagy pedig úgy kezelik, mintha elhanyagolhatók lennének az információ-visszakeresés új megoldásainak megjelenésével.

A természetes nyelvi feldolgozási technikákat, a kulcsszavas indexelést, és a kifinomult keresőprogramokat egyre gyakrabban említik az Internet elérési problémáira való potenciális megoldásként. Az AltaVista keresőrendszere jó példa a gyors keresésre. A gyors keresés azonban nem feltétlenül jelent jobb minőségű keresést is [11]. Nem mindenki van tehát meggyőződve arról, hogy a szoftver az egyetlen megoldás.

## A dokumentum fogalma a digitális könyvtárak korában

A digitális könyvtárak alapvető funkciója nem magának az Internetnek, hanem az Interneten megtalálható dokumentumoknak a katalogizálása. A katalogizálás szempontjából az Internet nem katalogizálási egység, mivel nem tekinthetjük egyetlen önálló könyvtárnak. Hagyományos értelemben az Interneten elérhető anyagok nem alkotnak könyvtári gyűjteményt, hiszen az Internet nem egy bizonyos felhasználói kör igényeinek a kielégítésére törekszik. Az Internet része lehet azonban a könyvtár műszaki infrastruktúrájának, de továbbra sem lesz könyvtár, abban az értelemben sem, ahogyan egy könyvtári intézmény hozzáférést biztosít egy gyűjteményhez [3].

Az elektronikus dokumentumok világában a könyvtárosok, az információs szakemberek éppúgy, mint a hagyományos nyomtatott dokumentumok kiadói, kissé zavarodottan szemlélik a technológiában, szerzői jogi kérdésekben bekövetkező radikális változásokat. Az egyik legalapvetőbb fogalom az információs szakemberek számára, ami egyértelműen átgondolásra szorul, a *dokumentum* fogalma. Az elektronikus források rend-

szerezésénél az első felmerülő kérdés az, hogy az elektronikus környezetben mit is értünk dokumentumon. A hagyományos fogalom olyan feltételezéseken alapul, amelyek nem veszik figyelembe az elektronikus formátum jellegzetességeit és használatát. Nem foglalkozik olyan kérdésekkel, hogy vajon a hipertextkapcsolatok (linkek) által egymáshoz kapcsolt dokumentumok egy dokumentumnak tekinthetők-e, az elektronikus üzenetek (e-mail) egy sorozata egy dokumentumot alkot-e, ki a szerelmi birtokosa a hiperlinkekkel egymáshoz kapcsolt dokumentumok bizonyos csoportjának.

Mi tekinthető tehát dokumentumnak? Nyilvánvaló, hogy az elektronikus dokumentumoknak vannak bizonyos egyértelmű jellegzetességei, amelyek megkülönböztetik őket a nyomtatott dokumentumoktól. Ilyen jellegzetességek például a könnyű manipulálhatóság, a külső és belső kapcsolódási pontok hozzáadásának lehetősége, az egyszerű transzformálhatóság, az azonnali adatátvitel, a korlátlan másolhatóság. Az elektronikus publikálásban tehát a legnyugtalanítóbb aspektus a változtathatóság. A bibliográfiai kontroll szempontjából a dokumentumok legfontosabb jellemzője azok „megfoghatatlansága,” átmeneti, képlékeny, folyton fejlődő, változó természete, egyszerre több felhasználó számára való elérhetősége, és az egyes felhasználók igényeihez való adaptálhatósága.

Az elektronikus dokumentumok természete a dokumentum fogalmának átgondolására késztet. Hogyan módosíthatjuk a dokumentum fogalmát úgy, hogy az pontosabban fejezze ki a jelenleg használatos valamennyi dokumentumtípus jellemzőit? L. Schamber a következő kategóriák átgondolását javasolja a dokumentum újradefiniálásával kapcsolatban [12]:

1. *Homogén dokumentum.* A dokumentumot hagyományosan egyetlen fizikai entitásként értelmezzük, amelyet egy vagy több szerző alkotott, és amely véges terjedelmű, állandó, lineáris tartalommal bír. Ezen típusú dokumentumok heterogén gyűjteményét általában tárgy szerinti csoportosításban vagy médium szerint tárolják.
2. *Hiperlinkekkel összekapcsolt heterogén dokumentum.* Az információs központok katalógusokon és indexrendszereken keresztül biztosítanak információ-hozzáférést. Ezek az eszközök lehetővé teszik a dokumentumrekordok valamely közös jellemző alapján való visszakeresését. Az így visszakeresett halmazt felfoghatjuk úgy, mint egy céltudatosan egyesített metaadatot. Jelenleg az információs központok gyakran használnak külön tárgy és médium szerinti katalógusokat. Ideális esetben egy elektronikus rendszer teljes mértékben képes integrálni a heterogén dokumentumokhoz való hozzáférést. Az összekapcsolt számítógépes

hálózatok világában a linkekkel összekapcsolt dokumentumrészek leírásának magába kell foglalnia a weboldalak elérési útvonatát.

3. *A kapcsolódó dokumentumok kontextuális megjelenítése.* A kifinomult elektronikus rendszerek vissza tudnak hívni olyan dokumentumhalmazokat, amelyekben az egyes rekordok a tárgyi hasonlóság mértéke alapján rangsorolva vannak. Az igazán kreatív tervezők kifejlesztettek olyan háromdimenziós megjelenítési formákat (piramisok, szobák stb.), amelyek segítenek elképzelni a rangsorolt halmazokat. Ebben az esetben az egész halmaz egyetlen metaadatot képvisel.
4. *A felhasználó által létrehozott heterogén dokumentum.* A hipertext világában a felhasználó összekapcsol és manipulál dokumentumokat vagy dokumentumok részeit, és így új dokumentumot hoz létre. Az új dokumentum továbbra is változtatható marad, és a végtelenségig vagy legalábbis mindaddig módosítható, amíg a felhasználó új kapcsolatokat (kapcsolódó linkeket) fedez fel. A dokumentumnak e dinamikus fogalma alapján a tartalom a felhasználó kontrollja alatt áll. A dokumentum tehát egyedi és nem állandó, és ezért valahogyan a dinamikus linkek segítségével kell leírni.

Mindezek az elképzelések olyan megközelítéssel alapulnak, amely a dokumentumokat mint alkotásokat/tárgyakat veszi figyelembe. Ha arra gondolunk, hogy a végső cél nem az alkotások/tárgyak, hanem az információ biztosítása a felhasználó számára, akkor a dokumentum definiálásakor egy folyamatorientált megközelítést célszerű használnunk. Természetesen a legfontosabb dolog nem az, hogy egyetlen definíciót alkossunk, amely kifejezi az elektronikus dokumentumok valamennyi jellemzőjét. A hangsúly a fogalom újragondolásán van. A bibliográfiai kontroll újragondolása pedig az olyan új fogalmak beépítését jelenti a mindennapi gyakorlatba, mint a hipertext, linkek, weboldalak, megjelenítési formátumok, verziók, navigációs lehetőségek stb.

## Az elektronikus dokumentumok

Egy Internet-katalógus létrehozásának egyik fő problémája az Internet-dokumentumok és -szolgáltatások tűnékeny természete. Az Internet-használók egyik állandó frusztrációja a nem érvényes linkek gyakorisága. Az Internetnek a jövőben valószínűleg nagyobb stabilitása lesz, jelenleg azonban még mindig átalakulóban van. J.-C. Guedon úgy véli, hogy az elektronikus kiadás esetében egy külön osztályozási rendszert kell kialakítani, amely megfelel az elektronikus dokumentumok jellemzőinek [13]. Ezek a tulajdonságok

pedig meglehetősen eltérnek a nyomtatott dokumentumokéitól. Az elektronikus kiadást inkább kommunikációs eszköznek fogja fel, nem pedig az információ állandóvá és széles körben való elérhetővé tételének.

Az Internet-dokumentumok flexibilitása és nyitottsága az interaktív párbeszéd és a módosíthatóság iránt egyértelműen pozitív dolgok, viszont a katalógizáló könyvtárosok szemében ezek inkább akadályok, mintsem előnyök. Woodward szerint az OCLC kissé naivnak tűnik, amikor azt feltételezi, hogy az Interneten információt szolgáltató intézményekkel való megegyezések révén az elektronikus dokumentumok leőhelye állandósítható [1]. Valójában egy központi katalógus szolgáltatójának rendszeresen ellenőriznie kell a rekordok érvényességét. Nagyon sok forrás vándorol egyik szerverről a másikra a weben való jelenléte során. A URN, a PURL vagy az Egységes Forrásazonosítók (Uniform Resource Identifier = URI) segíthetnek minimalizálni ezt a problémát, de az Internet-dokumentumok mozgása elkerülhetetlen. A dokumentum változása is igen gyakori. Módosítások, kiegészítések, törlések sűrűn előfordulnak, és felmerül a kérdés, hogy mekkora mértékű változás után tekinthető egy módosított forrás új dokumentumnak. Az elektronikus dokumentumok esetében érvénytelen az az állandóság, amit a nyomtatott dokumentumokkal asszociálunk. Ezeket a változásokat pedig tükrözni kell a katalógusrekordokban, ami azt jelenti, hogy valakinek fel kell vállalnia a változások követését [1].

