

## Digitális könyvtárak a világban\*

***A „digitális könyvtár” kezd olyan, sokak számára nehezen érthető vagy félreértett divatos bűvszóvá válni, mint az információs társadalom. Az alábbi írás megkísérli, hogy meghatározásokkal és megvalósított vagy fejlesztés alatt lévő projektek ismertetésével tisztázza, mit is értsünk ezen a kifejezésen.***

### Bevezetés

Digitális könyvtárakkal vagy legalábbis megvalósításukra irányuló (sok esetben nagyszabású) kezdeményezésekkel Európa több országában, az Amerikai Egyesült Államokban, Japánban, Ausztráliában és Új-Zélandon is találkozunk. Ezek többségéhez közvetve vagy közvetlenül hozzátartozik a digitális könyvtár fogalmának meghatározása. A fejlesztés sok esetben egyelőre csak a szervezeti kereteket és a fejlesztés elveit jelenthette, de ezek is igen tanulságosak.

### A digitális könyvtárak főbb jellemzői

*Lisa Dallape Matson és David J. Bonski, egy könyvtáros és egy információtechnikával foglalkozó szakember szerint „a digitális könyvtár olyan könyvtár, amely állományának egészét vagy annak lényeges részét számítógéppel feldolgozható formában tartja fenn, alternatívaként, pótlásként vagy kiegészítésként a hagyományos nyomtatott és mikrofilmen rögzített anyagok mellett, amelyek jelenleg meghatározóak a könyvtári gyűjteményekben” [1].*

Érdemes e meghatározás mellé állítanunk a British Library definícióját, amely szerint a digitális könyvtár digitális technikák használatát jelenti információk és dokumentumok beszerzésére, tárolására, megőrzésére. A digitális könyvtárakkal hozzáférést biztosíthatunk ezekhez, függetlenül attól, milyen formában publikálták őket eredetileg. A digitális könyvtár digitálisan tárolt dokumentumok (szavak, álló- és mozgóképek, hang és ezek bármely kombinációja) kritikus tömegét is jelenti. Ezeket a dokumentumokat szervezett és rendezett keretben a világon bárhol található felhasználónak kívánságra elérhetővé kell tenni. A szolgáltatott forma és médium a piac függvényében különféle

lehet. A dokumentumok több helyen is tárolhatók lehetnek, és több helyről származhatnak [2].

A digitális könyvtárak elsődleges célja, hogy a hálózatot egyetlen virtuális gyűjteménynek tekintsük, amelyből a felhasználók releváns részeket kereshetnek ki. Ennek megfelelően a méretezés, így az objektumok és a repozitóriumok számának kérdései igen fontosak [3]. A méretezés nyilvánvalóan alapkérdés, hiszen a túlzottan kis méretű tesztek eredményei utóbb sokszor nem bizonyulnak országos, még kevésbé globális rendszerekben alkalmazhatóknak. Ehhez pedig a digitális dokumentumok kritikus tömegének megállapítására van szükség, ami annak meghatározását fedi, hogy mennyi dokumentumot digitalizáljanak és vásároljanak meg a digitális könyvtár létrehozásához.

A digitális könyvtár tehát több helyen keletkezhet, de végső célként egyetlen egységként lehet majd hozzáférni. Képes adatok hatalmas mennyiségét kezelni: különböző formátumú, különböző forrásból származó szöveget és képeket, digitális hangot és videót, amelyek jelentős tárolókapacitást igényelnek.

A digitális könyvtárban a felhasználó maga kereshet információt, és önállóan rendelhet meg dokumentumokat, ami a könyvtárak közötti együttműködés és a finanszírozás megváltozását fogja magával hozni [4]. Egyetlen integrált és különböző platformokon működő felhasználói interfésszel kell rendelkeznie, és a dokumentumok eredeti leőhelyéhez képest a használatot jobban szolgáló szervezést, kategorizálást és indexelést kell nyújtania.

A digitális könyvtárnak megfelelő védelmet kell adnia a szerzői jog és más védelem tekintetében,

\* A Nemzeti Kulturális Alap Digitális Kultúra ad hoc kuratóriumi pályázatának keretében – a Neumann-ház, a Magyar Elektronikus Könyvtár és a KFKI Rézecske- és Magfizikai Kutató Intézet Számítógép Hálózati Központ felkérésére – készült tanulmány rövidített változata.

képesnek kell lennie arra, hogy átfogó kimutatást nyújtson az összes egyedi dokumentum használatáról. Dokumentumait úgy kell tárolni és megőrizni, hogy azok elérhetőek legyenek jóval a közvetlen aktualitásuk elvesztése után is.

A digitális könyvtárban a globális hozzáférés fontossága mellett sem feledkezhetünk meg a helyi felhasználók igényeinek fontosságáról, és a helyi felhasználói közösségről mint a szolgáltatást formáló erőről. Biztosítani kell a biztonsági tárolás és az információ helyreállításának mechanizmusait, méghozzá minél kevesebb emberi ráfordítással. Nem kell minden adatnak azonnal hozzáférhetőnek lennie, egyes dokumentumok lassabban elérhető, így olcsóbb hordozókon is tárolhatók.

A hosszú távú megőrzésnek az információ frissítését is magában kell foglalnia. A fizikai struktúrán kívül meg kell őrizni a művek intellektuális integritását is. Meg kell határozni, hány helyen kell tárolni az elektronikus információkat, hogy megőrzésük és a hozzáférhetőség biztosítva legyen, de valószínű, hogy a duplikáció mértéke nem lesz nagy. Meghatározó lesz ebben a kérdésben a nemzeti és regionális érdekek érvényesülése, a hálózatok sávszélessége, a gyorsaság, és ezek költségvonzata is, ami viszont redundanciához vezet. Foglalkozni kell dokumentumok részeinek visszakeresésével és elérésével, egyes dokumentumfajták elavulásának, hatályon kívül helyezésének jelzésével is.

A digitális könyvtár legtöbbször által osztott víziója olyan osztott repozitóriumok hálózata, amely kereshető indexelt gyűjteményekben és azok között. Rövid távon tehát olyan technikákat kell kifejleszteni, amelyek transzparens keresést tesznek lehetővé ezek között a repozitóriumok között, és képesek kezelni a formátumok és protokollok különböző változatait.

