



Nemzetközi körkép a webarchiválás gyakorlatáról

NÉMETH Márton

Bevezetés

A webarchiválás olyan dinamikusan fejlődő terület, mely számos vonatkozásban már a korábbiakban is felbukkant a Könyvtári Figyelő hasábjain, különösen a nemzetközi szakirodalom szemlézése kapcsán. (Például 2014-ben *Hegyközi Ilona* tekintette át a webarchiválással kapcsolatos nemzetközi trendeket.) Úgy éreztük, eljött az ideje egy újabb összegzésnek. Ennek különös hangsúlyt ad, hogy számos korábbi kezdeményezést követően, idén tavasztól megteremtődtek az alapjai az OSZK fejlesztési projektjén belül egy olyan kísérleti projekt elindításának, melyben felmérjük a webarchiváláshoz szükséges hardver és szoftver igényeket, valamint szakmai ismereteket. A fő cél, hogy jól megalapozott módon integrálni tudjuk e területet hosszú távon is az OSZK szolgáltatási tevékenységei közé.

Az OSZK Elektronikus Könyvtári Szolgáltatások Osztályán létrehoztunk egy Magyar Internet Archivum honlapot (<http://mekosztaly.oszk.hu/mia>), melyen tanulmányozhatók a webarchiválás különféle módszerei, alapfogalmai, meg a nemzetközi szakirodalom. Továbbá a projekttel kapcsolatos aktuális információkkal is szolgálunk és fel lehet iratkozni a webarchiválás szakmai kérdéseit tárgyaló levelezőlistára is.

Ennek a cikknek nem az a célja tehát, hogy a webarchiválási tevékenységek szakmai alapjait járja körül (amelyre a honlapot böngészve nyílik lehetőség), hanem, hogy áttekintést adjunk a webarchiválási szolgáltatásokat megalapozó nemzetközi jó gyakorlatokból.

Az Európai Unió belül már csak igen kevés tagállam maradt, amelyek még nem léptek ebbe az irányba. Az európai példák mellett a kísérleti projekt előkészítése közben érdemes volt felmérnünk a teljes összkép kedvéért néhány Európán kívüli ország szakmai gyakorlatát is. A következőkben rövid ízelítőt nyújtunk a 2017 tavaszán elvégzett szakirodalmi elemzés eredményeiből. Áttekintjük a webarchiválás néhány kiemelt nemzeti modelljét, amelyekből tanulságokat meríthetünk az itthoni gyakorlat kialakítása kapcsán.

Ki kell emelnünk a webarchiválás kapcsán a legtöbb általunk is használt szoftver fejlesztésében vezérszerepet játszó, s módszertani irányítóként is működő amerikai *Internet Archive* tevékenységét. Sokszínű együttműködést folytatnak az egyes nemzeti keretrendszerek képviselőivel. Az egyes nemzeti modellek tulajdonképpen arra válaszul jöttek létre, hogy a web fejlődése egyetlen szervezet számára már követetlenné vált archiválási szemszögből. A nemzeti szereplők, illetve az *Internet Archive* egymást kiegészítő tevékenységei adják meg a webarchiválási szolgáltatások teljességét. A nemzeti webarchiválási tevékenységeket pedig az IIPC nemzetközi konzorcium (*International Internet Preservation Consortium*, <http://netpreserve.org>) fogja közös keretbe. Hamarosan az OSZK is tagja lesz ennek a konzorciumnak. Ennek révén hatékony tudásmegosztási, kompetencia fejlesztési és szakmai együttműködési csatornák válnak elérhetővé számunkra is.

Előljáróban az egyes nemzeti modellek rövid ismertetése előtt érdemes leszögezni, hogy a nemzeti könyvtáraknak szinte minden vizsgált esetben központi szerepköre van a webarchiválási szolgáltatások szervezésében. E tevékenységek szervezeti keretei azonban rendkívül változatosak. A siker biztos garanciája a tartalomszolgáltatók és közgyűjtemények közötti széleskörű összefogás. A webarchiválás közös ügy, nem egyetlen intézmény felelősségi körébe tartozó szolgáltatási feladat. Az adott ország közigazgatási szerkezete, a közművelődési intézményi struktúra jellege alapvetően meghatározza a webarchiválási tevékenység szervezeti kereteit. Másrészt érdemes utalnunk arra is, hogy a webtechnológia robbanásszerű folyamatos fejlődésével egyre nagyobb kihívást jelent a különféle webes tartalomszolgáltatások archiválhatóságának biztosítása mind a tartalomszolgáltatók, mind az archiváló intézmények részére. A fejlődési trendeket figyelve megjelent egy olyan irány, hogy a tartalmakat egyre nehezebben begyűjteni képes aratórobotok helyett a böngészőprogramok működését

imitáló, a böngészőmotoron alapuló új szoftverek jelennek meg, ahol már a felhasználó képernyőjén látható tartalom rögzítése válik céllá. Ez részleges paradigmaváltást is hozhat a jövőben a webarchiválás területén, alapvetően átalakítva a továbbiakban ismertetett nemzeti modellek gyakorlati kereteit.