Az online információs források legalább két. jól elkülöníthető kategóriába sorolhatók: elektronikus adatforrások, valamint online rendszerek/ szolgáltatások. Az első kategória az online vagy offline elérhető elektronikus forrásokat foglalja magába, amelyeket viszonylag könnyen le tudunk írni a jelenlegi MARC bibliográfiai formátumban. Ezeknél az anyagoknál felmerül a probléma, hogy hogyan határozzuk meg a „leőhelyüket,” amikor az adatok egy távoli rendszer vagy szolgáltatás révén érhetőek el, nem hagyományos könyvtáron vagy archívumon keresztül. A második kategória sokkal inkább a szolgáltatói közösség információs formátumához hasonlít, mintsem a bibliográfiai adatokhoz [14].

Az elektronikus adatforrások kategóriájába olyan dokumentumok tartoznak, mint a számítógépes szoftverek, géppel olvasható formátumban tárolt szövegek vagy képek, bibliográfiai, numerikus, vagy egyéb adatok adatbázisai. Ezek a dokumentumok lehetnek csak elektronikusak, de előfordul, hogy vannak nyomtatott megfelelőik is. A számítógépes rendszerek vagy szolgáltatások külön kategóriát képviselnek. Lehetnek teljesen nyilvánosak, de lehetnek csak egy bizonyos kö-

zösség számára elérhetőek. A fő jellemzőjük az, hogy többnyire csak hálózaton át érhetőek el.

## Metaadatok

A könyvtárak már régóta használják a katalógustételeket és a MARC rekordokat, illetve biztosítanak elérést a nyomtatott irodalomhoz. Mivel a digitális gyűjtemények rég túlnőttek a könyvtárak falain, ezek a jól ismert és kedvelt bibliográfiai eszközök túlságosan részletesek, ezért sokak szerint alkalmatlanok a digitális dokumentumok leírására. A legközismertebb kezdeményezés, amely új formátumot nyújt a digitális dokumentumok leírására, egy 1995 óta tartó nemzetközi konferenciasorozat eredményeképpen alakult ki, és „Dublin Core-metaadatok” néven vált ismertté [15].

Különböző típusú intézmények egyik legkurrensebb nemzetközi vitatémája az, hogy a kutatók hogyan tudják hatékonyabban megtalálni a számukra szükséges információkat az Internet- és a webalapú források használata során, illetve a metaadatok mennyire lehetnek hasznosak ebben a keresésben. Könyvtárosoknak, levéltárosoknak, múzeumi szakembereknek éppúgy, mint az információs szabványok, hálózatok szakértőinek a metaadat kifejezés különböző konnotációkat hordoz. Caplan definíciója szerint a metaadat nem más, mint adat az adatról [14]. Ennek megfelelően egy katalógusrekord is metaadat, és könyvtáros berkekben már régóta folyik a metaadatok létrehozása. Ezeket az adatokat azonban hagyományosan katalógusadatoknak nevezzük. A metaadatok létrehozását tehát továbbra is hívhatnánk katalógizálásnak, de ez a kifejezés sokak számára olyan jelentéseket hordoz, amelyek a hagyományos könyvtári szabványokhoz és gyakorlathoz kapcsolódnak, így ez az elnevezés régimódinak tűnik. Ezért az Internet világában a semlegesebb és újabb „metaadat” kifejezést preferálják. Részben tehát a MARC formátum használata is metaadattal való foglalkozást jelent, manapság azonban egyre inkább a digitális fájlok fejlécében (header) található deskriptív információ azonosítását értik ezen.

S. *Weibel*, akit a Dublin Core kitalálójaként szoktak emlegetni, a metaadatokat egyszerű leírási formátumként definiálja, amelyeknek a célja a weben való keresés javítása nemzetközi megegyezés segítségével [16]. Nagyon sok eltérő adatmodellel rendelkező adatbázis létezik, de bizonyos fokig valamennyi leírható a Dublin Core-nak nevezett közös adatmodellel. A Dublin Core fő célja, hogy a *tartalom* és a tartalmat leíró *szintaktikai kifejezés* közti függetlenséget fenntartsa. Korábban – a katalógizálás során – a könyvtárosok valamennyire megengedték a szintaktikai kifejezé-

sek összevegyítését az adattartalom-modellel. A Dublin Core esetében azonban e két dolog elkülönítésére most ők is figyeltek.

Eredetileg a metaadatok létrehozása egyetemi kutatók ötlete volt, de azóta széles körű gyakorlati témává vált, ami a webarchitektúrával foglalkozók körében jelenleg a legnagyobb prioritású. Egy ilyen szabvány létrehozásakor felmerül a kérdés, hogy a metaadatok problémája elsősorban a kutatói környezetre korlátozódik-e. Ha figyelembe vesszük, hogy az üzleti világban mindenkinek képesnek kell lennie saját metaadathalmazának létrehozására más metaadathalmazok ismerete nélkül, hiszen az erre irányuló együttműködés időigényes és drága lenne, akkor a válasz egyértelműen nem. Ezenkívül a kódot, amely a Dublin Core használatát teszi lehetővé, nem egyetemi kutatók, hanem olyan cégek fogják beépíteni a keresőkbe (browser) és a szerverekbe, mint például a Microsoft és a Netscape. A könyvtárosok szerepe az információ-rendszerkezés terén a legfontosabb.

A metaadatoknak nemcsak az a funkciójuk, hogy a web indexeiben való keresést megkönnyítsék, hanem az is, hogy a hagyományos dokumentumok metaadatait is megjelenítsék a weben. A felhasználók ugyanis hajlamosak azt hinni, hogy ami nincs a weben, az nem is létezik, ezért emlékeztetnünk kell őket arra, hogy hatalmas mennyiségű kitűnő információforrás létezik a weben kívül is. Ebből a felismerésből kiindulva az 1997 nyarán megalakult Metadata Working Group javaslatot is tett az egyes Dublin Core-definíciók olyan jellegű módosítására, amely a Dublin Core-elemek használatát kiterjeszti más információ típusokra is.

1997 júliusában a Research Library Group találkozójának az volt a fő témája, hogy a Dublin Core elemei hogyan használhatók olyan anyagok elérésére, amelyek nem a weben keresztül érhetők el. A kérdés az, hogy milyen típusú metaadatokat kell az olyan különböző típusú forrásokhoz kapcsolni, mint a hagyományos könyvtári anyagok, online könyvtári katalógusok, elektronikus szövegek és múzeumi tárgyak. Jelenleg az a fő probléma ezekkel a forrásokkal, hogy a weben keresztül ugyan elérhető bizonyos információ róluk, de a dokumentumok nem webalapúak, és gyakran csak önálló, inkompatibilis indexelő és kereső protokollokon keresztül érhetők el.

## A Dublin Core-elemek

A Dublin Core-elemek, azaz a Dublinban 1996-ban született megegyezés azon 15 alap információs elem – más néven metaadat-deszkriptor – kialakításáról, amelyet a weboldalakon való indexelés és kivonatképzés során használnak a keresési eredmények javítása céljából, igazi mérföld-

kőnek számít. A Dublin Core már de facto szabványként használatos, és remélhetőleg (a mostani érdeklődésből kiindulva) hamarosan formális szabvánnyá is válik. Jelenleg az egyik legkurrensebb megoldásra váró feladata az Internet-dokumentumok leírásainak nemzetközivé tétele. A web jelenlegi infrastruktúrájában több helyen is előfordulnak különböző nyelvekhez kapcsolódó, egymástól eltérő karakterhalmazok, amelyeknek a kezelésére ugyan jelenleg még nincs megoldás, de 16 ország adateírói folyamatosan dolgoznak egy nemzetközi Dublin Core-prototípus kialakításán. Eszerint van esély arra, hogy egy nemzetközileg egységes megoldás szülessen a hálózaton elérhető források katalogizálására.

A Dublin Core-metaadatelemek jelenlegi verzióját 1996 decemberében véglegesítették. Annak ellenére, hogy a Dublin Core folyamatos fejlesztés alatt áll, és az érdekelt területek szakértői rendszeres találkozókön vizsgálják felül, valószínű, hogy a közeljövőben az adatelemek és definíciójuk már nem változnak meg alapvetően. Néhány adatelem alkalmazása ugyan jelenleg még kísérleti szinten folyik, és értelmezése implementációként változhat. Az elemek elnevezései szemantikai értelmezésüket próbálják tükrözni. Minden elemhez kapcsolódik egy egyszavas név is, amely az elemek szintaktikai meghatározását teszi egyszerűbbé a kódolási rendszerek számára. Valamennyi elem opcionális és ismételhető [16].

**Cím (cim):** A forrás létrehozója vagy kiadója által meghatározott neve.

**Szerző vagy létrehozó (létrehozó):** Az(ok) a személy(ek) vagy szervezet(ek), amely(ek) elsődlegesen felelős(ek) a forrás szellemi tartalmának létrehozásáért. Az írott dokumentumok esetében például a szerzők, vizuális források esetében a művészek, fényképészek vagy illusztrátorok.

**Tárgy- és kulcsszavak (tárgy):** A forrás témája. Általában a dokumentum tárgyát kulcsszavak fejezik ki. A kontrollált szótárak és formális osztályozási rendszerek használata preferált.

**Leírás (leírás):** A forrás tartalmának szöveges leírása, amely dokumentum típusú tárgyak esetében a ki vonatot, vizuális források esetében a tartalmi leírást jelenti.

**Kiadó (kiadó):** Az(ok) a személy(ek) vagy szervezet(ek), amely(ek) a forrás jelen formájában való közre adásáért felelős(ek).

**Egyéb közreműködő (közreműködő):** Az(ok) a létrehozóként nem említett személy(ek) vagy szervezet(ek), amely(ek) jelentősen hozzájárult(ak) a forrás szellemi tartalmához, de közreműködése(iük) másodlagos a létrehozóként megnevezett személy(ek) vagy szervezet(ek) szerepéhez képest. Példa: szerkesztő, fordító, illusztrátor stb.

