

Webarchiválási politikák

Immáron öt esztendeje annak, hogy a Magyar Elektronikus Könyvtár egyik atyjaként számon tartott Drótos László a miskolci Networkshopon előrukkolt egy Magyar Internet Archívum létrehozásának gondolatával.¹ A kérdés már akkor több volt, mint időszerű. A magyarországi szaklapokban azóta nem egy, a külföldi fejleményekről tudósító cikk, referátum látott napvilágot, közöttük olyanok is, amelyek a „nagy nemzetek” helyett a „kis európai nemzetek”, így például a csehek vagy litvánok webarchiválási gyakorlatát ismertetik.² A hazai szakemberek is igyekeztek a problémát napirenden tartani.³ A kérdés jogi aspektusaival is foglalkozott a Tudományos és Műszaki Tájékoztatásban megjelent cikkében⁴ Rónai Iván, aki konkrét javaslattal is élt a hatályos köteleispéldány-rendelet (60/1998) megváltoztatásának tárgyában. A szerzői jog kérdésköre felől közlített a témához Amberg Eszter, aki – ugyancsak a TMT-ben megjelent – tanulmányában⁵ arra is felhívta a figyelmet, hogy a jelenleg érvényben lévő köteleispéldány-rendelet elviekben rendelkezik a távoli hozzáférésű dokumentumok beszolgáltatásának kötelezettségéről, más kérdés, hogy a szolgáltatásra kötelezettek többségének értelmezésében az elektronikus dokumentum rendeletbe foglalt definíciója nem vonatkozik az online kiadványokra.

A témát taglaló írások rendre kitérnek azokra az alapelvekre, amelyek mentén a különböző országok jogalkotói, illetve érintett intézményei (legtöbbször a nemzeti könyvtár vagy annak társintézményei) a webarchiválást megoldani igyekeznek, illetve amelyeket koncepcióként a vonatkozó jogszabályokba ágyaznak. Az elképzelések a főbb pillérek tekintetében gyakorta egyeznek. Ilyenek például a nemzeti webterület archiválása (esetünkben a *.hu* végű doménnevek alá tartozó online tartalmaké) általában a webaratás módszerével; a távoli hozzáférésű kiadványokra, illetve a born-digital dokumentumokra (e-folyóiratokra, e-könyvekre) vonatkozó köteleispéldány-szolgáltatás jogi szabályozása; valamilyen szelekciós elv alkalmazása a gyűjtés során (pl. bizonyos online kiadványtípusok kizárása az archiválандók köréből). Jelen írásomban a kurrens webarchiválási politikákat, valamint a mögöttük húzódó megfontolásokat szándékozom sorra venni, mindenekeelőtt *Thomas Chaimbault* „L'archive du Web”⁶ című tanulmányára alapozva.

* * *

Alapelvek

Teljes körű gyűjtés és archiválás

E megközelítés a teljes webterület learatását célozza meg. E koncepcióra példa az 1996-ban indult Internet Archive (IA) kezdeményezés, amely mindenféle válogatás nélkül végzi a weben megjelenő tartalmak archiválását, méghozzá úgy, hogy tevékenységéhez nem kéri a tartalmakat előállítók hozzájárulását, okot szolgáltatva az eljárással kapcsolatban felmerülő szerzői jogi aggályokra. Az IA, amelynek felhasználói sajátos Robin Hoodokként szabadítják fel a mások által digitalizált, szerzői jogi hatály alá eső műveket⁷, ugyanakkor egy nemzetközi konzorcium (International Internet Preservation Consortium, <http://netpreserve.org/about/index.php>) tagjaként, nagy nemzeti könyvtárakkal, archívumokkal karöltve igyekszik választ találni az interneten elérhető tartalmak hosszú távú megőrzésével és szolgáltatásával kapcsolatban felmerülő kérdésekre.