PANDORA: Az ausztrál webarchiválás modellje

Általános keretek, gyűjtőkör

Az Ausztrál Nemzeti Könyvtár 1996-tól végez webarchiválási tevékenységet a PANDORA: Ausztrál Webarchívum keretei között (részletes összefoglaló a gyűjtőköri alapelvekről: <http://pandora.nla.gov.au/selectionguidelines.html>).

A webarchiválás keretei Ausztráliában némiképp sajátosan alakultak ki. A tartalmak válogatott archiválása nem egyetlen szervezet, hanem egy intézményi konzorcium keretei között zajlik. Ennek tagjai az egyes tagállamok állami könyvtárai, illetve számos tudományos könyvtár és archívum (ideértve az Ausztrál Háborús Archívumot, továbbá a filmarchívumot is). Egy nemzeti osztott rendszerbe tölti be mindegyik intézmény a saját szelektív módon learatott tartalmait. Az aratások gyűjtőköre a helyi igények szerint intézményi szinten kerül szabályozásra. Az egységes webarchiválási szabványkörnyezet kialakítása, a rendszer infrastruktúrájának működtetése, az együttműködési keretek kialakítása nemzeti könyvtári hatáskör.

A szelektív archiválási tevékenységek mellett 2005-től kezdve a teljes ausztrál domain learatására is sor kerül. Ezt a tevékenységet az Ausztrál Nemzeti Könyvtár megbízásából az *Internet Archive* végzi el.

A szelektív begyűjtés módszerei

A konzorcium a szelektív mentésekre összpontosít. Ennek keretei között inkább a tartalom begyűjtésének minőségére, a megfelelő tartalmi mélységű aratás biztosítására fókuszálnak a mennyiségi paraméterek helyett. A hivatalos kormányzati dokumentumok tagállamonként kerülnek learatásra, illetve beszolgáltatásra. Ezt az anyagot egészíti ki az ausztrál élet teljességéről számot adó szelektív aratások köre (konferenciák, oktatási anyagok, zenei élet webhelyei, vallási oldalak, a napilapok webhelyeinek egyedi webes tartalmai stb.). A kiadókkal történő egyeztetés során törekedni kell arra, hogy a már kereskedelmileg nem értékes szolgáltatásokhoz kapcsolódó weboldalak teljes köre szabadon elérhető legyen a

nagyközönség számára. Ugyancsak egyeztetések szükségesek a keresőmotorokkal nem leartható tartalmak begyűjtéséhez is. Az egyes webhelyeket leíró adatokat igyekeznek feldolgozni s beilleszteni az Ausztrál Nemzeti Bibliográfiába.

Összegzés

Összefoglalásként elmondható, hogy a webarchiválás keretét adó konzorciumi szervezeti forma, bár némiképp a később tárgyalandó britre emlékeztet, mégis egyedi jelenség. A British Library jóval erősebb koordináló szakmai szerepet tölt be, mint az Ausztrál Nemzeti Könyvtár. A brit nemzeti intézmény a konkrét webaratási tevékenységekből is jobban kivieszi a részét, különösen, hogy a teljes nemzeti webtér aratást az Internet Archive látja el az ausztrálok részére. A gyűjtőkori alapelvek világosan lehatároltak, egységes egészként kezelik a teljes gyűjtőkori tevékenységet, melynek szerves részeként jelenik meg a webarchiválás.

Webarchiválás Nagy-Britanniában

Általános keretek, gyűjtőkör

Nagy-Britanniában a szelektív webarchiválás (<http://webarchive.org.uk>) 2004-ban indult el a British Library gondozásában, tematikus gyűjtemények létrehozásával. A teljes .uk webteret 2013-tól kezdték el aratni, összhangban az akkor megjelent új köteleispéldány szabályozással. A webaratási tevékenység a következő intézmények egyenrangú együttműködésével zajlik: the National Library of Scotland, the National Library of Wales, Cambridge University Library, the Bodleian Library in Oxford, the Library of Trinity College Dublin. A webarchívumhoz csak az említett intézmények fizikai tereiben lehet zárt dedikált hálózaton keresztül hozzáférni a szerzői jogi rendelkezésekkel összhangban. Az intézmények összeállítanak egy szűkített mintát is, melyet nyilvánosan elérhetővé tesznek. Az egyes intézmények a törvényi keretek figyelembevételével saját maguk alakítják a webarchiválás gyűjtőkörét. Az általános szempontok az alábbiak: tudományos jelentőség, innovatív webtechnológiai megoldások bemutatása, a brit élet minél teljesebb körű bemutatása szociális, kulturális, politikai, vallási, tudományos és gazdasági szemszögből. Külön-

leges méltánylást igénylő esetekben kérvényezni lehet egy adott webhely anyagainak teljes törlését is a webarchívumból.