**Dátum (dátum):** Az az időpont, amikor a forrás a jelenlegi formájában elérhetővé vált. A dátum ajánlott formátuma a következő: év (4 karakter) – hónap (2 karakter) – nap (2 karakter). Sok egyéb formátum le-

hetséges. Használatuk esetén az egyértelműséget biztosítani kell.

**Forrástípus (típus):** A forrás kategóriája, mint például „weblap,” „regény,” „vers,” „tanulmány,” „szótár” stb. Az egységesség és a rendszerek közti átjárhatóság érdekében a típust egy megadott listából kell kiválasztani. Jelenleg ez a lista még nem végleges. Legújabb verziója a <<http://sunsite.berkeley.edu/Metadata/types.html>> URL címen érhető el.

**Formátum (formátum):** A forrás adatformátuma. Azonosítja a szoftvert, és optimális esetben a hardvert is, amely a forrás megjelenítéséhez vagy működtetéséhez szükséges. A forrástípushoz hasonlóan – az egységesség érdekében – a formátum típusát is egy előre összeállított listából kell kiválasztani.

**Forrásazonosító (azonosító):** Olyan betű- vagy számkombináció, amely egyértelműen azonosítja a forrást. A hálózati források esetében ilyen azonosító lehet a URL vagy a URN. Nem online források esetében ez lehet a világszerte használt azonosító, az ISBN, vagy egyéb formális név.

**Forrás (forrás):** Olyan betű- vagy számkombináció, amely egyértelműen azonosítja azt a művet, amelyből a forrás származik. Például egy regény PDF verziója forrásként tartalmazhatja annak a fizikai könyvnek az ISBN számát, amelyből a PDF verzió származik.

**Nyelv (nyelv):** A forrás szellemi tartalmának nyelve(i).

**Kapcsolat (kapcsolat):** Az adott forrás kapcsolata más forrásokkal. Ezen adatelem célja, hogy lehetőséget biztosítson az egyébként önálló források közti formális kapcsolatok jelölésére. Például egy dokumentum képei, egy könyv fejezetei, vagy egy gyűjtemény egységei között. Ezen adatelem használata és specifikációinak kialakítása jelenleg kísérleti fázisban van.

**Lefedett téma (téma):** A forrás helyi és/vagy időbeli jellemzői. Ezen adatelem használata és specifikációinak kialakítása jelenleg kísérleti fázisban van.

**Szerzői jog (jog):** Link a szerzői jogi megjegyzéshez, vagy ahhoz a szolgáltatóhoz, ahol a forráshoz való hozzáférésre vonatkozó információ érhető el. Ezen adatelem használata és specifikációinak kialakítása jelenleg kísérleti fázisban van.

## A metaadatok jelentősége

A katalógusrekordok előállításának költségei jelentősen csökkenthetők, ha a MARC rekordoknak legalább egy része automatikusan előállítható egy szoftver használatával, amely ki tud választani bizonyos adatelemeket a dokumentumokból, és tárolni tudja őket a megfelelő MARC mezőkben. Ezek a programok a MARC rekordok létrehozásához már különböző szabványokon (például TEI = Text Encoding Initiative vagy URC = Uniform Resource Citation) alapuló metaadatokat használnak [1].

A Spectrum, az InterCat projekt részeként kifejlesztett szoftver például olyan szoftverrendszer, amely el tudja végezni ezeket a feladatokat, és így kevésbé költséges Internet adatbázist hoz létre. Ez

a költségmegtakarítás lehetővé teszi az OCLC számára, hogy mindenki számára elérhető, nyilvános katalógust működtessen az Interneten, valamint az InterCat projektben részt vevő intézmények számára egy sokkal kifinomultabb OPAC-elérést biztosítson. Az egyik legérdekesebb dolog a Spectrummal kapcsolatban az, hogy különböző formátumú adatokból képes MARC rekordokat létrehozni.

A TEI fontos nemzetközi projekt, amelyet az Association for Computers and the Humanities (ACH), az Association for Computational Linguistics (ACL) és az Association for Literary and Linguistic Computing (ALLC) támogat. Feladata irányvonalak kifejlesztése és terjesztése a géppel olvasható szövegek kódolására, közvetíthezőségére és cserélhetőségére, valamint javaslatok tétele új szövegek kódolására. A TEI javaslata a szövegek kódolására a Standard Generalized Markup Language (SGML) metanyelv. Az SGML nemzetközi szabvány az elektronikus szövegek kódolására, amely lehetővé teszi, hogy sok különböző szöveg egymással kompatibilis formátumban legyen kódolva, és elemezhető legyen az SGML-t támogató szoftverekkel. Ez a kódolási formátum az elektronikus szöveghez jól használható, de még nem MARC-kompatibilis. A Spectrum lehetővé teszi a MARC rekordok, TEI fejlécek, vagy a URC-k egy adatbázisba való összegyűjtését, és hagyományos technikákkal való kereshetőségét. A Spectrum három fontos folyamatot tud elvégezni: rekordok létrehozását, adatbázis-építést és rekordvisszakeresést. A bibliográfiai rekordok konvertálása TEI vagy URC formátumból MARC formátumba a legjelentősebb funkció.

## Metaadatszabványok

Az inkompatibilis szabványok és technológiák nagy problémát jelentenek az Interneten elérhető dokumentumok katalogizálásában. A MARC szabvány igazán kifinomult eszköz a könyvtári források leírására, de a könyvtárakon kívül más intézmények nem nagyon használják. Annak érdekében, hogy a MARC elfogadhatóbb legyen a tágabb közönség számára, az OCLC és a National Center for Supercomputing Applications (NCSA) egy metaadat-workshopot kezdeményezett, amely különböző információszolgáltatókat (például kiadók, szoftverfejlesztők és kutatók) is bevont a beszélgetésekbe. A workshop eredményeképpen megszületett a Dublin Core, amely a USMARC egyszerűsített változatának tekinthető.

A Harvest nevű másik metaadat-szerkezetet a Colorado Egyetemen fejlesztették ki az Advanced Research Projects Agency (ARPA) támogatásával. A Harvestnek olyan eszközei vannak, amelyek

megkönnyítik az információ megtalálását a fájlokban, és automatikusan tudnak indexet építeni. Saját keresőprogramját és protokollját felhasználva különböző felhasználói csoportok számára készít katalógusokat. Annak ellenére, hogy nem MARC-alapú, közös platformként szolgálhat a hálózati adatbázisok között. A Harvest tartalmi összefoglalást is készít minden általa összegyűjtött információs tárgyról. A rekordokat Summary Object Interchange Format (SOIF) formátumban tárolják. A tartalmi összefoglalások bizonyos mértékig megfelelnek a MARC rekordokban alkalmazottaknak. Neuss és Kent a bibliográfiai kontroll hagyományos módszereinek a Harvesttel való kombinációján dolgoznak. A könyvtártudomány és a fogalmi elemzés ötleteinek kombinációjával a hálózati információs források, különösen a World Wide Web fogalmi elemzésére fejlesztenek ki eszközöket. A hálózati információs források szövegfájlokat, WAIS adatbázisokat és webdokumentumokat foglalnak magukba. Neuss és Kent szerint ezeket a forrásokat célszerűbb inkább fogalmi osztályoknak, mintsem tárgyakként tekinteni, ezért a hálózati információs források rendszerezésére és leírására fogalomorientált megközelítést alkalmaznak [1].

A TEI által létrehozott szabvány, a TEI P3 a humán tudományok szövegeinek digitális kódolására irányul, és egy elektronikus fejléc létrehozását javasolja, amely minden egyes dokumentum esetén a címdalhoz hasonló információkat tartalmaz. Steve Kirsch, az InfoSeek keresőszolgáltatás kifejlesztője – más keresőprogramok szolgáltatóival együttműködve – szintén folyamatosan dolgozik egy metaadatszabvány kifejlesztésén. Ez a szabvány nem MARC-alapú, de mint a Z39.50 egyik kiterjesztése támogatja a MARC formátumot. A Working Group on Document Identifiers, amely az Internet Engineering Task Force része, az Egységes Forrásjellemezők (URC = Uniform Resource Characteristic) szabványosításával foglalkozik.

A fent említett szabványosítási kísérletek mellett egyre nagyobb szükség van olyan szabványokra, amelyek a különböző rendszerek közti átjárhatóságot biztosítják. Az elektronikus források számának növekedésével együtt súlyosbodik a különböző rendszerekben tárolt információ elérhetőségének problémája. Ez a probléma nemcsak a felhasználók, hanem az eltérő adatbázis-kezelő rendszerek és az elektronikus információcsere szintjén is jelentkezik. A Z39.50 amerikai nemzeti szabvány egy új lépés a formátumok nagyobb kompatibilitása felé. A Z39.50 feladata azon üzenetek kódolása, amelyek két számítógépes rendszer közti kommunikációhoz szükségesek az információkeresés és visszahívás során. A szabvány eredetileg ugyan könyvtári célokra lett kifejlesztve, de teljesen független a közvetített információ típusától. Jelenleg több egyetemi intézmény

és információszolgáltató használja. A Kongresszusi Könyvtár, az OCLC és a Research Library Information Network (RLIN) például bibliográfiai rekordok cseréjére alkalmazza, de a szabvány alkalmas az Interneten való használatra is.