Hosszú távon nemcsak ezek, hanem a tartalom és jelentés variációinak kezelésére is ki kell dolgozni a módszereket. A végső cél tehát a mély szemantikai együttműködés (interoperabilitás), vagyis az a képesség, hogy a felhasználó konzisztensen és koherensen érhesse el heterogén repozitóriumokban elosztott digitális objektumok és szolgáltatások hasonló (bár autonóm módon definiált és kezelt) osztályait, amelyek olyan módon vannak kapcsolva egymáshoz, hogy egyetlen, szervezett gyűjteménynek látszanak. Ez utóbbi a *federating*, amely egyelőre nemigen fordítható magyarra (talán a *közös működés* kifejezés megközelíti az együttműködő, mégis önálló egységek értelmét), és túlterjed a repozitóriumok felhasználásán, amelyek maguk indexelt objektumok gyűjteményei.

Sok meghatározásban a digitális könyvtár mellett szinonim módon szerepel a *virtuális* és az *elektronikus* könyvtár. A mi használatunkban az

elektronikus könyvtár a számítógépek könyvtári alkalmazásának összességét jelenti; a virtuális könyvtár pedig azoknak az azonosítóknak rendezett gyűjteménye, amelyek másutt tárolt, és hálózaton elérhető dokumentumokat jelölnek.

Formai szempontból a digitális könyvtár gyűjteményébe tartozik a könyvtári anyagok sokféle gyűjteménye, digitális formában így

- a retrospektív módon konvertált anyagok éppúgy, mint újonnan közölt dokumentumok, amelyek lehetnek nyomtatott művek származékai, vagy eredeti elektronikus kiadványok;
- a dokumentumtípusok széles köre, beleértve a folyóiratokat, monográfiákat, referenciaműveket, kép-, hang- és mozgóképanyagokat;
- mindenféle digitális formátum a bitmap képektől az SGML-kódolású szövegeken át a speciális formában kódolt kép- és hanganyagig. (Megjegyezzük, hogy a továbbiakban a *dokumentum* kifejezést bármilyen multimédia-állományra érvényesen használjuk.)

## A metaadatok kérdése

A digitális könyvtárak létrehozása kapcsán kezelni kell a metaadatok (metadata) meghatározásának és használatának kérdését, amely magába foglalja ezen adatoknak az objektumokból való kinyerését vagy kiszámítását, és az objektumok számítógépes leírását. A metaadatok legismertebb formája a katalógusrekord, metaadatokat azonban a könyvtárakon kívül is használnak.

A metaadatok sok esetben magukból a dokumentumokból is kinyerhetők, ami különösen igaz a funkcionálisan strukturált (pl. SGML) dokumentumok esetében, de alapul szolgálhat a dokumentumok teljes szövege is.

A sokféle metaadat-formátum megléte hívta életre a *Dublin Core (DC)* adatkészletet, amely mindössze 13 elemből áll. Nevének megfelelően ez a mag a következőket tartalmazza:

- tárgy (Subject),
- cím (Title),
- szerző (Author),
- kiadó (Publisher),
- más ágens (pl. szerkesztő, Other Agent),
- dátum (a publikáció dátuma, Date),
- objektumtípus (műfaj, Object Type),
- forma (adatrepresentáció: Postscript fájl, Windows exe-fájl stb., Form),
- azonosító (egyedi azonosító, Identifier),
- reláció (más objektumokhoz, Relation),
- forrás (Source),
- nyelv (Language),
- az objektum térbeli elhelyezkedése és időbeli tartama (Coverage).

A DC a részletesebb katalogizálás alapjául szolgál. A metaadatok származhatnak a szerzőtől, pl. HTML fejlécekből, a kiadótól és természetesen katalogizálóktól. Elemei a dokumentum belső tartalmára szorítkoznak, a külsődleges (pl. az elérési jogokkal, árakkal kapcsolatos) információkat a DC-hez kapcsolható kiegészítő metaadatokban kell rögzíteni. Ebből következik az is, hogy a DC bővíthető (az előbbieken kívül pl. helyi állományadatokkal). A DC független a szintaxistól. Ezért szándékosan nem határozták meg az elemek szintaxisát, viszont szemantikájuknak a felhasználók széles köre számára érthetőnek kell lennie. Az elemek ismételhettek és opcionálisak.

A DC elemeit nem kívánják megváltoztatni és bővíteni. Szükség van azonban az ezt kiegészítő információkra is. Ezek a következő típusúak lehetnek (és természetükből következően opcionális információcsomagokat jelentenek):

- szakterülettől függő leírás,
- a felhasználás feltételei,
- címkék és besorolás,
- biztonsági információ, autenticitás, aláírások,
- összetett dokumentumokra vonatkozó információ,
- a különböző manifesztációkra és változatokra vonatkozó információ,
- a teljes dokumentum,
- az archiválásra vonatkozó információ (ki archiválja?).

Ez az úgynevezett warwicki keretmegállapodás (*Warwick Framework*) [5].

## Néhány digitális könyvtári projekt

### Amerikai Egyesült Államok

Az Egyesült Államokban 1996 óta a Kongresszus támogatásával folyik a Nemzeti Digitális Könyvtár projekt a Kongresszusi Könyvtár, több egyetemi könyvtár és a magánszféra részvételével. Tekintettel a fejlesztés méreteire és irányultságaira, a folyamatban lévő munkát három csoportban tárgyaljuk:

- a Kongresszusi Könyvtár Történeli Gyűjteménye programja,
- az egyetemi keretek között folyó Digitális Könyvtári Kezdeményezés (Digital Library Initiative) projektjei,
- egyéb kezdeményezések az Egyesült Államokban.

### *A Kongresszusi Könyvtár Nemzeti Digitális Könyvtár Történeli Gyűjteménye programja*

A Kongresszusi Könyvtár (LC) ötéves programja azt célozza, hogy saját nyomtatott és nem nyomtatott anyagából elkészítse az amerikai történelmi és kulturális forrásanyag kezdeti magját digitális formában, amelyet bárki elérhet hálózaton keresztül.

A munka egyik tényezője a heterogén anyag kiválasztása és digitalizálása. A másik tényező a kényelmes és hatékony hozzáférés kialakítása széles használói kör számára az iskolás gyerekektől a kutatókig. További szempont az, hogy bár a könyvtár nem kutatási és fejlesztési intézmény, mégis az ilyen munkák elkerülhetetlenek, ezért a legújabb technikát kell alkalmazni, tekintettel kell lenni a szellemi tulajdonra vonatkozó jogokra, végül az LC speciális feladataira. A munka során támaszkodhattak korábbi szellemi eredményeikre, a MARC és a Z39.50 szabvány kidolgozására.