Írányelvek

Az archívum weboldalán javaslatokat lehet megfogalmazni a szolgáltatásba bevonni kívánt webhelyekről. A dinamikus tartalmak, a csak belső keresőmotor által elérhető tartalmak (linkek nélkül), illetve a Javascript menüszervezetek némelyike kimarad az aratásból. Így is egy hónap átlagában 28 TB adatmennyiség kerül learatásra a Web Curator szoftver segítségével. Audio és videoanyagok (pl. Youtube, flash video) nem esnek bele a webarchiválási projekt hatókörébe. Az egyszerűen leartható hangfájlok viszont rögzítésre kerülnek. A könyvtár technikai ajánlásokat is megfogalmaz a weblap üzemeltetők számára az aratás megkönnyítése végett. A .uk domain hatókörébe eső weblapokat általános szabály szerint félévente aratják le, de a fontos weblapok aratására egyéni mérlegelés szerint ennél sűrűbben is sor kerülhet. Nem gyűjtik be azokat a tartalmakat, amelyek jelszóval védettek, illetve azokat a tartalomelemeket sem, amelyeket a webhely robots.txt fájlja kizár a megtekintésből.

Összegzés

Nagy-Britanniában igen jól szervezett webarchiválási tevékenység zajlik, számos közgyűjteményi résztvevővel, illetve a tartalomszolgáltatók bevonásával. Az egységes alapelvek szerint hatékonyan működő regionális szervezetrendszer hatékonyan egészíti ki a British Library országos koordinációs tevékenysége. Vezető szerepet játszanak a szakmai területet érintő szoftveres fejlesztések előmozdításában is.

Netarkivet.dk – dán netarchívum

Általános keretek

Dániában a koppenhágai székhelyű Dán Királyi (Nemzeti) Könyvtár, illetve az Aarhusban található nemzeti könyvtári rangú Állami könyvtár partnerségén alapul a webarchiválás modellje*. Az internet archiválást érintő gyűjtőkori és szervezeti keretek a két könyvtár együttműködésével külön szabályzatban kerültek rögzítésre: http://netarkivet.dk/wp-content/uploads/2015/10/Politik_for_indsamling_af_materiale_til_Netarkivet.pdf

* A két intézményt éppen napjainkban integrálják közös szervezeti keretek közé.

Írányelvek

A szabályozás irányelvei háromévente áttekintésre kerülnek a megadott, később részletezett törvényi kereteken belül. Maga a konkrét webarchiválási tevékenység az aarhusi intézményre összpontosul.

A webarchiválással foglalkozó munkacsoportban mindkét intézmény képviselteti magát gyakorlati szakemberekkel (IT és könyvtári oldal egyaránt). A munkacsoport a két főigazgató alárendeltségébe tartozik, mindkét intézményben külön digitális kurátorok foglalkoznak a begyűjtött, illetve begyűjtésre kijelölt anyaggal, állománnyal.

A Kulturális Minisztérium a nagy médiavállalatok és kutatási szakemberek bevonásával működtet egy tanácsadói munkacsoportot az audiovizuális (nem szövegalapú) tartalmak archiválására. Ez a tevékenység tehát a netarkivet keretein belül, de eltérő szabályrendszerben, külön ajánlások alapján történik a privát szereplők aktív részvételével. A képi, illetve hangzóanyagok learatásához a szokásostól eltérő szoftverkönyvet is igénybe vehető (pl. ftp protokoll használata médiatartalmak aratására a műsorszolgáltató webhelyéről).

Törvényi rendelkezések a begyűjtésre vonatkozóan

A gyűjtőkori szabályozás törvényi úton került rögzítésre (<http://pligtaflivering.dk/loven/index.htm> 3. fejezet 8–12. paragrafusok), ebbe épülnek be a webarchiválásra vonatkozó kitételek is. A törvény hatálya kiterjed az összes elektronikus kommunikációs hálózaton nyilvánosan elérhető dán tartalomra. Nem tartoznak tehát a törvény hatálya alá az intraneten, zárt hálózatokon elérhető tartalmak. Az adott tartalomnak mindig a nyilvánosság felé szántnak kell lennie. Egy olyan szolgáltatás például, ahová bárki regisztrálhat nyilvánosan, a törvény hatálya alá esik, a meghívással működő vagy teljesen zárt regisztrációhoz kötődő tartalmak viszont nem.