## Metaadatok és az elsődleges források

Az olyan elsődleges forrásokkal foglalkozó intézmények, mint az irattárak, speciális gyűjteményekkel, kéziratokkal, ritka könyvekkel, múzeumi tárgyakkal, művészeti alkotásokkal, vagy egyéb ritka és egyedi anyagokkal foglalkozók, szintén erősen érdeklődnek a metaadatok felhasználása iránt, különös tekintettel azok lehetséges használatára az egyedi anyagok weben való elérhetővé tételében. 1997 őszén a Research Libraries Groupon belül külön archívumi metaadatszoport alakult azzal a céllal, hogy megállapítsa ennek lehetőségeit. A csoport egyik feladata, hogy megvizsgálja a már létező metaadatelemeket, és megállapítsa, hogy azokon kívül milyen más elemekre van még szükség az archívumok számára, valamint hogy megállapodásra jusson a különböző nemzetközi metaadatszabványokról.

A metaadatokkal kapcsolatos viták tehát nem korlátozódhatnak nemzeti szintre, hiszen az egyes nemzeti kulturális örökséget reprezentáló anyagok a nemzetközi közösség részei. Ebből a felismerésből kiindulva, valamint szem előtt tartva azt, hogy csak a széles körű érdekek figyelembevételével érhetőek el a legjobb eredmények, a Canadian Heritage Information Network (CHIN) például a metaadatok nemzetközi felhasználásával foglalkozik a szöveges és képi adatok összekapcsolásában. A CHIN projektjének célja, hogy létrehozza a kanadai múzeumok gyűjteményeinek nemzeti online elérésű leltárát. Jelenleg ez még csak egy szöveges gyűjtemény, amely több mint 25 millió szépművészeti, irodalmi, természettudományi és egyéb tárgyat reprezentál. A legújabb cél viszont az, hogy a szöveges adatokhoz tartozó képeket is elérhetővé tegyék. Ehhez azonban új adatelemekre van szükség, ami új problémákat vet fel a források leírásában.

A metaadatokkal illetően felmerül a probléma, hogy vajon létrehozható-e olyan univerzális elemhalmaz, amely valamennyi információforrás-típusra alkalmazható, és a kereső számára általános áttekintést ad az információról, függetlenül annak típusától. Az információforrások leírása azonban nem az egyetlen probléma. A hagyománhoz hasonlóan a metaadatoknak több szintje létezik a kereső és a forrás között. Ezzel kapcsolatban többen aggodalmukat fejezik ki, vajon a felhasználóknak hogyan lehet egyértelműen jelezni, hogy adott helyzetben az információnak melyik szintjével ta-



lálkoznak. Ez a szint lehet a leíró rekordok teljes adatbázisa, az adatbázis egyetlen rekordja, vagy pedig egy rekord által leírt dokumentum képe. Ahogyan Lyn Elliot Sherwood megfogalmazta: „Meg kell találni a módját annak, hogy hogyan lehet nagy katalógusokat a hálózati rendszerben elérhetővé tenni, és célszerű eldönteni, hogy milyen szinten legyenek azok láthatók. Vajon érdemes-e katalógustételek millióit megjeleníteni a weben való keresés első fázisában? Valószínűleg nem. Mindannyian belefulladnánk az adatokba. A kérdés tehát az, hogyan segítsük az iteratív keresési folyamatot” [17].

További megoldásra váró probléma, hogy néhány, a web számára kialakított Dublin Core-elem kevésbé hasznos az olyan elsődleges források leírásánál, amelyeket a weben azok digitális reprodukciói képviselnek. A nem webdokumentumokat illetően a két legproblémásabb Dublin Core-elem a *dátum* és a *kiadó*. Ezek olyan specifikusan vannak definiálva, hogy használatuk a digitális szurrogátumok leírásánál zavaró lehet. A *dátum* adatelem definíciója a következő: „Az a dátum, amikor a forrás a jelenlegi formájában elérhetővé vált.” Ugyanakkor, ha valaki például *Van Gogh* egyik festményét keresi a weben, akkor nagy valószínűséggel inkább az fogja érdekelni, hogy a festmény mikor készült, nem pedig az, hogy mikor lett digitalizálva. Hasonló a helyzet a *kiadó* esetében is, amit úgy definiáltak, hogy „az a személy/testület, amely az információforrást a jelenlegi állapotában létrehozta”. El kell gondolkoznunk azon, hogy a metaadat létrehozásakor hogyan lehet különbséget tenni egy műalkotás digitalizált szurrogátuma és az alkotás között. Vajon a *forrás* információ magára az alkotásra vonatkozik, vagy pedig annak a fotójára? A *dátum* az alkotás elkészültének a dátuma, vagy pedig azt jelöli, hogy a fotót mikor készítették? A *létrehozó* a művész vagy a fényképész? A Dublin Core ma még nem alkalmas a következőes alkalmazásra, és a használata problematikusvá válik, ha nem eredetileg elektronikus módon létrehozott információval, vagy ha műalkotások elektronikus reprezentációival van dolgunk [17].

## Rendszerezési törekvések az Interneten

Jelenleg nagyon sok tevékenység irányul az Internet rendszerezésére. Meglehetősen nagyszabású katalógusok és indexek vannak használatban már évek óta. A Yahoo! vagy az Alex csak két példa erre. Folyamatos kísérletek folynak a digitális dokumentumok katalogizálására. A legismertebb példák erre a British Library által támogatott

CATRIONA projekt, valamint az OCLC Internet-források katalogizálásának programja, az InterCat.

Új szabványok jelennek meg, és a már meglévők módosulnak, amelyek a digitális dokumentumokat, és a róluk készített metaadatokat reprezentálják. Egyre több az Internet katalogizálásával foglalkozó konferencia és egyéb szakmai találkozó. Cikkek jelennek meg ebben a témában az Interneten éppúgy, mint a nyomtatott időszaki kiadványokban. Sok Internet-listán (Autocat, Pacs-I, Emedia vagy Intercat) folyik a vita katalogizálási témákban. 1996 óta a Haworth Press, Inc. kiadásában nemzetközi elektronikus folyóirat jelenik meg az Interneten *Journal of Internet Cataloging* címmel (<http://jic.libraries.psu.edu>).

Az Internet megjelenésével a könyvtáraknak már nem elegendő az Internet-dokumentumok csak bibliográfiai adatainak elérése, hanem a teljes szövegekhez való hozzáférés biztosítására is törekedniük kell. A felhasználók eligazítása a nem helyileg tárolt információk világában fokozatosan átalakítja a katalógusról kialakult képet. A hálózati információk elterjedésével olyan eszközre van szükség, amely egyaránt tudja kezelni mind a helyi, mind pedig a hálózaton tárolt információkat, szöveget és képeket, és hozzáférést biztosít adatokhoz, információkhoz és információtároló eszközökhöz. Összehasonlítva a hagyományos katalógust a digitális könyvtári katalógussal, a következő alapvető különbségeket állapíthatjuk meg. A hagyományos katalógus segítségével elérhető a helyileg tárolt anyagok, amelyek elsősorban tárgyi dokumentumok. A hozzáférést lelőhelyekre mutató információ formájában biztosítja. A hagyományos katalógus önálló bibliográfiai eszköz. Ezzel szemben az új, hálózati forrásokat is számba vevő katalógus egyaránt biztosít elérést mind a helyileg, mind pedig a máshol (valahol a hálózaton) tárolt anyagokhoz. Nemcsak az információt tartalmazó tárgyak bibliográfiai adataihoz, hanem az elektronikus információkhoz is hozzáférést biztosít. Magát az elektronikus információt szolgáltatja, és nem csak annak lelőhelyadatát.

A felhasználó közös felhasználói felülettel találkozik, amely többféle információ keresésére alkalmas. A könyvtárosoknak minden okuk megvan arra, hogy nagyra értékeljék az információ megjelenítés következetességét, hiszen a felhasználóknak kevesebb előzetes tájékoztatásra van szükségük, és gyorsabban tudnak mozogni a különböző források között. Mindenképpen grafikus felhasználói felületre van szükség, amelyet a nyilvános elérésű számítógépes rendszer (Public Access Computer System) irányít, és amelynek egyik összetevője az online elérésű katalógus keresőprogramja. A rekordok definíciója kibővül a máshol elérhető fizikai dokumentumok, és a helyileg vagy máshol elérhető elektronikus dokumentumok in-

formációival [18]. A dokumentum fogalmát itt tágan értelmezzük. Azt a feljegyzett információt értjük dokumentumon, amely lehet elektronikus vagy nyomtatott: könyv, cikk, vagy egyéb szöveg; grafikus vagy audioinformáció. A felhasználó szempontjából a dokumentumok két csoportba sorolhatók: azok a dokumentumok, amelyek közvetlenül elérhetők lesznek a kapcsolódó linkek segítségével, valamint azok a dokumentumok, amelyekhez a katalógus csak elérést biztosít. A felsorolt különbségek ellenére a katalógus fogalma olyan értelemben nem változik meg, hogy továbbra is rendszerezett információt szolgáltat; egységesített neveket, címeket, tárgyszavakat, és következetes utalórendszert alkalmaz; valamint továbbra is prekoordinált és következetes marad. Mindezek pedig hozzáértő könyvtárosok munkáját követelik meg.

### Az Internet forrásainak rendszerezése: katalógizálói megközelítés

Miért katalógizáljuk az Internet-dokumentumokat? Egyrészt azért, mert számos értékes információ érhető el az Interneten. Másrészt azért, mert ezeket a forrásokat valahogyan rendszerezni kell ahhoz, hogy hozzáférhetőek legyenek. Valamint azért, mert a már létező könyvtári technikák és munkafolyamatok felhasználása, és visszakeresésre alkalmas rekordok létrehozása a már létező online katalógusok számára a leghatékonyabb e források elérésére.