A program 1994. évi indulásakor 200 americana gyűjteményt vettek számba mint lehetséges elemet, és figyelembe vették az elosztott gyűjtemények lehetőségét is, együttműködve más könyvtárakkal. A projekt előzményei közé számítják az 1982–87 közötti Optikai Lemez kísérleti programot, és az 1989–94 között futott Amerikai Emlékezet kísérleti programot, ez utóbbi CD-ROM-on és analóg videolemezen rögzített americanákat.

Egy 1996. áprilisából származó beszámoló 12 történelmi gyűjteményt jelez elérhetőnek az Amerikai Emlékezet hálóoldalakon az 1920-as évekből származó hanganyagokkal és mozifilmrészletekkel, több mint 30 000 képpel és szöveges dokumentumokkal (<http://www.dlib.org/dlib/april96/loc/04c-arms.htm/>).

A gyűjtemények egységesek, de integritásuk változó. Egy részük koherens archív anyag egy-egy személyről vagy intézményről; a második csoportra a speciális eredeti forma jellemző, pl. eredeti dagerrotípiák gyűjteménye; a harmadik csoportot a tematikus antológiák jelentik. A gyűjtemények mérete is változó, vannak fotógyűjtemények 600 és 25 000 képpel. Új irányként jelentkezik az oktatási gyűjtemények képzése.

A program többek között a következő problémákat vetette fel, és az alábbi tanulságokkal szolgált:

1. *A történelmi gyűjtemények digitalizálása nem szokványos feladat.* A szabványos letapogatóbereendezések (szkenner) papírlapoldalak és diafilm digitalizálására készültek. Az ilyen munkára vállalkozó cégeknek nincs gyakorlatuk nagyformátumú negatívok vagy mikrofilm digitalizálásában, ugyanakkor a sérülékeny eredetiek helyett általában ezeket kapják meg.



A kötött művek letapogatására csak mostanában jelentek meg készülékek, ezek lassúak és drágák. (Már az eddigiekből is látható, hogy az LC külső cégekkel digitalizáltatta a gyűjteményt vagy annak egy részét, bár a felvételek a helyszínen készültek.)

2. *A tervezés, a projektmenedzsment és a minőségbiztosítás időt vesz igénybe.* A szkennelés csak a projekt egy része. Igen munkaidőfaló és szaktudásigényes a munkafolyamatok tervezése, a szerződések specifikálása, az anyagok előkészítése digitalizálásra, az előrehaladás figyelése és a minőség-ellenőrzés. A külső szerződések részletes utasításokat kell elkészíteni, és ezek végrehajtását kisebb mintákon kell ellenőrizni a kellemetlen meglepetések elkerülésére.

3. *Az elnevezési sémákat korán kell megalkotni.* A hálózati erőforrások konzisztens, szervezett eléréséhez robusztus, általános elnevezési sémák szükségesek. Egyedi és állandó azonosítókat kell használni a katalógustételeket és más leíró elemeket a tárolt dokumentummal való összekapcsoláshoz. A korai elnevezési rendszer lehetővé teszi a digitalizálás és a leíró anyagok párhuzamos elvégzését, megkönnyíti a projekt- és minőség-ellenőrzést minden szinten.

Az LC által kifejlesztett általános megközelítésben minden tétel két részből álló logikai nevet kapott, az egyik a gyűjteményt, a másik azon belül a tételt azonosítja. A digitális archívum hierarchikus rendbe szervezett állományok halmaza, és a logikai nevek egy táblázatban kereshetők. A tervek szerint a jövőben a logikai nevek lesznek az ún. Uniform Resource Name azonosítók, amelyek globálisan egyediek, állandóak és helyfüggetlenek. A Corporation for National Research Initiatives (CNRI) az LC-vel együttműködve kialakított egy olyan rendszert, amelyben a lokális nevekhez ún. *handle* (kapaszkodó?) adódik. A CNRI handle-szerver fogja feloldani az URN-eket, és leképezi azokat a világhálón is értelmezett speciális helyfüggő azonosítókká, mint amilyenek az URL-ek.

4. *A leírások egyneműsége nem praktikus.* A digitális másolatok igénylik a keresés és a tallózás támogatását az azonosítást megkönnyítő leírásokkal. Ez időigényes feladat, megalkotásuk jelentős kutatómunkát is igényelhet. Még az LC-re is igaz, hogy a legtöbb archív anyag nincs tételszinten katalogizálva.

Sok gyűjteménynek van egyfajta (gyakran nem is publikált) leírása (*finding aid*), amely tartalmazza a gyűjtemény célját, tartalmát, szellemi kontextusát stb., és egy strukturált listát az elemekről. (Létezik egy, a Berkeley Finding Aid Projectből származó szabványtervezet is, az Encoded Archival Description, amely egy SGML DTD az ilyen *finding aid* típusú leírások számára.)

Egy másik megközelítés a ún. csoportszintű katalogizálás, amely a tételek logikai csoportjaira vonatkozik. Ez gyakran elegendő a felhasználó számára, míg a tételes katalogizálás keresztülvihetetlen. (Néha gyakorlati okokból: a Berkeley egyetem számításai szerint 3,5 milliós képanyaguk katalogizálása jelenlegi munkaerő-kapacitásukkal 400 évig tartana.) Tehát az archiválási gyakorlatot kell integrálni az automatikus rendszerekbe. A probléma az, hogy ez a gyakorlat nem egyforma, és gyakran jó indokkal. A következtetés egyértelmű: a digitális archívumot azt felismerve kell építeni, hogy a leírások szintje és szerkezete (a metaadatok forrásai) igen változatos.

5. *Hol kell tárolni a leíró információkat?* Az LC modell elválasztotta a hozzáférés eszközeit (indexek, katalógusok, leírások) a forrásanyagot tartalmazó digitális archívumtól. Így az utóbbira közvetlenül rámutathatnak az előbbieket. A modell azt is tartalmazza, hogy a digitális tétel mellett más változatok (film, papír stb.) is létezhetnek, és természetesen az eredeti is. (A modellalkotás számos egyéb kérdést is felvet, amelyekre nem térünk ki.)