A „dán” fogalmának meghatározása a következők szerint történik: 1. Dániában bejegyzett domainelemekhez kötődő tartalmak (tehát nemcsak a .dk, hanem az összes dán regisztrátor által bejegyzett domainelemekhez kötődő tartalom). 2. A nem dán bejegyzésű domainelemek közül azokra terjed ki, melyeknek dániai célközönsége van. 3. A Kulturális Minisztérium egyedileg is kijelölhet a webarchiválás keretébe tartozó tartalmakat (8. paragrafus).

A Dániában regisztrált domain név tulajdonosának törvény által megszabott kötelezettsége az archivál-

hatóság követelményeinek megteremtése. Amennyiben nem Dániában regisztrált domainről van szó, akkor a tartalmat publikáló személynek, szervezetnek kötelessége a tartalmat archiválhatóvá tenni, archiválható formátumot előállítani (például a másolásvédetten szolgáltatott anyagok rendelkezésre bocsátásával). Az eredeti szöveg az „archiválható kiadás” kifejezést használja (9. paragrafus).

Amennyiben szükséges, rendelkezésre kell bocsátani a tartalom archiválásához elengedhetetlen hozzáférési jogosultságokat, felhasználóneveket, jelszavakat az archiválást végző intézmény részére (ilyenek lehetnek pl. a személyre szabott tartalom eléréséhez szükséges felhasználónevek, jelszavak). Az adott intézmény köteles ezeket az adatokat bizalmasan kezelni, nem adhat hozzáférést azokhoz külső személyek számára (10. paragrafus).

A dániai illetőségű domain név regisztrátoroknak az általuk regisztrált domain nevek adatait, illetve a domain nevek tulajdonosaira vonatkozó információkat elektronikus formában el kell küldeni a webarchiválásért felelős intézmény számára (11. paragrafus). Így könnyedén ellenőrizhető a begyűjtendő tartalmak köre.

A beküldés gyakoriságának mértéke nincs törvényileg rögzítve. A gyakorlatban háromféle aratási stratégiát alakítottak ki, melyeket az adott honlapok jellegéhez igazítanak. A legfontosabb hírértékű, tudományos értékkel bíró tartalmakat akár naponta aratják (80 honlap szerepel ebben a körben). Az általános gyakorlat az évi négy aratás elvégzése valamennyi, a törvény hatálya alá tartozó honlapról (cross-cutting harvest). Adott aktuális témakörökhöz kapcsolódó szelektív begyűjtést is folytatnak. Kiválasztanak évente három eseményt, s begyűjtik a releváns weboldalak tartalmait külön az eseményekhez rendezve (forrás: <http://netarkivet.dk/til-webstedejere/pligtaflivering/>).

Fontos kitétel az aratás folyamata kapcsán, hogy minden webhelyről teljes aratás történik. A robotok nem respektálják a tartalomszolgáltató által robots.txt fájlban rögzített korlátozásokat, hanem mindent learatnak! Az utolsó paragrafus pedig azt rögzíti, hogy a webarchiválási kötelezettség teljesítésével (esetleges járulékos tartalom előállításával) kapcsolatban felmerülő költségeket a tevékenységet végző intézménynek kell fedeznie (12. paragrafus). (A fent említett tevékenységek általában a statikusan aratható tartalom előállításának plusz kötelezettségét foglalják magukban a beszállító által a webaratást végző intézmény számára.)

Hozzáférés

A szerzői jogi és a személyes adatokat védő rendelkezések figyelembevételével adhat hozzáférést a két dán nemzeti könyvtár a leartott tartalom egyes szeleteihez a kérelmezők számára. Általában kutatási, oktatási céllal kérvényezett igényeket elégítenek ki.

Összegzés

Dániában széles társadalmi egyeztetés előzte meg a kötelezpéldány törvény vonatkozó rendelkezéseinek megalkotását, mely igen kemény kitételeket tartalmaz. Ennek révén viszont a webarchiválási tevékenység hatékonyan ellátható, a technikai feltételek biztosítását a domain név tulajdonosára, illetve a tartalom előállítójára hárítva át! A könyvtárak a webarchiválás tevékenységi kereteit a tartalomszolgáltatókkal aktív partnerségben dolgozták ki, s a rendszeres felülvizsgálat is az érdekeltek párbeszédével zajlik. Így elérhető a szigorú törvényi rendelkezések betartatása is az érdekeltekkel.