A következő kérdés az, hogy mit katalógizáljunk. Vannak, akik az adott könyvtár vagy felhasználói közösség számára hasznos weboldalak katalógizálását javasolják az Interneten lévő anyagok elérhetővé tétele első lépéseként. Mások a projektben részt vevő intézmények saját elektronikus anyagainak a katalógizálását látják hasznosabbnak. A Missouri-Columbia Egyetem könyvtárának belső használatra kialakított katalógizálóprioritásai az Internet-források rendszerezését és katalógizálását illetően például a következők: a helyi rendszerben tárolt és fenntartott Internet-források; a helyi felhasználók számára fontos kutatási anyagok; azok az eszközök, amelyek fejleszthetik a tájékoztató szolgáltatásokat; és végül azok az eszközök, amelyek a könyvtári személyzet készségeit és tudását fejlesztik.

Valószínűleg a számítógépes fájlok katalógizálása az egyik legproblematisabb terület. Ez a médium nagyon gyorsan változik, fejlődik. Mire megismerkedünk az éppen használatban levő technológia terminológiájával és jellemzőivel, addigra a legújabb fejlesztések kiszorítják őket. Az ISBD szabványok számítógépes fájlokra második

kiadása, amelynek a jelenlegi címe „ISBD elektronikus forrásokra” – International Standard Bibliographic Description for Electronic Resources, ISBD (ER) –, 1994 óta folyamatos módosítás alatt áll [19].

Általános alapelv azonban, hogy bármit is katalógizálunk, a katalógizálóknak figyelembe kell venniük a nemzeti és nemzetközi szabványokat. Ha helyi okok miatt bizonyos külön terminológiára, leírásra, megjegyzésekre stb. van szükség, akkor a rekordokat a helyi gyakorlatnak megfelelően ki lehet egészíteni. Az Egyesült Államokban például az AACR2 által előírt katalógizálási szabályok adnak általános útmutatást. Ezek a szabályok bármilyen típusú dokumentumra alkalmazhatók. Az általános szabályok és irányelvek áttekintése mindig hasznos lehet, ha valamilyen új problémával állunk szemben.

### CATRIONA

Több kutatási projekt témája az Internet forrásainak katalógizálása. Erre a két legnagyobb és legújabb példát a CATRIONA [8], amely Nagy-Britanniában indult, valamint az OCLC InterCat projektje [20], amely az Egyesült Államokban folyik. Mindkét projekt célja az Internet dokumentumait számba vevő katalógus létrehozása, amely hasonlít a könyvtári nyilvános online katalógusok helyi adatbázisaira, egyben kompatibilis velük. Mindkét projekt a MARBI által kezdeményezett Elektronikus Hely és Elérésre kijelölt 856-os mezőt használja [21]. (A MARBI az Amerikai Könyvtáros Egyesület [ALA] azon testülete, amely a Kongresszusi Könyvtár tanácsadó szerveként működik a USMARC formátumot érintő kiegészítésekben és módosításokban.)

A CATRIONA (Cataloging and Retrieval of Information Over Networks Applications) projekt a BUBL (Bulletin Board for Libraries) Subject Tree Initiative kezdeményezésére jött létre 1993-ban, és alig egy évvel később már több mint 8000 Internet-linket tartalmazott. A projekt fő támogatója a British Library Research and Development Department. Célja azon technikai, szervezeti és pénzügyi követelmények vizsgálata, amelyek hálózatokon át közvetített dokumentumok és egyéb források katalógizálására, osztályozására és visszakeresésére szolgáló alkalmazási programokhoz és folyamatokhoz szükségesek. A projekt egy könyvtári rendszer használhatóságát is vizsgálja. Az OCLC-vel ellentétben CATRIONA az Internet-források olyan katalógusának létrehozására törekszik, amelyet a Z39.50 szabványon alapuló könyvtári OPAC felület fog össze, nem pedig egy központi adatbázis. A projekt egy Z39.50 GUI OPAC klienst választott ki, amely képes a távoli Z39.50 OPAC-okban való keresésre, és olyan MARC rekordok visszahívásá-

ra, amelyek a 856-os mező \$u almezőjében URL információt tárolnak [21], és az elektronikus dokumentumok megjelenítésére automatikusan betöltik a megfelelő megjelenítő programot (Mosaic, Netscape, Geac image viewer).

### **InterCat**

Az InterCat az Egyesült Államok nemzeti szinten koordinált projektje a felsőoktatási intézmények és könyvtárak között olyan USMARC formátumú rekordokból álló adatbázis létrehozására, bevezetésére, tesztelésére és értékelésére, amely az Interneten elérhető anyagok esetében kiegészül az elektronikus tárolásra és elérésre vonatkozó információval (USMARC 856-os mező). A projekt azt vizsgálja, hogyan lehet a könyvtári szolgáltatásokat kiterjeszteni az Internet forrásaira, azaz hogyan lehet olyan szintű hozzáférést biztosítani az Internet-forrásokhoz, amelyet egy könyv esetében elvárunk a könyvtártól [22]. A projekt főkoordinátora, Eric Jul hangsúlyozza, hogy az InterCat csak egy lehetséges módszer az Interneten elérhető dokumentumok keresésére, amely az információt keresők számára viszonylag egyszerű keresést biztosít. Jelenleg azonban ez a legnagyobb szabású projekt az Internet-dokumentumok elérhetővé tételére. A projekt elindítása előtt az OCLC tanulmányozta az elektronikus információ tulajdonságait, és szisztematikus elérésükkel kapcsolatos problémáit. A projekt céljai a következők:

- lokalizálni és beazonosítani az Interneten elérhető elektronikus információk típusait;
- kialakítani ezen információk taxonómiáját;
- összegezni azokat a problémákat, amelyekkel a könyvtáraknak meg kell küzdeniük ezen információk beszerzése, katalogizálása, indexelése, tárolása, visszakeresése és továbbítása során.

Az elektronikus irodalom értékének megállapítására szolgáló módszerek kialakítása után a projekt célja annak megállapítása, hogy ez a folyamat mennyire automatizálható. Végül a projekt javaslatokat tesz az ilyen típusú információk elérési módszereire és szintjeire.

Az InterCat projekt részeként az OCLC azt is vállalta, hogy megvizsgálja a bibliográfiai rekordon belüli kódolt információ alkalmazhatóságát a közvetlen felhasználói elérés biztosítására. A 856-os USMARC mezőt a Uniform Resource Locator információjának tárolására hozták létre [21]. A katalógusban megtalált URL címet a felhasználók valószínűleg nem szívesen másolnák le és gépelnek be újra, ezzel lehetőséget adva a gépelési hibákra. A tapasztalt Internet-használók hiperlinkek használatát várják el, amelyek a katalógustól közvetlenül a kívánt dokumentumhoz vezetnek. Mind a CATRIONA, mind az OCLC projekt az OPAC-ok kifinomult visszakeresési módszereinek, a MARC

rekordszabványinak, és a hipertext kapcsolatainak a kombinációját képzelel el. Az OCLC, a részt vevő könyvtárak, és egyéb felsőoktatási intézmények által létrehozott rekordokból egyetlen nagy adatbázist épít, amely közvetlen elérést biztosít az Interneten tárolt dokumentumokhoz. A projekt fő célja, hogy megállapítsa a USMARC formátum lehetséges használatát az Interneten elérhető dokumentumokhoz való hozzáférése során, azaz teszteli a 856-os MARC mező használhatóságát a nem helyileg tárolt dokumentumok elérésében (OCLC).

### **Alcuin**

A North Carolina State University Libraries Alcuin projektje [23] szintén a 856-os MARC mező használatán alapul [21]. Már 1994-ben kísérleteztek a 856-os mező használatával, amikor a helyi OPAC adatbázisába beépítették a URL információt. A projektet az teszi azonban igazán érdekessé, hogy az Alcuin fejlesztői a gopher elérésre kialakított Internet-címek „dBaselV” adatbázisát, amit Alexnek hívtak, olcsón, és kevés kiegészítő input hozzáadásával konvertálták egy új, MARC-alapú OPAC adatbázissá, amelyet Alcuinnak neveztek el.

Az Alcuin weboldalai webfelhasználói felülettel rendelkező katalógusokat tartanak nyilván. A projekt nem törekszik teljességre. Elsődleges célja az, hogy segítséget nyújtson azoknak a könyvtárosoknak, akik egy World Wide Web-alapú közvetlen elérésű online katalógus létrehozásának a lehetőségei iránt érdeklődnek. Másik célja annak megállapítása, hogy a könyvtárosság hagyományos gyakorlata hogyan alkalmazható az Internet-források gyűjtésére, rendszerezésére, osztályozására és terjesztésére.

### **Catholic University of America**

J. Hsieh-Yee, a Catholic University of America (Washington D.C.), School of Library and Information Science professzora, 160 Internet-forrás katalogizálása után arra a következtetésre jutott, hogy a források teljes leírása nagyon időigényes, és több, az AACR2R szerint definiált adatelem fölösleges a keresők számára [24]. Úgy véli, hogy a jelenlegi szabványok alapján egy eltérő mélységű leírási formát kell kialakítani, ami tükrözi az Internet dinamikus jellegét. A rekordok létrehozásának sebessége és a rekordok minősége, valamint a rekordkészítés sebessége és az Internet-források gyors hozzáférése közti egyensúly érdekében Hsieh-Yee egy minimális leírást biztosító szabványra tett javaslatot. E módosított szabvány célja, hogy a katalógus kereső, értékelő, rendszerező és dokumentumlokalizáló funkcióit biztosítsa csak azoknak az adatelemeknek a felhasználásá-

val, amelyek az Internet-források azonosításához és tárgyi rendszerezéséhez a legszükségesebbek. Ez az M szintűnek nevezett katalogizálás továbbra is az AACR2R adatcsoport elválasztójeleit és adatcsoportjait használja, de az adatelemek közül több is egyszerűsített változatban szerepel.