Az LC projekt során gyakran felvetődött kérdés volt a digitális másolat eredetiségének megőrzése. Ennek két aspektusa is van: mennyire hű másolat az eredetihez képest, és szabad-e az olvashatóság érdekében technikai javításokhoz – pl. kontrasztkiemeléshez – folyamodni.

Az LC a technikai kereteit lépésről lépésre fejlesztette ki, úgy, ahogy azok jelentkeztek, és úgy vette használatba az új technikákat, szabványokat és a nemzeti infrastruktúra újdonságait, ahogyan azok megjelentek.

Az LC dokumentumait és könyveit bitmapképekké alakították, és ezek képezték a kereshető változatok alapját is, amelyeket SGML-ben jelöltek az ún. Text Encoding Initiative által meghatározott dokumentumtípus-definíciót használva. Mivel a szerződő feleknek előírták a 99,95%-os karakterpontosságot a kereshető szövegekre, ők mindannyian a kézi beírást választották (1996-ig) az OCR technika helyett.

### **A Digitális Könyvtári Kezdeményezés (Digital Library Initiative) projektjei**

Az NSF, az ARPA és a NASA által támogatott, és hat egyetemi kutatási-fejlesztési projektet felölelő kezdeményezés célja: létrehozni digitális gyűjteményeket, az ezekhez szükséges technológiát, és az egészet beilleszteni a globális információs infrastruktúrába. A projekt 1994 őszén kezdődött, és négy év alatt kell végrehajtani. A támogatás egyenként és évente 1 millió USD. Miközben mindegyik fejlesztő csoport elvégez egyfajta „prototípus- vagy tesztrendszer- (testbed) vizsgálá-

tot", kutatási terveik, módszereik, gyűjteményeik és megcélzott használói körük lényegesen különböznek.

A projekt támogatók kívánságára az összes program közös tevékenységei közé tartozik a felhasználókra vonatkozó kutatás. A digitális könyvtár és a multimédia-dokumentumok használata új helyzetbe állítja a felhasználót is, így a programok azt is vizsgálják, hogy miként reagálnak a felhasználók az új technikákra, illetve miként lehet az információkeresést kényelmesebbé és hatékonyabbá tenni. A hat projekt a következő:

#### **Carnegie Mellon Egyetem *Informedia* Digital Video Library projekt**

Célja létrehozni egy interaktív, online digitális videókönyvtárat, amelyben a felhasználók archivált tudományos és matematikai videoanyagokhoz tudnak hozzáférni. Hanganyagot, képeket, természetnyelv-megértő technikát integrálva, különböző helyekről származó, különböző tartalmú 1000 órányi videoanyagot dolgoznak fel. A kutatás kiterjed az ember-gép kapcsolatra, a digitális videó használatának árképzésére, a személyiségi jogokra és a biztonságra (<http://informedia.cs.cmu.edu>).

#### **Berkeley Egyetem *Electronic Environmental Library* projekt**

Nagyméretű, osztott könyvtárat kíván létrehozni, amely egyaránt tartalmaz képeket, szöveget, térképeket, hangot, videofilmeket, numerikus adatbázisokat, hipertextesített kompozit dokumentumokat a környezetgazdálkodás témakörében, kifejezetten olyan hozzáférési technikával, amelyet a képzetlen olvasó is használni tud.

A kutatási témák közé tartozik az automatizált indexelés és intelligens visszakeresés, a videoalkalmazásokat támogató adatbázis-technika, kliens-szerver protokollok, osztott erőforrás-keresés, a felhasználó és a rendszer közti párbeszéd új paradigmái. A felhasználókkal kapcsolatos vizsgálatok elsősorban a szakemberekre, és nem a nagyközönségre irányulnak.

A végső soron a HTTP protokollra, illetve a hámlóra alapozott program korszerű feldolgozási és keresési eljárásokat épített a tesztrendszerbe. Az igen vegyes állomány egy része csak a kutatásfejlesztést szolgálja, de nagy része szabadon elérhető (<http://elib.cs.berkeley.edu>).

A szöveges keresést a Chesire II-re alapozták (<http://chesire.lib.berkeley.edu>), amely egy új generációs rendszer, a Z39.50-nel és az Infobusszal is együttműködik.

#### **Santa Barbara Egyetem *Alexandria* projekt**

Felhasználóbarát digitális könyvtár létrehozása térképek, képek és térinformatikai adatok archi-

vumaiból. Az adatok és dokumentumok kezdetben az egyetem környezetére vonatkoztak, és ezt szélesítik a program során, bevonva más könyvtárakat is. A rendszerben szükség szerint helyeznek el térinformatikai interfészeket, elektronikus katalógusokat és az adatkeresés eszközeit. Kutatási területeik: térinformatikai adatok indexelése, amely széles körű keresést tesz lehetővé; a nagy adat-tömeg miatti teljesítményproblémák; térinformatikai metaadatmodellek; a térinformatikai rendszerek követelményei a felhasználó szemszögéből.

#### **A Stanford Egyetem *Infobus* nevű digitális könyvtár projektje**

A résztvevők között a Stanford Egyetem figyelmét a rendszerek együttműködő-képességére fordítja. Olyan technikákat dolgoznak ki, amelyek egyaránt hozzáférést nyújtanak mindenféle információ forráshoz, legyen az személyes, hagyományos könyvtári gyűjtemény, vagy kutatók nagyméretű adathalmaza. Gyűjteményük elsősorban a számítástechnikai szakirodalom, de a világhálót mint egészet tekintik. A projekt magja az Infobus protokollt futtató próbarendszer. Ez biztosítja a különböző szolgáltatásokhoz és információ forrásokhoz az egységes hozzáférést ún. „proxikon keresztül, amelyek fordítóként működnek az Infobus protokoll és az eredeti protokollok között. Kutatási témáik felölelik az információelosztó és kommunikációs modelleket, az információ kliens-interfészeket, az információkereső rendszereket és a kereskedelmi tranzakciós modelleket, ezek jogi és közgazdasági vonatkozásait.

#### **Michigan Egyetem *Digital Library* projekt**

Tartalmilag a föld- és űrtudományokra irányul. Több ezer felhasználót és információforrást kíván összekötni (felsőoktatási campusokat, főiskolákat, közkönyvtárakat). Rendszerbe kívánja foglalni azt a hatalmas tömegű információt, amely a hálózaton található, és a legkülönbözőbb formátumú, a könyvtári katalógusoktól az enciklopédiákon, szótárakon, folyóiratcikkeken és adatbázisokon át a levelezőcsoportokig (<http://www.si.umich.edu/UMDL>). A Hálón már Java-alapú oldalakkal jelennek meg, az Artemis interfészt ajánlják.