Webarchiválás Csehországban

Törvényi háttér

A kötelezpéldány rendelkezések hazánkhoz hasonlóan Csehországban sem terjednek ki a webarchiválással kapcsolatos tevékenységekre. A webarchiválás feladatát nemzetközi példák alapján a nemzeti könyvtárra bízva modellezik kísérleti projekt keretében, ami 2001-ben indult el (<http://www.webarchiv.cz>). A cseh webarchiválási projekt alakulását az OSZK különféle szintjein már nagyon régóta figyelemmel kísérték korábban is az itthoni webarchiválási tevékenységek megalapozása érdekében.

Szolgáltatási keretek

A cseh webtérre és a cseh vonatkozású tartalmakra irányuló teljes domain aratásokat a .cz domainre koncentrálva a CZ. NIIC céggel való együttműködés keretében végzik. Ennek révén jelenleg mintegy 1 200 000 domain tartalmát aratják le (<http://www.webarchiv.cz/en/comprehensive-harvests> és <https://www.slideshare.net/webarchivCZ/esk-webov-archiv-68200875>), illetve különféle eseményekhez, témákhoz kapcsolódó tematikus, szelektív aratásokat is végeznek (például: <http://www.webarchiv.cz/en/topic-collections>). A webarchiválási projekt még további kiteljesedéséhez szükség lenne a finanszírozási feltételek rendezésére és a kötelezpéldány ren-

delkezések felülvizsgálatára is. Jogi feltételek akadályozzák a kiválasztott tartalmaknak a Cseh Nemzeti Bibliográfiához történő hozzáadását is. Nagy kihívást jelent a számukra, hogy a robots.txt fájlok tartalmának figyelmen kívül hagyásával rengeteg feleslegesnek bizonyuló adatmennyiség is leartásra kerül. A Cseh Nemzeti Könyvtár 5143 különféle tartalomszolgáltató partnerrel írt már alá megállapodást a webarchiválás biztosításáról. A széleskörű együttműködés biztosítása tehát itt is alapfeltétele a webarchiválási tevékenységek kiteljesedésének. A cseh webarchiválási projektről számos Slideshare prezentáció is rendelkezésre áll az alábbi címen: <https://www.slideshare.net/webarchivCZ>.

Hozzáférés

A jogtulajdonosokkal történt megállapodások alapján egy szűk adatkészlet elérhető nyilvánosan online (<http://www.webarchiv.cz/en/browse>), az egyéb szeletek csupán a nemzeti könyvtár épületében, zárt hálózaton. A szerzői jogi törvény az európai szerzői jogi rendelkezésekkel összhangban jelenleg kifejezetten tiltja a szerzői jog által védett leartott anyagok nyilvános szolgáltatását.

Webarchiválás Szlovákiában

Alapvető információk

Szlovákiában a pozsonyi székhelyű nemzeti könyvtári ranggal is bíró Egyetemi Könyvtár koordinálásával zajlik a webarchiválási tevékenység (<http://www.webdepozit.sk>). Az első tesztek még 2006-ra nyúlnak vissza, üzemszerűen pedig 2015-től archiválják a web szlovák szeletét. A webarchiválással három főállású munkatárs foglalkozik, de további három vezető beosztású munkatárs hatáskörét is érinti e terület. A szerver, illetve a technikai infrastruktúra terméktámogatását kiszervezték piaci alapon egy külső partnercég számára. A szolgáltatás megfelelő jogi kereteit is sikerül megteremteni, különös tekintettel a kötelezpéldány szabályozásra. A Szlovák Kulturális Minisztérium betervezte a kormány elé az új kötelezpéldány rendelet tervezetét, benne a webarchiválásra vonatkozó kitételekkel, ami jelenleg társadalmi egyeztetés alatt áll, s még az idén tervezik az elfogadását. Szorosan együttműködnek a cseh partnerekkel, havi szintű személyes szakmai konzultáció zajlik a szakemberek között. Széleskörű együttműködési hálózat kialakítására törekszenek a tudományos és közkönyvtárakkal, valamint a tartalomszolgáltatókkal is.