Célzott (exhausztív) gyűjtés és archiválás

Az ezt a politikát követő webarchiválásban érdekelt intézmények a webterület egy kijelölt szegmensének, jellemzően az egy nemzeti doménnév alá tartozó webtartalomnak a gyűjtését és megőrzését végzik, azt azonban teljes körűen, „kimerítő” jelleggel. Viszonylag széles körben alkalmazott stratégiáról van szó. Két példa: a 2010-ben lezárult svéd Kulturarw3 projekt, illetve a finnországi EVA projekt, de ugyancsak domén alapon gyűjt és archivál Franciaországban a nemzeti könyvtár és a Nemzeti Audiovizuális Intézet, valamint az osztrák jogi szabályozás⁸ is lehetővé teszi az .at doménnév alatt működő kiadványok learatását. Ez a megoldás kínálkozik az e-kötelezpéldány-szolgáltatás kötelezettsége alá nem eső, dinamikus tartalmú online elektronikus kiadványok (blogok, levelezőlisták stb.) archiválása tekintetében. Mindenképp előnyös tehát átültetnünk e politikát a magyar gyakorlatba, mindenekelőtt megteremtve a szükséges jogszabályi hátteret, a kötelezpéldány-rendelet szintjén rögzítve a nemzeti könyvtár jogosultságát a teljes magyar webterület „learatására”.

Válogató gyűjtés és archiválás

E megközelítés a webterület egy vagy több meghatározott szelekciós elv alapján történő archiválását foglalja magában. A válogatás alapja lehet minőségi kritérium, valamilyen tartalmi szempont, esetleg célzottan bizonyos típusú weboldalak összegyűjtése. A kiválasztott weboldalakról az archiváló intézmény meghatározott időközönként pillanatfelvételt készít. A válogató archiválásra jó példa az Ausztrál Nemzeti Könyvtár PANDORA elnevezésű projektje⁹. (A kapcsolódó útmutató részletes tájékoztatást ad a válogatás szempontjairól: <http://pandora.nla.gov.au/selectionguidelines.html>.)

Tematikus gyűjtés és archiválás

Az előző egy speciális esete általában bizonyos eseményekhez kötődik. A Francia Nemzeti Könyvtár (BnF) a 2002-es elnöki és törvényhozási, valamint a 2004-es regionális és uniós választások alkalmából készített tematikus gyűjteményt, míg a washingtoni Kongresszusi Könyvtár például az egyes választásokról és a 2003-as iraki háborúról készített hasonló archívumot¹⁰. Tematikus gyűjtemények természete-

tesen másféle alapokon is létrejöhetnek. Ilyen például a Heidelbergi Egyetem DACHS (Digital Archives for Chinese Study) projektje¹¹, vagy az ARCHIPOL projekt¹², amely a holland politikai pártok weboldalait gyűjti össze.

Kombinált gyűjtés és archiválás

Miután önmagában egyik megközelítés sem kínál tökéletes megoldást az interneten található kulturális örökség megőrzése tekintetében, több országban is a koncepciók kombinálása mellett döntöttek. Franciaországban és Dániában például a nemzeti doménnevek alá tartozó weboldalak teljes körű, célzott gyűjtése kiegészül a meghatározott kritériumok alapján történő, válogató jellegű, valamint a bizonyos eseményekhez kapcsolódó, tematikus gyűjtéssel.

A webarchiválási politikák jogszabályi háttérét jelentő kötelespéldány-törvények legtöbbször a korábbi, elsősorban a papíralapú kiadványokra vonatkozó jogszabályok kiegészített változatai, és országonként eltérő módon határozzák meg az általuk érintett távoli hozzáférésű kiadványok értelmezési körét. A különböző (nyelvi, földrajzi, tematikus) meghatározások azonban nem szükségszerűen jelentenek szelektív jellegű gyűjtést, a köteles példányok gyűjtését végző intézmények ugyanis a törvénybe foglalt definíciók lehető legtágabb értelmezésében érdekeltek.