Egyik mellékprogramjuk az ún. Journal Storage projekt, amely 10 közgazdasági folyóirat összes számának digitalizálását tűzte ki célul (<http://index.umd.umich.edu/jstor>). Tesztrendszerük a TULIP interfészen alapszik, amely az Elsevier Science anyagtudományi lapjainak kereső és böngésző felülete.

Kutatási területük többek közt az intelligens ágensek és a rendszerintegráció. A tesztrendszerben háromféle egymással együttműködő intelligens ágenst dolgoztak ki, az első a felhasználóval párbeszédet folytat annak információigényéről, a

közvetítőagens koordinálja a kereséseket a távoli hálózati forrásokban, a gyűjteményágensek a különböző gyűjteményekre jellemző specifikus kereséseket végzik.

### Illinois Egyetem Interspace Digital Library projekt

Az egyetem műszaki közössége számára fejleszt digitális könyvtárat. A gyűjtemény magját az SGML formátumú folyóiratok képezik. Kialakítanak egy „könyvtártól a közösség felé” modellt: olyan gyűjteményt, amelyben a dokumentumoknak egyedi nevük van (*handle*), szerepel a szerzők által generált indexelés (*metaadat*), egyenrangú állomások közötti architektúra, amely támogatja a többszörös forrásból származó hálózati objektumok kölcsönhatását és analizését.

A kutatási témák közé tartozik még a tesztrendszer szociológiai értékelése, a szemantikai dokumentum-visszakeresés műszaki fejlesztése, a jövő skálázható információs rendszereinek prototípus-fejlesztése, az Interspace.

A DL jól szervezett alkalmazott kutatási program. A jelentős központi támogatás miatt pénzügyi és szakmai elszámolási rendszere alaposnak tűnik. Ez utóbbi a kívülállónak is észlelhetően megnyilvánul a résztvevők rendszeres munkáuléseiről, konferenciáiról és időszaki jelentéseiről készült Hálóoldalakon.

### Néhány egyéb kezdeményezés az Egyesült Államokban

1. Az Amerikai Egyesült Államok Oktatási Minisztériuma (Department of Education) finanszírozásában 1996 szeptemberében kezdődött meg a háromévesre tervezett National Digital Library of Theses and Dissertations, később **Networked Digital Library of Theses and Dissertations** (Szakdolgozatok és Disszertációk Hálózati Digitális Könyvtára, NDLTD, <http://www.nltld.org>). A kezdeményezés a SURA (Southeastern Universities Research Association, a Délkeleti Egyetemek Kutatási Szövetsége) sikeres programjára épül.

Az NDLTD elsősorban oktatási célú projekt: egyetemi hallgatóknak kívánja megtanítani, hogyan legyenek elektronikus kiadók, és hogyan használhatják a kutatásban a digitális könyvtárakat, azaz végső soron hogyan válhatnak az „*information literate*” (*információértő*) fogalmának megfelelő értelmiségekké. Az NDLTD-hez való hozzáférés lehetővé fogja tenni, hogy a hallgatók könnyen megtalálják az őket érdeklő szövegeket, elolvassák társaik szakdolgozataiból a szakirodalmi szemléket, és hipertextkapcsolókat (linkeket) kövessenek a releváns adatok és eredmények megtalálása céljából. Az oktatók rámutathatnak a

szakterületükön folyó kutatás legjobb példáira egészen a konkrét táblázatok, illusztrációk stb. szintjéig.

Az első évben húszra nőtt a résztvevők száma. A projekt vezető intézménye, a Virginia Tech Egyetem úgy döntött, hogy a Cornell Egyetemen a Networked Computer Science Technical Report Library (Számítástechnikai Műszaki Jelentések Hálózati Könyvtára, <http://www.ncstrl.org/>) projekt számára kifejlesztett Dienst rendszert adaptálják.

Az IBM cég több számítógép adományozásával járult hozzá az NDLTD megvalósításához. Az IBM Digital Library szoftverrel futó egyik gép 4 terabájt, ami 40 millió átlagos terjedelmű elektronikus disszertáció tárolására elegendő. Egy másik IBM gépen az OCLC SiteSearch fut. Az OCLC ingyenesen bocsátotta rendelkezésre a WorldCat adatbázis szakdolgozatokra és disszertációkra vonatkozó több mint egymillió MARC rekordját. Ez jól kiegészíti a projektben szintén részt vevő UMI anyagait.

A Virginia Tech hallgatóinak egy sor eszköz áll rendelkezésére:

- az elektronikus szakdolgozatokra és disszertációkra kifejlesztett Document Type Definition, azaz az SGML-en alapuló dokumentumszerkesztési módszerrel készített implementáció;
- MS Word űrlap a Microsoft SGML Author for Word használatához (a Microsoft adománya) elektronikus szakdolgozatok és disszertációk SGML-verziójának elkészítéséhez;
- Word Perfect és AuthorEditor űrlapok szakdolgozatok és disszertációk SGML-verziójának elkészítéséhez;
- LaTeX űrlap a LaTeX2e használatához, és SGML-verziók készítéséhez;
- stíluslapok elektronikus szakdolgozatok és disszertációk SGML-verziójának nézegetéséhez;
- SGML–HTML konverter elektronikus szakdolgozatokhoz és disszertációkhoz;
- multimédia oktatóanyag a PDF-eszközök használatáról (az Adobe az Acrobat szoftvert ajánlódokta a projektnek) [6].

2. A Cornell Egyetem számítástudományi tanszéke neve alatt fut a **Networked Computer Science Technical Reports Library**, amely az ARPA által szponzorált, elosztott könyvtár. A számítástudományi műszaki jelentések gyűjteménye a részt vevő intézmények, ipari és kormányzati kutatóbázisok anyagait tartalmazza, és együttműködő szerverek hálózatán van elhelyezve.

3. A *Michael Hart* nevével fémjelzett **Gutenberg Project** a hálózati irodalmi könyvtárak őse. Hart célja 2001-re 10 000 szépirodalmi művet elhelyezni a hálózaton. A szövegek ASCII-szövegállományok, és valamennyi ingyenesen



hozzáférhető, ennek megfelelően többnyire 70 évnél régebbi művekről van szó. E programnak van esélye arra, hogy a legnagyobb hatást gyakorolja a közönségre.