Gyűjtőkör, szolgáltatási keretek

Tematikus, eseményalapú és általános .sk domainre kiterjedő aratásokat egyaránt végeznek. A gyűjtőköri szabályozást évente felülvizsgálják. A nemzeti domain adiminisztrátor cégtől megkapják évente az aktuálisan bejegyzett .sk domainek listáját. Ezt saját maguk egészítik ki a nem. sk domain alatt lévő webhelyekkel. A 352 ezer bejegyzett szlovák domainből 279 ezret sikerült a 2017 februárjában lezajlott második általános aratás során begyűjteni. A robots.txt előírásait betartják, csupán az adott honlap tulajdonosával történt megállapodást követően hagyják azt figyelmen kívül. A szelektív archiválás keretében 550 intézményt kerestek meg együttműködést kérve, ezek közül idáig 111 intézménnyel sikerült szerződést kötni. A brit példához hasonlóan, a projekt honlapján itt is bárki javasolhat weblapokat archiválásra a megfelelő űrlap segítségével. A webarchívum egyes szeletei a pozsonyi intézményben dedikált munkaállomásokon érhetőek el. A nyilvános hozzáférés a tartalom jogtulajdonosának hozzájárulásával biztosítható. A learatott honlapokról, részben automatizált módon, begyűjtik a metaadatokat s MARC 21 formátumban tárolják, valamint hozzáférhetővé teszik a nyilvános szolgáltatási felületükön. Így, ha a teljes tartalom nem is érhető el, de képet lehet kapni a begyűjtött honlapok jellemzőiről. A metaadatok tárolását, visszakeresést, publikálását saját fejlesztésű keretprogram segíti. Különös jelentőségű a szlovák projekt számunkra azért is, mert a szlovákiai magyar webes tartalmak is a gyűjtőkörbe tartoznak. A HÍD-MOST párt honlapja például elérhető nyilvános szolgáltatás keretei között is.

Összegzés

A fiatal, ám dinamikusan fejlődő szlovák projekt infrastruktúrális hátterét és szervezetségét tekintve is feltétlen figyelemre méltó. Igyekszünk átvenni szakmai tapasztalataikat a magyar szolgáltatás kialakítása kapcsán, illetve megteremteni az együttműködési lehetőségeket is.

Webarchiválás Szlovéniában

Alapvető információk, gyűjtőkör

Az összefoglaló *Alenka Kavčič-Čolič*nak a Szlovén Nemzeti és Egyetemi könyvtár webarchiválásért felelős vezetőjének az IFLA 2017-es konferenciáján Wrocławban elhangzott előadásán alapul.

Szlovéniában a nemzeti és egyetemi könyvtár a fővárosban Ljubljánában végez webarchiválási feladatokat. 2005-ben kezdtek el a témával foglalkozni, tehát már több mint tíz éves tapasztalatokkal rendelkeznek a szelektív webarchiválás terén. A webarchiválásra is kiterjedő új digitális kötelezpéldány törvény 2006-ban született meg. 2007 tavaszától vált az Egyetemi és Nemzeti Könyvtár az IIPC tagjává. A teljes szlovén webtér (.si domain) aratását a megfelelő jogi és műszaki háttér megteremtését követően pedig 2014–2015-ben kezdték el. 2016-ban 1375 webhelyet, illetve 2897 weboldalt arattak le ennek keretében 4.2 TB terjedelemben. Idén 117 ezer URL-t tartalmazó lista aratását tervezik, amely tartalmazza például az összes államigazgatási aldomain nevet is pl. .gov.si. A teljes gyűjtés általános kritériumai a következők: szlovén szerző, szlovén nyelvű honlap, Szlovéniában bejegyzett honlap, illetve Szlovéniában publikált honlap. A szelektív gyűjtések specifikus kritériumainál az adott tartalom lehet önállóan publikált vagy nagyobb egység része, emellett kulturális, tudományos, illetve szellemi értékkel kell rendelkeznie. Ez a specifikus kritériumrendszer az általános aratásnál is részben érvényesül, erre hivatkozva zárják ki pl. az erotikus honlapok archiválását.

Szervezeti, szolgáltatási keretek

Összesen két főfoglalkozású munkatárs foglalkozik a webarchiválási feladatokkal. Ehhez képest különösen szép a tematikus gyűjteményeik magas száma, pl. az államigazgatási, gazdasági, egészségügyi, kulturális honlapok rendszeres aratása, az általános webarchiválás megszervezése. A Web Curator Tool keretrendszer mellé saját kiegészítőt fejlesztettek, amivel a twitteren megjelenő szlovén vonatkozású tartalmak egyes szeleteit is próbálják begyűjteni. A szelektív archiválás keretében 105 967 domaint archiválnak 25,2 TB terjedelemben. Az archívum a <http://arhiv.nuk.uni-lj.si/> címen érhető el.