### **A Missouri Columbia Egyetem**

A Missouri Columbia Egyetem Jogi Karának Könyvtárában a következő irányelveket fejlesztették ki az Internet forrásainak katalogizálására. Az Internet forrásai két csoportba oszthatók: (1) rendszerek és szolgáltatások, beleértve a direktorykat és aldirektorykat, amelyek egynél több bibliográfiai egységet tartalmaznak; (2) önálló direktoryk, aldirektoryk és fájlok, amelyek egy bibliográfiai egység egyetlen példányát tartalmazzák. A könyvtár minden (1) kategóriába tartozó dokumentumot nyitott monográfiaként katalogizál. A (2) kategóriába tartozó dokumentumokat sorozatként katalogizálják, ha azok elektronikus sorozatok, és az egyes dokumentumok tartalmazzák információt arról, hogy a jövőben új változatok jelennek meg (pl. ver. 1.0). Ettől eltérő esetben ez utóbbi dokumentumokat is monográfiaként írják le. Ha az elektronikus források lelőhelye gyakran változik, akkor csak a rendszert vagy szolgáltatást katalogizálják, a 859-es mezőben jelezve a direktorykat és/vagy a fájlokat. Az elektronikus dokumentumok valamennyi verzióját (HTML, ASCII) ugyanabban a rekordban írják le, külön 856-os mezőt használva minden egyes verzióra [21]. Ha a rendszert és az önálló direktoryt vagy fájlt egyaránt katalogizálják, akkor használják az 580-as és a 773-as mezőket az 1 első indikátorral, vagy az önálló direktory, illetve fájl esetén a 773-as mezőt a 0 első indikátorral [25].

### **Az Internet forrásainak rendszerezése: nem katalogizálói megközelítés**

Bármilyen nyomtatott indexről vagy katalógusról legyen szó, a hatékony kereséshez elengedhetetlen a kereső mechanizmus megértése, az adott adatbázis méretének, tartalmának, és az alkalmazott indexelési stratégiának az ismerete. Ugyanez igaz a hálózatokon elektronikusan tárolt információ esetében is. Valamennyi keresőprogramnak megvannak a saját tartalmi korlátai. Az adatelemhalmaz, amelyen a keresés folyik, szintén korlátozott. Ráadásul mindegyik keresőprogram esetében adott, hogy a keresési eredményeket milyen formában fogja megjeleníteni. A legnagyobb problémát azonban az jelenti, hogy nagyon nehéz megállapítani, mi az, amit a keresés magába foglal. Sok keresőprogram írja ki a keresési eredményekre

vonatkozóan, hogy „mennyi” a releváns találat, de azt nem, hogy „miből.” Ha ugyanazt a keresést elvégeztetjük több keresőprogrammal is, akkor hamar kiderül, hogy egyetlen indexelő szolgáltatás sem képes mindent elérni az Interneten, a keresések pedig vegyes és kiszámíthatatlan eredményeket szolgáltatnak.

Valamennyi internetes keresőprogramnak megvannak az előnyei és hátrányai. Az előnyök általában a következők: ezek az indexek naponta 24 órán át használhatók; minden keresés eredményként egy listát kapunk, amelynek az elemei hozzá vannak kapcsolva a dokumentumokhoz, így szinte azonnal elérhetők; az Internet-források folyamatosan bővülnek; ezek az indexek többdimenziósak, azaz rajtuk keresztül más könyvtári katalógusokhoz vagy indexekhez is eljuthatunk. A hátrányok között a következőket sorolhatjuk fel: a keresési eredmények között sok az ismétlődő elem; az eredmények kiszámíthatatlanok; az eredmények félrevezetőek lehetnek, hiszen egyes keresések üres halmazt eredményeznek egy bizonyos keresőprogrammal, ugyanakkor egy másikkal több releváns forrást is találnak; egyes keresőprogramok nem adnak információt adatbázisuk tartalmáról, vagy arról, milyen szempontok szerint válogatnak a dokumentumok között, hogy egy adott dokumentum bekerüljön-e az adatbázisukba, vagy nem; szótáraik nem ellenőrzöttek; az írásjelek, valamint a kis- és nagybetűk használata nem szabványosított; a kapcsolatok és a relevancia megállapítása sokszor nem lehetséges a dokumentumok megvizsgálása nélkül [11].

Az Internet világában való eligazodásra használt jelenleg elérhető segédeszközök (archie, veronica, jughead, és ezek World Wide Web megfelelői) hasznosak, de összehasonlítva egy átlagos könyvtári katalógus vagy folyóiratindex kifinomultságával és precízségével, még mindig nagyon kezdetlegesek. A problémát részben az jelenti, hogy a könyvtárosok még mindig inkább csak reagálnak a számítógépes szakemberek által javasolt technikai újításokra, ahelyett, hogy maguk is javaslatokat tennének. Ugyanakkor a számítógépes szakemberek nem igazán foglalkoznak a könyvtár-tudomány megismerésével, és így folyamatosan próbálják újra feltalálni azokat a fogalmakat, amelyeket a könyvtárosok már évtizedek óta használnak. Talán ennél nagyobb probléma az a számítógépes területen elterjedt nézet, hogy minden human értékelés nagyon időigényes és drága, ezért hasznosabb a robotok használata ezekre a feladatokra.

P. Riddle, a RiceInfo adminisztrátora a Rice Egyetemen kísérletet tett arra, hogy a könyvtárosi szelekciós és osztályozási elveket alkalmazva számítógépet használjon az Internet-források rendszerezésére [26]. Egy programot írt arra, hogy az

előre kiválasztott, nem helyileg tárolt, egymással kompatibilis tárgyi gopher menüket beolvassza a helyi gopher menübe, így létrehozva egy igazán átfogó gopheres fastruktúrát. Ennek a megközelítésnek azonban több hátránya is van. Egyrészt az információs zaj, másrészt pedig az, hogy ez a program a gopher szerkezetén alapszik, amit nehéz lenne a World Wide Webre is alkalmazni. A legnagyobb probléma pedig az, hogy az Internet-források besorolása egy maréknyi kategóriába az Internet jelenlegi gyors növekedése mellett lehetetlen.

A könyvtári osztályozás rendszereit felhasználó projektek többsége a hierarchikus osztályozást választotta az Internet forrásainak rendszerezésére. A gopher menük felépítése egyértelműen a hierarchikus osztályozási rendszerek hatását mutatja. A World Wide Weben a CyberDewey [27] a nevében is jelzi, hogy a Dewey Tizedes Osztályozás elveit követi. A WWW Virtual Library [28] és a CyberStacks [29] a Kongresszusi Könyvtár osztályozási rendszerét adaptálta. Nagy-Britanniában a BUBL Subject Tree projekt az Egyetemes Tizedes Osztályozást használta fel [30]. A fent említett projektekben az Internet-forrásokhoz a megfelelő osztályozási rendszereken alapuló jelzetek kapcsolódnak. Megjegyzendő, hogy a CyberDewey, a CyberStacks és a WWW Virtual Library esetében a kiválasztott osztályozási rendszert egyszerűsített jelzetek képviselik. A források azonban nem teljesen katalogizáltak, és a tárgyi hozzáférés is meglehetősen korlátozott. A felhasználó az alapul szolgáló osztályozási rendszer szintjeinek többé-kevésbé megfelelő hierarchia minden egyes szintjén egyre szűkíti a keresést mindaddig, amíg minden egyes képernyőn kiválasztva a megfelelő témát, végezetül megérkezik a keresett információhoz. Ez tulajdonképpen olyan, mintha a felhasználó az osztályozási táblázatban követne egy útvonalat az általánostól a specifikusig. Annak ellenére, hogy a felhasználónak sok képernyőn kell áthaladnia, ami lelassítja a keresés folyamatát, a keresés ezen típusa többnyire sikeresebb, és kevesebb zsákutcába vezet, mint a kulcsszó szerinti keresés.

### CyberDewey

A CyberDewey láthatóan nem könyvtárosok által készített projekt. Létrehozója, *David Mundie* [31] saját számítógépes fájljai rendszerezése során ismerkedett meg a Dewey Tizedes Osztályozással, miután sok egyéb, nem könyvtári rendszerezési formát kipróbált. Lelkesedését csak növelte a felismerés, hogy a Dewey Osztályozási Rendszer a könyvtárosok állandó revíziója alatt van, így soha többé „nincs egyedül” az osztályozási jelzetek létrehozását érintő döntésekben [27].

### WWW Virtual Library

A WWW Virtual Library a CyberDewey-nél jóval kiterjedtebb projekt, amely a Kongresszusi Könyvtár Osztályozási Rendszerét felhasználva sokkal hatékonyabban is osztályozza a szakreferensek és a végfelhasználók által kiválasztott forrásokat. A Virtual Library a web egyik legrégebbi katalógusa. Készítője *Tim Berners-Lee*, aki magának a webnek is a kitalálója. Az egyéb webkatalógusokkal és keresőprogramokkal ellentétben a Virtual Libraryt olyan önkéntesek építik, akik jártasak valamely témában, és csak az abban a témában hasznosnak tartott weboldalak URL címeit gyűjtik. A Virtual Library által lefedett terület talán nem olyan nagy, mint az egyes kereskedelmi, robotokat alkalmazó keresőoldalaké, de az általa gyűjtött weboldalak általában sokkal hasznosabbak, mivel azok az egyes szakterületek szakértőinek minőségi ellenőrzésén mentek át. A Virtual Library nem egyetlen adatbázis. Önálló indexek szerverek százain találhatóak az egész világon. Ezeknek a listáknak egy URL-gyűjteményét tartja fenn *Gerard Manning* a Stanford Universityn [28].