## Japán

1993 nyarán a japán Kereskedelmi és Ipari Minisztérium, a MITI 7,45 milliárd jenes informatikai fejlesztési programot indított, amelynek része volt egy elektronikus könyvtár kiépítésére vonatkozó kísérleti program (<http://www.cni.org/regconfs/1997/ukoln-content/repor~17.html>, Matsumura 1997). Még ugyanabban az évben a Keio Egyetemen megnyitották az ún. Információs Infrastruktúra Központot, amelyben elindult a program. Két részből állt, az **Union Catalog Network Project**ből, és egy elektronikus könyvtár demonstrációs kísérletéből.

Az elsőről csak annyit, hogy 1997 februárjában 7,6 millió bibliográfiai rekordot tartalmazott. Az elektronikus könyvtár próbaprogramjában 1996-ig közel 9,5 millió oldalt digitalizáltak, a legkülönbözőbb információs forrásokból és típusokból, beleértve a National Diet Library egyes gyűjteményeit is. Hierarchikus tárolási rendszert és kísérleti multimédiás távközlési kapcsolatot építettek ki. 1996-ban a felügyelő intézmény, az Information Processing Technology Agency (IPA) egy értékelő kísérletet végzett 100 különböző könyvtárban, tartalomszolgáltatóknál, bizottsági tagoknál elhelyezett 100 terminálon keresztül.

Más úton indult az **Electronic Library System of the National Centre for Science Information Systems** (NACSIS-ELIS) prototípus-fejlesztés. Ez integrálta az 1987 óta működő NACSIS-IR-t, az intézmény létező bibliográfiai adatbázisát a digitalizált tudományos folyóiratokkal és konferenciakiadványokkal. Osztott feldolgozási környezetben működik, kliens-szerver kapcsolatban. Kiterjesztett változata Z39.50 protokollt használ a képadatok kezelésére. A fejlesztés még a nyolcvanas évek elejére nyúlik vissza, amikor egy Online Libraryt kezdtek fejleszteni optikai lemezen, faxterminállal. A sok intézmény részvételével folyó projektben 1993-tól 1997 márciusáig 62 folyóiratról kb. 800 000 oldalt digitalizáltak. 1997 végéig a próbaszolgáltatás ingyenes volt. 1998-ban tervezik a teljes körű hozzáférés indítását. Itt is, mint az előző kezdeményezésnél, a szerzői jogi kérdéseket menet közben igyekeznek tisztázni, emiatt késik a teljes nyilvános szolgáltatás bevezetése.

1997-ben az egyetlen működő egyetemi elektronikus könyvtár a Nara Institute of Science and Technology (NAIST) Mandala Library volt. Ebben az ismert elektronikus szolgáltatásokon kívül már megjelentek digitális könyvtárként tárolt anyagok.

Japán, koreai és thai nyelven írott, és ezeknek a nyelveknek a karakterkészletét használó szövegek többnyelvű böngészését teszi lehetővé a **Multilingual HTML Browser Project** keretében kifejlesztett HTML-böngésző (<http://mhtml.ulis.ac.jp/gwy/>).

## Kanada

A Kanadai Digitális Könyvtári Kezdeményezés (**Canadian Initiative on Digital Libraries**, CIDL, <http://www.nlc-bnc.ca/cidl/>) a Kanadai Nemzeti Könyvtár kezdeményezésére indult meg 1997-ben. Első lépésként kérdőívet küldtek ki a könyvtárak érdeklődésének és véleményének feltérképezésére, amelyre mintegy ötven válasz érkezett. 1997 márciusában húsz könyvtár részvételével tanácskozást tartottak, amelyen meghatározták a júniusban meghirdetett CIDL fő irányvonalait. Szeptemberre 53 különböző típusú könyvtár csatlakozott a kezdeményezéshez. Megválasztották a CIDL irányító testületét. Ennek feladatai a következők:

- a tervezési tevékenység irányítása;
- a tagság körének szélesítése, és a tagok aktív részvételének előmozdítása;
- a szakbizottságok (albizottságok) munkájának felügyelete;
- ajánlások kidolgozása a szakbizottságok és a tagság számára;
- a munkaterv betartásának ellenőrzése, és a tagság tájékoztatása;
- informálás a WWW-n a CIDL tevékenységéről és eredményeiről;
- a pénzügyi források kezelése;
- a CIDL hatékonyságának elemzése.

A CIDL tevékenységének korai szakaszában egy **Digitális Források Fejlesztése** (*Development of Digital Resources*) elnevezésű albizottság működik a következő feladatkörrel:

- ajánlások készítése a nemzeti és helyi érdekeket egyaránt figyelembe vevő koordinált szolgáltatás- és gyűjteményfejlesztési megoldásokra;
- útmutatók kidolgozása a digitalizálás gyakorlata, a szerzői jogok kezelése, az archiválás témakörében;
- a hozzáférési politika és gyakorlat legelfogadottabb és leghatékonyabb módszereire vonatkozó ajánlások kidolgozása;
- kapcsolatfelvétel és konzultációk a kapcsolódó területekkel (többek között kiadókkal, levéltárakkal, múzeumokkal, kutatóintézetekkel).

A **Hozzáférés és Megőrzés** (*Access and Preservation*) albizottság feladatai:

- a felhasználói hozzáférés által megkövetelt technikai előfeltételek összegzése;

- útmutatók kidolgozása a digitális források leölyhelyének azonosításához, és a források leírásához szükséges minimális metaadatokra;
- útmutatók kidolgozása a megörzés, a szerzői jogok technikai kezelése legelfogadottabb és leghatékonyabb módszereire;
- stratégia kidolgozása a szükséges kalauzok, útmutatók (directoryk) összeállítására;
- a megfelelő gateway-ek és csereprotokollok kifejlesztésének vagy alkalmazásának elősegítése;
- kapcsolatfelvétel a kapcsolódó területekkel.