Webarchiválás Észtországban

Alapvető információk, gyűjtőkör

Az észt Nemzeti Könyvtár webarchiválási projektjének (<http://veebiarhiiv.digar.ee/>) alapvető célkitűzése az észt kulturális örökség szempontjából fontos webes tartalmak archiválása (<http://www.nlib.ee/index.php?id=21581>). A nemzeti könyvtárról szóló törvény nevesíti kötelezettségként a nemzeti könyvtár számára a webarchiválási feladatokat (gyűjtés, feldolgozás,

szolgáltatás). A gyűjtőkör alapelemei a következők: észti nyelven megjelenő weblapok, Észtsorszáiban publikált weblapok, Észtsországra vonatkozó információkat tartalmazó weblapok. A webarchiválás szabályozását a köteleespéldány törvény rendelkezései közé is beillesztették. A weben megjelenő publikációk is köteleespéldánynak minősülnek, melyeket le kell aratni és nyilvánosan közzé kell tenni. A tartalom tulajdonosának azonban lehetősége van intézkedni arról, hogy a tartalom csak zárt hálózatban a Nemzeti Könyvtár épületében legyen elérhető tekintettel a szerzői jogokra. 2017. január 1-jétől ez már alapértelmezetté is van, az archívum a kijelölt intézményekből érhető csupán el. A teljes törvény elérhető angolul is: <https://www.riigiteataja.ee/en/eli/514092016001/consolide>.

Szolgáltatási keretek

A pdf alapú kiadványok begyűjtése 2006-ban, a webaratás 2008-ban kezdődött el. Az archívum 2013 óta érhető el a felhasználók számára. 2010-től 2015-ig 56 millió URL-ről arattak le adatot mintegy 4,2 TB mennyiségben. A webarchiválást szelektív aratásokkal kezdték el, ám mivel a releváns tartalomnak csak kis mennyiségét tudták így begyűjteni, 2015–16-ban lebonyolították az észti.ee domain első webes aratását. Egy adott webhelyről maximum 300 MB adatot arattak le, hogy kezelni tudják az aratási folyamatot. Összesen 4 TB tömörítetlen adatot arattak le ily módon. Az új köteleespéldány törvény szerint amennyiben a könyvtár nem tudja learatni a gyűjtőkörbe eső tartalmat, akkor a tartalom tulajdonosa kötelezett arra, hogy eljuttassa azt a Nemzeti Könyvtár számára. A beszállítás költségeit a tartalom tulajdonosának kell állnia! Őt könyvtárat jelölték ki országosan a köteleespéldányok elérésére az adott intézményeken belül (beleértve a webarchívumot is, zárt dedikált hálózatban).

Szerkezeti keretek

Az Észti Nemzeti Könyvtárban három főállású munkatárs foglalkozik webarchiválási feladatokkal, ketten az archiválásért felelős szakemberek, illetve az alkalmazások kezeléséért felelős adminisztrátor. Az ő munkájukat egy olyan munkacsoport segíti, mely tíz kutatási és kulturális örökség kezelésével foglalkozó

intézmény 24 munkatársát tömöríti. A munkacsoport tanácsokat ad a learatandó anyag kiválasztására, hozzáférhetővé tételére az aratás számára, emellett képviseli a kutatói szféra webarchiváláshoz kötődő érdekeit is. 2011-ben stratégiai dokumentumban fogalmazta meg a munkacsoport a webarchiválásra vonatkozó gyűjtőkörü alapelveket. 2012-től az Észti Nemzeti Könyvtár tagjává vált az IIPC-nek is.

Összegzés

Az észti köteleespéldány szabályozás megítélésünk szerint európai szinten is példaértékűen sikerült. Az Észti Nemzeti Könyvtár megfelelő szervezeti kereteket és erőforrásokat is képes garantálni a webarchiválási tevékenység biztosításához, miközben kialakítottak olyan közgyűjteményeket is magukban foglaló szélesebb körű együttműködési hálózatot is, ami a tartalmak begyűjtésének hatékonyságát növeli meg.

Webarchiválás Hollandiában

A holland webarchiválási modell számos különleges tulajdonsággal bír. Ezekről első kézből értesülhettünk az OSZK-ba látogató magyar származású *Kees Teszelszkytől**, aki a Holland Nemzeti Könyvtár webarchiválással foglalkozó csapatának tagja. Kifejtette nekünk, hogy Hollandiában történeti okokból nem alakult ki a hagyományos papíralapú dokumentumokra sem a köteleespéldány rendszer, a kiadók önként juttatják el kiadványaikat a Holland Nemzeti Könyvtár részére. Ebből kifolyólag nincs webes köteleespéldány szabályozás sem. Ennek hiányában viszont kifejezetten tiltott törvényileg az általános célú aratás a holland domainről. A Holland Nemzeti Könyvtár csak egy a sok közgyűjtemény közül, amelyek szelektív webaratást végeznek, bár egyfajta koordinációs funkciót tölt be a gyűjtőkörök egyeztetésénél, illetve szoftverfejlesztési együttműködésekben is részt vesz nemzetközi partnerekkel. A holland domain volt az egyik első, amely kialakult Európában, s ma is nagyon jelentős tartalmi gazdagsággal bír. Igyekeznek különféle eseményekhez, témakörökhöz kötődő aratással minél több szöveget begyűjteni ennek a webes térnek. Sok év helyben járás után e terület most kiemelt figyelmet élvez a

* Az OSZK webarchiválással foglalkozó workshopján az előadó részletesen ismertette a belga és a holland szakmai tevékenységek kereteit. A prezentáció és a videofelvétel linkje is elérhető az alábbi címen: http://mekosztaly.oszk.hu/mia/404_workshop.html

digitális könyvtári fejlesztések között, ami pozitív fejleményeket sejtet a jövőre nézve.