### CyberStacks<sup>SM</sup>

A CyberStacks<sup>SM</sup> 1995-ben World Wide Web virtuális könyvtárként jött létre, azzal a céllal, hogy megvizsgálja a hagyományos könyvtári osztályozási rendszerek alkalmazhatóságát az Internet forrásainak rendszerezésére [29]. A CyberStacks a Kongresszusi Könyvtár osztályozási rendszerét adaptálta a kiválasztott Internet-források közti kapcsolatok jelölésére. Annak ellenére, hogy a CyberStacks a természettudományok és a technika témaköreire korlátozódik, a projekt a hagyományos osztályozás felhasználásának nagyon érdekes prototípusává fejlődhet, mivel létrehozója, *Gerry McKiernan* egy hipertext teaurusz kialakítását is tervezi [32, 33].

### Beyond Bookmarks

Az Internet-források Beyond Bookmarks nevű gyűjteménye a CyberStacks projekthez kapcsolódóan jött létre [34]. Hogy a hagyományos könyvtári módszereknek az Internet forrásainak rendszerezésére való alkalmazhatóságát növeljék, és hogy a CyberStacks modell fejlesztéséhez új, alternatív megközelítéseket találjanak, a CyberStacks projekt résztvevői több fórumon (levelezőlistán, hírcsoporton) kértek információt olyan forrásokra vonatkozóan, amelyek szabványos és nem szabványos könyvtári osztályozási rendszereket és ellenőrzött szótárakat alkalmaznak a web forrásainak rendszerezésére. A javasolt forrásokat átnézték, kategóriákba sorolták, és a hiperlinkek segítségével

beépítették egy gyűjteménybe, amelyet Beyond Bookmarksnak neveztek el.

### **Project Aristotle**

A hagyományos könyvtári osztályozási rendszerek és az ellenőrzött szótárak hasznos szerkezeti keretet kínálnak a webforrások azonosítására és használatára, alkalmazásuk azonban a jelenlegi környezetben igen nagy erőfeszítéseket igényel, mind a létrehozás, mind a fenntartás terén. A CyberStacks személyzete tehát elhatározta, hogy megvizsgálja a webforrások automatizált kategorizálását. 1996 nyarán több tucat olyan projektet és szolgáltatást vizsgáltak meg, amelyek az automatikus rendszerezés valamilyen formáját alkalmazzák [35].

### **BUBL Subject Tree**

A BUBL (Bulletin Board for Libraries) Subject Tree tűnik a leginkább könyvtárorientált projektnek. A BUBL volt az első olyan szolgáltatás az Egyesült Királyságban, amely tárgy szerinti felosztású hozzáférést biztosított az Internet-forrásokhoz. Létrehozásakor a projekt szintén egyedi volt abból a szempontból, hogy gopher források és webdokumentumok feldolgozását egyaránt magába foglalta [36]. Jelenleg a BUBL személyzete és az Egyesült Királyság egész területéről származó önkéntes szakreferensek közösen dolgoznak a tárgyi hierarchia részeinek fenntartásán. Csak olyan forrásokat dolgoznak fel, amelyek az egyetemi körök érdeklődésére számot tarthatnak. Meggyőződésük, hogy az egyetemi körök számára hasznos Internet-források felfedezése és nyilvántartása a helyi szintű, valamint a nemzeti és nemzetközi kooperatív katalogizálással együttesen oldható meg.

A BUBL Subject Tree érdekes változata a NISS (National Information Services and Systems), amely a hagyományos katalogizálás előnyeit próbálja hasznosítani a merev katalogizálási szabályok, a professzionális katalogizálói szaktudás, és egy nagy központi intézmény nélkül. Az Egyesült Királyság könyvtárosait felkérték, hogy egy Forrásleírási Űrlap kitöltése révén küldjenek be forrásleírásokat (forrásrekordokat) a NISS-hez. Ezek az űrlapok meg sem kísérelték, hogy a MARC rekordok formátumát utánozzák, bár egyaránt tartalmazzak fizikai jellemzőket leíró, és tartalomra vonatkozó mezőket. A BUBL Subject Tree koncepciója a NISS projektben a tárgy és kulcsszó szerinti kereséssel egészül ki.

### **ROADS**

A ROADS (Resource Organisation and Discovery in Subject-Based Services) egy másik

## **Gőz Á.: Az Interneten elérhető információforrások...**

olyan osztályozáson/katalogizáláson alapuló projekt, amely tárgy szerinti hozzáférést és bibliográfiai információt egyaránt tartalmaz. Fő célja úgy biztosítani hozzáférést a hálózati forrásokhoz, hogy a felhasználók számára észrevétlen maradjon a különböző szolgáltatások közti mozgás. A ROADS közös platformot kínál több tárgyi felosztású információszolgáltató rendszerben való kereséskor.

### **Összegezés**

Az információrendszerezés elvei hagyományosan a következőket foglalják magukba: meg kell állapítani, hogy milyen források léteznek, és közülük kiválasztani a felhasználói igényeknek megfelelőeket; a kiválasztott források leírása; az elérési pontok biztosítása és egységesítésük; a kiválasztott források tartalmi elemzése; valamint ezeknek a forrásoknak a lokalizálásához szükséges információ biztosítása. Összehasonlítva a könyvtárakat és a keresőprogramokat, vagy az Internet keresőszolgáltatásait, a könyvtárak sokkal alkalmasabbnak tűnnek a források kiválasztására, hiszen hagyományosan sokkal nagyobb gyakorlatuk van a különböző formátumú anyagok felhasználói igények alapján történő kiválasztása és beszerzése terén.

Az Internet-források értékelése, kiválasztása, leírása és tárgyi elemzése mindenképpen szükséges a hatékonyabb hozzáférés érdekében. A könyvtárosok gyarapítási szaktudása biztosítja a felhasználók számára hasznos Internet-források értékelését és kiválasztását is. A katalogizálók már régóta növelik az információhordozó anyagok értékét azzal, hogy rendszerezik őket, és ahogyan azt néhány katalogizálási projekt is mutatja, az általuk használt információrendszerezésre vonatkozó irányelvek hatékonyan alkalmazhatók az Internet-források esetében is [24]. Függetlenül tehát attól, hogy minek is nevezzük az Internet dokumentumainak rendszerezésére tett kísérleteket (bibliográfiai, katalógusok, listák stb.), létrehozásuk, fenntartásuk mindenképpen humán erőfeszítéseket igényel.

Ezenkívül természetesen sok nyitott kérdés marad, amire az egyes digitális könyvtárak célkitűzései adják meg a választ. Milyen típusú anyagokat kell, illetve érdemes katalogizálnunk? Hogyan különböznek a digitális katalógusok a hagyományostól? Képzett katalogizálók fogják fenntartani őket? Milyen ismeretek szükségesek az új típusú katalógusok létrehozásához, és hogyan lehet azokat elsajátítani? Milyen mértékben alapulnak majd ezek a katalógusok a már létező (MARC vagy AACR) szabványokon? Lesz-e, elképzelhető-e

egyáltalán a digitális anyagok egyetemes katalógusa? [3]

Az Internet egészét rendszerező információs rendszer létrehozása csábítónak tűnik, azonban egy ilyen rendszer létrehozása jelenleg nem lehetséges, és nem is lenne hasznos a következő okok miatt: az Internet túl sok forrást tartalmaz; sok forrás minősége erősen kétséges, vagy tartalmánál fogva értéktelen a felhasználók számára. Sokkal hasznosabbnak és megvalósíthatóbbnak tűnik az OCLC által követett modell, amely több intézmény együttműködésén alapul. A részt vevő könyvtárak saját használói körük igényei alapján értékelik az Interneten elérhető információs forrásokat, és a helyi felhasználók számára releváns dokumentumokról szolgáltatnak rekordokat a közös adatbázisba. Ez a gyakorlat nagyon hasonlít az OCLC eredeti programjához, itt azonban a katalógizálás leíró része jelentősen leegyszerűsített. Ez az együttműködés nagy mennyiségű minőségi Internet-forrás számbavételét teszi lehetővé. Egy ilyen, vagy ehhez hasonló információs rendszer minőségi rekordjai és hatékony információ-visszahívása révén válik közkedvelté a felhasználók körében.

A metaadat jelenleg a legjelentősebb dolog, ami a jövőbeni könyvtári osztályozási és katalógizálási projekteket befolyásolja majd az Interneten. Az új metaadat-specifikációk és a katalógusrekordok közti párhuzamok nyilvánvalóak. A TEI fejléc és a MARC formátum egymásba konvertálhatósága, vagy legalábbis az erre mutató törekvések egyértelműen jelzik, hogy a katalógizálási gyakorlat és a katalógizálási szabványok nagymértékben hozzájárulhatnak az Internet-források rendszerezéséhez. A könyvtári metaadatszabvány, a MARC módosítások során ment keresztül, hogy tükrözze az elektronikus kor követelményeit. Különösen fontos az új 856-os mező [21], amely lehetővé teszi a felhasználók számára, hogy a könyvtári online katalógusból vagy a hálózaton elérhető katalógusból közvetlenül eljussanak az elektronikus dokumentumokhoz. A metaadatok nemcsak a könyvtárosok, hanem valamennyi Internet-szolgáltató számára alapvető fontosságúak. Sokan úgy vélik, hogy a Harvest vagy a TEI szabványok a legalkalmasabbak a hálózati használatra [1]. Az OCLC viszont azt szeretné, ha a Dublin Core válna nemzetközi szabvánnyá. Most azonban még nem tudhatjuk, hogy mely szabványok nyernek széles körű elismerést, és válnak nemzetközileg elfogadottá.