A *Reklám és Finanszírozás (Promotion and Funding)* nevet viselő albizottság feladatai közé tartozik:

- a digitális könyvtárakkal kapcsolatos költségek meghatározása és racionalizálása;
- a források nemzeti és helyi szintű biztosítása stratégiájának és legelfogadottabb, leghatékonyabb módszereinek azonosítása;
- az együttműködő finanszírozás modelljének a kidolgozása (a kezdeményezés közös finanszírozása a kanadai könyvtárak vagy azok egy részének együttműködésével);
- koordinált reklámozási stratégia kialakítása;
- támogatási-képviselési (advocacy) stratégiák kidolgozása és alkalmazása a helyi és tartományi kormányzatok meggyőzése (lobbyzás) érdekében;
- kapcsolatfelvétel a kapcsolódó területekkel.

Az *Oktatás és Kutatás (Training and Research)* albizottság feladatai a következők:

- képzési és továbbképzési programok kidolgozása;
- kommunikációs stratégiák és források kialakítása;
- a digitális könyvtári oktatás integrálása a meglévő képzési programokba;
- alkalmazott kutatási területek felderítése;
- kapcsolatfelvétel a kapcsolódó területekkel.

A CIDL mind ez idáig a kidolgozott akciótervnek megfelelően működött. Az előrehaladást a meghirdetés után egy évvel, 1998 nyarán vizsgálják majd meg részletesebben.

## Nagy-Britannia

Nagy-Britanniában több jelentős program fut, de a British Librarynek (BL) nyilvánvalóan döntő szerepe van. A British Library Digitális Könyvtári Programját már említettük a meghatározásokkal kapcsolatban. A *British Library Digital Library Programme* (<http://www.bl.uk/services/diglib>) a British Library Research and Innovation Centre vezetésével alapvetően a magánszektorral való együttműködésre épül, mivel a könyvtár vezetői úgy látják, hogy a költségek igen

nagyok, és nem fedezhetők kizárólag állami forrásokból. Ennek megvalósítására szolgál a *Public Private Partnership*, amelynek keretében nemzetközi tendert írtak ki a digitális könyvtár megvalósítására. A potenciális együttműködők 1997 októberében nyilváníthatták ki érdeklődésüket, majd november 24-ig nyújthatták be pályázataikat, amelyek értékelése nyomán alakulnak majd ki a végleges ajánlatok és az azokon alapuló szerződések. Ez a folyamat az ajánlatok nagyságának függvényében 1998 őszéig is eltarthat. A program prioritásai a következők:

- a BL dokumentumellátó szolgáltatásainak kiterjesztése digitális forrásokból származó, folyóiratcikkre vonatkozó jeladó információs szolgáltatásra, továbbá az eddiginél jobb rendelési és küldési szolgáltatásra alapozva;
- a szabadalmi szolgáltatások kiterjesztése;
- jobb hozzáférés a BL történeti gyűjteményeihez;
- a fenti fejlesztések integrálása a BL fő gyűjteményeibe és szolgáltatásaiba;
- a BL növekvő digitális gyűjteményeinek kezelését biztosító infrastruktúra létrehozása, és a brit kötelempéldány-szabályozás kiterjesztése a nem nyomtatott anyagokra.

A BL három forrásból kívánja építeni a digitális könyvtárat:

- a könyvtár meglévő anyagainak digitalizálása,
- vásárlás és licencek vásárlása,
- köteles példányok.

Feladat még a *Digital Library Public Private Partnership Project* ellátása információkkal azokról a technikai és üzleti követelményekről, amelyet a BL támaszt a digitális könyvtárral szemben; annak biztosítása, hogy a BL már meglévő digitális gyűjteményei továbbra is fennmaradjanak, azokat megőrizték és elérhetővé tegyék. Ez utóbbit igen fontosnak kell itélnünk, mivel a fejlesztés és szolgáltatás többé-kevésbé üzleti alapokra történő helyezése nem jelentheti a világban már meglévő digitális gyűjtemények elhanyagolását.

## Németország

A Deutsche Forschungsgemeinschaft támogatásával folyik az *Elosztott Digitális Tudományos Könyvtár* ([http://www.dbi.berlin.de/projekte/d\\_lib/](http://www.dbi.berlin.de/projekte/d_lib/)) kialakítása. Ennek alprojektjei:

- A tudományos könyvtárak modernizálása és ésszerűsítése (Modernisierung und Rationalisierung in wissenschaftlichen Bibliotheken).
- Elektronikus publikálás a tudományos könyvtárak szakirodalmi és információs szolgáltatásában/kínálatában (Elektronische Publikationen im Literatur- und Informationsangebot wissenschaftlicher Bibliotheken).



- Könyvtári állományok retrospektív digitalizálása (Retrospektive Digitalisierung von Bibliotheksbeständen).

A **Tudományos könyvtárak modernizálása és ésszerűsítése** alprojekt különösen szoftverek és eljárások fejlesztését támogatja, ha azok

- releváns módon alkalmazhatók a könyvtári szolgáltatásokra;
- tudományos kutatásokon alapuló olyan alkalmazások prototípusait valósítják meg, amelyek nem kínálhatók mint a piaci versenyben részt vevő termékek;
- egyszeri könyvtári kísérleti programok és modellek kialakítására vonatkozó innovációt képviselnek, amelyeknél várható más alkalmazók későbbi érdeklődése.

Az **Elektronikus publikálás a tudományos könyvtárak szakirodalmi és információs szolgáltatásában** alprojekt az elektronikus publikációknak a tudományt és kutatást szolgáló információs és szolgáltatási kínálatba való integrálását célozza.

A támogatandó projektek:

- az elektronikus publikációk régiók feletti szolgáltatását és használatát biztosító tartalomorientált szolgáltatások;
- az elektronikus publikációk régiók feletti szolgáltatását és használatát elősegítő előkészítő, implementációs és struktúraképző intézkedések;
- technikailag a piacon beszerezhető vagy public domain termékeken alapulnak úgy, hogy azokat a felhasználásnak megfelelően konfigurálják és módosítják, nem képezve azonban prototípusokat.

A program megvalósítása magában foglalja:

- az elektronikus publikációkkal kapcsolatos információs és feltáró eszközök kidolgozását,
- a feldolgozással kapcsolatos normák megalkotását,
- navigációs eljárások kialakítását,
- az elektronikus publikációk és hagyományos dokumentumok integrálását,
- szakdolgozatok és disszertációk elektronikus publikációját és szolgáltatását,
- a szakirodalom és más információs anyagok digitalizálását,
- digitalizált dokumentumok archiválását.