Webarchiválás Ausztriában

Ausztriában a webarchiválási tevékenységet az Osztrák Nemzeti Könyvtár Digitális Könyvtári Osztálya szervezi. A tevékenység kereteit a médiatörvény szabályozza (https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Erw&Dokumentnummer=ERV_1981_314). A jogszabály évente négy általános aratást engedélyez a szelektív, illetve eseményalapú aratások mellett. A gyűjtőkör az osztrák webteret, illetve az ausztriai vonatkozású tartalmakat foglalja magában. Az osztrák webtér kiterjed a .at fő domain mellett az újonnan bejegyzett .wien és .tyrol új domain névtartományokra is. Mintegy 1,3 millió webhelyre terjed ki e webtér jelenleg, az osztrák webarchívum pedig mintegy 1,7 millió webhelyről őriz adatokat, a tárolt tartalom mennyisége mintegy 100 TB terjedelmű. A begyűjtött tartalom az Osztrák Nemzeti Könyvtárban két dedikált terminálon tekinthető meg, illetve négy másik intézményben is lehetőség van a tartalmak elérésére zárt dedikált hálózatban. Nyilvánosan kereshető a webarchívumban lévő webhelyek listája saját fejlesztésű keresőfelület révén (<https://webarchiv.onb.ac.at>). Ezen a felületen új tartalmat is lehet ajánlani archiválás céljából. A találati halmazt el lehet menteni, majd a könyvtárba betérve megtekinteni az adott webhelyeket. Sajnos a törvényi kereteket a technikai feltételek, illetve a munkaerő hiánya miatt nem tudják teljeskörűen kihasználni. A Nemzeti Könyvtárban két főállású munkatárs foglalkozik webarchiválási feladatokkal, ketten félállásban segítik a munkájukat. A tárhelynek meglehetősen szűkében vannak. Az általános aratás kapcsán szigorú méretkorlátokat kell alkalmazniuk az egyes webhelyekre vonatkozóan. Egy általános aratás során nem gyűjthetnek be 6TB-nál nagyobb adatmennyiséget, két évente kerül sor erre 2009 óta,

melynek időtartama mintegy fél évet vesz igénybe. A szelektív webarattással 2008 óta évente mintegy 2 TB adat begyűjtésére van mód. Széleskörűen, számos kategóriában gyűjtenek webhelyeket, emellett pedig például a legutóbbi választások kapcsán is eseményalapú gyűjtést is folytattak. A Dán Nemzeti Könyvtár által fejlesztett Netarchive Suite programcsomagot használják, melyet könnyít az a tény, hogy a webarattáshoz nem kell engedélyt kérniük, s a robots.txt szabályait sem kell figyelembe venniük. Így Dániához hasonlóan itt sincs szükség az engedélykézés adminisztrálására. Az osztrák webarchívum is tagja az IIPC-nek, a nemzetközi együttműködésben való aktív részvétel lényegesen megkönnyítette a webarchiválási tevékenység technikai, illetve szakmai megalapozását is.

Konklúzió

A nemzetközi körképet áttekintve megállapíthatjuk, hogy a jogi, szabályozási kereteket, illetve a webarchiválási tevékenységek szervezeti kereteit tekintve rendkívül vegyes az összkép. Az adott ország állami berendezkedése, a szakmai attitűdök, a pénzügyi háttér, az együttműködés kultúrájának erőssége erősen befolyásolja a webarchiválási feladatok szervezését. Kritikus elemet jelent a megfelelő jogi szabályozási keretek megléte. A kötelezpéldány szabályozásba itthon is bele kellene foglalni a webarchiválási tevékenységet. Az ésszt mintából sokat lehetne meríteni e téren. Pozitívumot jelent a jövőre nézve, hogy a magyar kísérleti projekt szakmai kereteinek kialakításakor számos jó példát lehet találni a szomszédságból is. A 2017. évi IFLA konferencián külön szekcióülés foglalkozott a webarchiválás aktuális kihívásaival. Reményeink szerint ez a rövid összefoglaló ráirányítja a figyelmet a webarchiválás nemzetközi dimenziójára, s elősegíti, hogy a külföldi tapasztalatok jelentős szeletét az itthoni körülmények között is sikerrel hasznosíthassuk.