Az Interneten levő hatalmas mennyiségű információ rendszerezésével kapcsolatos problémákat összegezve nyilvánvaló, hogy az online világnak szüksége van azokra a tapasztalatokra és ismeretekre, amelyeket a könyvtárosok az információrendszerezés során kifejlesztettek: az információhordozó anyagok beszerzésére; szerző, cím

és tárgyi hozzáférés biztosítására ezekhez a dokumentumokhoz; valamint információ szolgáltatására a dokumentumok „fizikai” elérhetőségéről. A használók átsegítése az online információs források labirintusán nagyon hasonlít a felhasználóknak a különböző referenzkönyvek, katalógusok és osztályozási rendszerek útvesztőjében való eligazításához [5]. A digitális gyűjtemények kialakításakor pedig az Interneten elérhető online dokumentumok „beszerzése” a hagyományos könyvtári gyűjteményépítéshez szükséges ismereteket, és a hagyományos osztályozási rendszerek használatát követeli meg.

## Irodalom

- [1] WOODWARD, J.: Cataloging and classifying information resources on the Internet. = M. E. Williams (Ed.). *Annual Review of Information Science and Technology*, 31. köt. 1996. p. 189–220. Medford, NJ: Learned Information.
- [2] BRAUN, L. W.–FLEMING, J.: From the desktop: Editor's note. = *Internet Trend for Libraries*, 1. köt. 4. sz. 1996. <<http://www.public.iastate.edu/~CYBERSTACKS/ITW.htm>>
- [3] LEVY, D. M.: Cataloging in the digital order. = *Digital Libraries '95, The Second Annual Conference on the Theory and Practice of Digital Libraries*, June 11–13, 1995, Austin, Texas, USA. <<http://www.csd.tamu.edu/DL95/papers/levy/levy.html>>
- [4] QUITTNER, J.: Internet plays growing role as world's electronic highway. *Newsday*, 1992. november 2. p. 3.
- [5] TAYLOR, A. G.: The information universe: Will we have chaos or control? = *American Libraries*, 1994. júl.–aug. p. 629–632.
- [6] CAPLAN, P.: Cataloging Internet resources. = *The Public-Access Computer Systems Review*, 4. köt. 2. sz. 1993. p. 61–66. <<http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/caplan.txt>>
- [7] JUL, E.: Why catalog Internet resources. = *Computers in Libraries*, 16. köt. 1. sz. 1996. p. 8–10.
- [8] *Cataloging and Retrieval of Information Over Network Applications (CATRIONA)*, 1995. CATRIONA Project: Documents. <<http://www.bubl.bath.ac.uk/BUBL/maincatriona.html>>
- [9] Argus Clearinghouse, 1996. Mission and Philosophy. <<http://www.clearinghouse.net/docs/mission.html>>
- [10] STEINBERG, S. G.: Seek and ye shall find (maybe). = *Wired*, 4. köt. 5. sz. 1996. p. 108–114., 172–182.
- [11] TAYLOR, A. G.–CLEMSON, P.: Access to networked documents: Catalogs? Search engines? Both? 1996. <<http://www.oclc.org/oclc/man/colloq/taylor.htm>>
- [12] SCHAMBER, L.: What is a document? Rethinking the concept in uneasy times. = *Journal of the American Society for Information Science*, 47. köt. 9. sz. 1996. p. 669–671.

- [13] GUEDON, J.-C.: Why are electronic publications difficult to classify?: The orthogonality of print and digital media. <<http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/guej1.txt>>
- [14] CAPLAN, P.: Providing access to online information resources: A paper for discussion. 1992. <<http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/caplan2.txt>>
- [15] KLEMPERER, K.-CHAPMAN, S.: Digital libraries: a selected source guide. 1994. = TMT, 45. köt. 7. sz. 1998. p. 265–271. <[http://www.lita.org/ital/1603\\_klempere.htm](http://www.lita.org/ital/1603_klempere.htm)>
- [16] WEIBEL, S.: Dublin Core metadata element set: Reference description. 1997. <[http://purl.org/metadata/dublin\\_core\\_elements\\_970711](http://purl.org/metadata/dublin_core_elements_970711)>
- [17] Research Libraries Group. Metadata: Connecting researchers with relevant resources. = The Research Libraries Group, 44. köt. 1997. p. 3–10.
- [18] GRAHAM, P. S.: The mid-decade catalog and its environment. 1994. <<http://aultnis.rutgers.edu/texts/cffc.html>>
- [19] OLSON, N. B. (ed.): Cataloging Internet resources: A manual and practical guide. Second edition. 1997. <<http://www.purl.org/oclc/cataloging-internet>>
- [20] Intercat-Internet cataloging project home page. 1996. <<http://www.oclc.org/oclc/man/catproj/catcall.htm>>
- [21] Library of Congress. Guidelines for the use of field 856. 1997. augusztus. <<http://www.loc.gov/marc/856guide.html>>
- [22] JUL, E.: Now that we know the answer, what are the questions? 1997. <<http://jic.libraries.psu.edu/jic1nr3-42.htm>>
- [23] Alcuin: Online catalogs with „Webbed” interfaces. <<http://www.lib.ncsu.edu/staff/morgan/alcuin/wwwed-catalogs.html>>
- [24] HSIEH-YEE, I.: Modifying cataloging practice and OCLC infrastructure for effective organization of Internet resources. 1996. <<http://www.oclc.org/oclc/man/colloq/hsieh.htm>>
- [25] SHA, V.: Guidelines for cataloging Internet resources. University of Missouri Columbia, School of Law Library. 1995. <<http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/sha1.txt>>
- [26] RIDDLE, P.: Library culture, computer culture, and the Internet haystack. 1994. <<http://is.rice.edu/~riddle/dl94.html>>
- [27] CyberDewey. A guide to Internet resources organized using Dewey Decimal Classification codes. 1989. <<http://ivory.lm.com/~mundie/DDHC/DDH.html>>
- [28] WWW Virtual Library. <<http://vlib.stanford.edu/Overview.html>>
- [29] CyberStacks. CyberStacks home page. 1997. <<http://www.public.iastate.edu/~CYBERSTACKS/homepage.html>>
- [30] BUBL (Bulletin Board for Libraries) Information Service. BUBL WWW Subject Tree – arranged by Universal Decimal Classification. 1996. <<http://www.bubl.bath.ac.uk/BUBL/Tree.html>>
- [31] MUNDIE, D. A.: Organizing computer resources: Or, how I learned to stop worrying and love the DDC. 1995. <[http://ivory.lm.com/~mundie/DDHC/organizing\\_computers.html](http://ivory.lm.com/~mundie/DDHC/organizing_computers.html)>
- [32] McKIERNAN, G.: Casting the Net: The development of a resource collection for an Internet database. 1996. <<http://www.library.ucsb.edu/untangle/mckiernan.html>>
- [33] McKIERNAN, G.: The once and future library. 1996. <<http://www.library.ucsb.edu/istl/96-fall/mckiernan.html>>
- [34] Beyond Bookmarks: Schemes for organizing the Web. <<http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>>
- [35] Project Aristotle<sup>SM</sup>: Automated categorization of Web resources. <<http://www.public.iastate.edu/~CYBERSTACKS/Aristotle.htm>>
- [36] STEELE, M.: The BUBL Subject Tree and Catriona (Cataloging and Retrieval of Information over Networks Applications). = Computers in Libraries, 15. köt. 5. sz. 1995. p. 63–65.

Béérkezett: 1998. VI. 4-én.

## Új online könyvesbolt

Megnyílt a *Corvina internetes könyvruház*. A Pesti Műsor és a Drótposta Kft. által indított online áruház nemcsak új könyveket árul, de a Sós Antikváriummal együttműködve antikvár köteteket is. A könyvesház más, könyvvel kapcsolatos szolgáltatásokat is nyújt: keres-kínál jellegű ingyenes apróhirdetéseket tesz közzé, és online könyvaukciókat szervez. *Kende Péter*, a Pesti Műsor főszerkesztője a könyvesház bemutatóján elmondta, hogy terveik között szerepel valós idejű, online könyvárverés szervezése és a könyvek mellett folyóiratok árusítása is. A Corvinában hitelkártyával is lehet fizetni, ezt azonban telefonon intézik (<http://corvina.hu>).

/INteRNeTTo heti hírlevél, 1998. 17. hét./

## Digitális azonosító a publikációkhoz

Az Amerikai Kiadók Szövetsége és a Nemzeti Kutatási Kezdeményezések együttműködve kifejlesztettek egy elektronikus azonosító rendszert az Interneten lévő anyagokra. A digitális anyagazonosító (digital object identifier = DOI) rendszer összekapcsolja a lehetséges tartalomhasználókat a tartalom szerzői jogának tulajdonosával, egy, a munkához kapcsolódó alfanumerikus címke segítségével. Az AAP azt reméli, hogy a rendszer alkalmas lesz a védett munkákhoz való hozzáférés korlátozására, de jelenleg még csak szigorúan becsületesbeli a rendszer. A DOI rendszert az újonnan alakuló Nemzetközi DOI alapítvány tartaná fenn és adminisztrálná (<http://www.doi.org>).

/Computer Technika, 5. köt. 7. sz. 1998. p.4–5./