1997-ben indult a **Könyvtári állományok retrospektív digitalizálása** alprojekt, amelynek keretében elsősorban a szerzői jogtól nem védett anyagok digitalizálását tervezik. A mintegy 30 ezer kötetnyi digitalizálandó dokumentum között vannak a kulturális örökség részét képező, továbbá a gyakran használt és nehezen hozzáférhető, valamint az egyes szakterületek alapvető irodalmát jelentő művek.

Az előtérben tehát a következők állnak:

- a kutatás és oktatás szempontjából fontos anyagokhoz való közvetlen hozzáférés,
- többszörös hozzáférés a sokat használt szakirodalomhoz,
- nehezen hozzáférhető anyagok digitális szolgáltatása,
- eddig kevésbé ismert anyagok szélesebb körű használata.

Az **Elosztott Digitális Tudományos Könyvtár** program keretében nemcsak eszközökre, hanem időleges állások létesítésére is lehet pályázni. Eszközöket azonban csak akkor finanszíroznak, ha azok a kísérleti üzemeltetés elvégzéséhez elengedhetetlenül szükségesek, és a meglévő alapfelszereltséggel nem valósíthatók meg.

A modell- és kísérleti projekteknek olyan fejlesztéseket kell szolgáltatniuk, amelyek e meglévő megoldásokhoz alternatívákat kínálnak, és később rendszeres használatukra is sor kerülhet. Ezekhez a DFG kezdeti finanszírozást nyújt, a projekteknek azonban önfinanszírozókká kell válniuk.

A modellek meghatározott innovatív kérdésfeltevésekre kínálnak olyan szabványos válaszokat, amelyeket jelentős kiegészítő illesztési, adaptálási munkák nélkül át tudnak venni más könyvtárak. Az előtérben a helyi és regionális felhasználáson túllépő modellek kialakítása áll. Döntő kérdés, hogy a modellek gyártótól függetlenek, és más rendszerekre átvihetők legyenek. Szükséges az is, hogy sok könyvtárban legyen megállapítható igény az adott modell felhasználására.

A kísérleti projektek eddig még nem létezett vagy új megoldások keresését jelentik. Végső céljuk szintén a későbbi felhasználásra alkalmas megoldások kimunkálása, elsősorban azonban új elképzelések gyakorlati megvalósíthatóságának bizonyítására irányulnak. Ennek megfelelően megengedhető, hogy a későbbi felhasználóknak nagyobb mértékben kelljen alkalmazkodniuk a kidolgozott megoldásokhoz.

Döntő kérdés, hogy mekkora a digitalizálandó állományok jelentősége a tudomány és kutatás számára. Ez a következők függvénye:

- az adott állomány tudományos értéke,
- viszonya az aktuális kutatási trendekhez,
- eddigi és várható jövőbeni használatága,
- a felhasználhatóság mértékének várható növekedése,
- a nemzetközi kutatási és digitalizálási trendekhez és projektekhez való kapcsolódása.

Biztosítani kell a digitalizálás során keletkezett nyers adatok hosszú távú megőrzését. A támogatott projektekben keletkezett állományokat általában ingyenesen kell szolgáltatni. Ingyenes és korlátozás nélküli hozzáférést kell biztosítani a metaadatokhoz. A technikai megvalósítás során a digitális képek mellé a dokumentumok fontos ele-

meinek (címlapok, tartalomjegyzékek, fejezetcímek, mutatók) teljes szöveges feldolgozására vagy tartalmi feltárására is szükség van. További előrelépést jelent a bibliográfiák szöveges és tartalmi feltárása annak lehetővé tételével, hogy a bennük tükrözött dokumentumokhoz építsenek ki kapcsolatokat.

A képként történő digitalizálás helyett alkalmazható a dokumentumok teljes szöveges bevitelle. Az első lépésben ez célzott szöveges keresésekhez teljesszöveg-keresési indexként szolgál, ezért megengedhető a bevitt szöveg bizonyos fokú hibás volta.

Második lépésként az adott szöveg elektronikus kiadásának tekinthetők a bevitt szövegek a szöveg szerkezetének (pl. SGML formátumú) leírásával.

A program megengedi a szöveg bevitelének bármely formáját, de tipikusnak a képként való bevitelnek kell lennie a fontos tartalmi elemekkel való kiegészítéssel. A további szövegfeldolgozás vagy SGML-strukturálás kivételnek tekintendő. A képernyőn történő megjelenítés mellett lehetőséget kell teremteni a jó minőségű nyomtatásra.

## Irodalom

- [1] MATSON, L. D.–BONSKI, D. J.: Online, 1997. november.  
<http://www.onlineinc.com/onlinemag/NovOL97/matson11.html> – 13-Oct-97.
- [2] <http://www.bl.uk/services/ric/diglib/diglib.html>
- [3] SCHATZ, BRUCE, CHEN, HSINCHUN: Building large-scale digital libraries. = Computer, 1996. május. <http://www.computer.org/dli/>
- [4] Denmark's Electronic Research Library. Skrifter fra Statens Bibliotekstjeneste 16. Copenhagen, 1997. <http://www.kum.dk/publ/def/eng-version/>
- [5] HAKALA, J. et al.: Warwick framework and Dublin core set provide a comprehensive infrastructure for network resource description.  
<http://www.ub2.lu.se/tk/warwick.html>
- [6] FOX, E. A. et al: Networked digital library of theses and dissertations. An international effort unlocking university resources = D-Lib Magazine, 1997. szeptember.  
<http://www.dlib.org/dlib/september97/theses/09fox.html>

Beérkezett: 1998. V. 6-án.

# TELDAN ONE SOURCE FOR ALL YOUR INFORMATION NEEDS

Information Systems Ltd.



## A Leading Information Supplier To Libraries:

- Academic
- Corporate
- Government
- Research
- Defense
- Medical

- Journal subscriptions to 180,000 titles
- Electronic journals
- CD-ROM & Online databases
- Global defense information on CD-ROM
- **Info 99** - The 14th International Conference & Exhibition, Tel-Aviv

*Inviting you to meet us at our stand at The 30th Annual Meeting  
Of The Hungarian Librarians' Association - 30.7-1.8.1998*



For more details:

TELDAN INFORMATION SYSTEMS

P.O.B. 18094 Tel-AVIV 61180, ISRAEL TEL. 972-3-6950073 FAX. 972-3-6956359 URL: <http://www.teldan.com> E-mail: [teldan@teldan.com](mailto:teldan@teldan.com)