Kivitelezés

A webarchiválási stratégiákat a kivitelezés szempontjából ugyancsak kategorizálhatjuk.

Automatizált gyűjtés

Az automatikus gyűjtés során egy keresőrobot végzi el a webaratást előre definiált kritériumok alapján. Általában a célzott gyűjtésben alkalmazzák, például az egy nemzeti doménnev alá tartozó weboldalak „learatására”. Ilyen módon gyűjtötték egybe a svéd Kulturarw3 projekt során a .se végű doménnevek alá tartozó tartalmakat.

Részben automatizált gyűjtés

Az előzőtől annyiban különbözik, hogy szigorúbbak a szelekció kritériumai, ennek megfelelően a – minőségi követelményeket is könnyebben érvényesítő – válogató gyűjtések során alkalmazzák. Példa rá a fentebb már megemlített ausztrál PANDORA projekt.

Manuális gyűjtés

Nem jellemző, hogy az archiválandó webtartalmak begyűjtését kizárólag az emberi tényezőre bízzák, de erre is van példa: a Québeci Nemzeti Könyvtár azokat a monográfiákat és sorozati kiadványokat gyűjti és indexeli ily módon, amelyeket a québeci kormányzat tett elérhetővé az interneten. Az erősen válogató gyűjtés során minden egyes publikációra vonatkozólag külön licencmegállapodást kötnek az érintett kormányzati szervvel vagy miniszterrel, amelynek értelmében utóbbi a vonatkozó kiadvány szerzői jogát átruházza a könyvtárra. A humán faktor megkerülhetetlenségét egyébiránt a francia nemzeti könyvtárban is felismerték, ahol az automati-

zált gyűjtés (webarátás) során felmerülő hibák korrekcióján túl a mély weben közzétett tartalmak azonosítása, válogatása és gyűjtése is manuális úton valósul meg.

Végül jegyezzük meg, hogy az ésszerűség keretein belül történő, lehetőség szerinti teljességre való törekvés az internetarchiválás terén azt igényli, hogy a nemzeti gyűjtemények összehangolják ez irányú erőfeszítéseiket. Ennek eszköze a már korábban említett, immáron 35 nemzeti könyvtárat tömörítő *International Internet Preservation Consortium* nevű szervezet, avagy az Európai Bizottság által kezdeményezett *Network European Deposit Libraries* (Letéti Könyvtárak Európai Hálózata) projekt.

Egy példa – Új-Zéland

Az Új-Zélandi Nemzeti Könyvtár (NLNZ) jól kidolgozott, komplex technológiai arzenálja kiváló eszközként szolgál a kötelepéldány-rendelet¹³ hatálya alá eső elektronikus dokumentumtípusok begyűjtésére és archiválására.

Az e-köteleként szolgáltatandó „born-digital” kiadványokat létrehozóik a Web Deposit Tool¹⁴ alkalmazás segítségével juttathatják el a Nemzeti Digitális Örökség Archívumba (National Digital Heritage Archive)¹⁵, amely utóbbi technológiai hátterét az Ex Libris csoporttal közösen fejlesztett Rosetta digitális megőrző rendszer¹⁶ jelenti. A weboldalak szelektív jellegű gyűjtésére a British Libraryvel együttműködésben kifejlesztett, nyílt forráskódú Web Curator Tool szoftvert¹⁷ kezdtek el használni. E gyűjtések eredményeiből nemzeti webarchívum (New Zealand Web Archive) épül. Az Új-Zélandi Nemzeti Könyvtár először 2008 októberében végezte el a teljes nemzeti webterület aratását, másodszor az Internet Archive-ot bízta meg a feladattal. Utóbbi 2010. május 12-e és június 5-e között, mintegy 24 nap alatt 170 millió URL-ről gyűjtött össze 6,1 terabájtnyi adatot, utóbbiakba beleértve a naplófájlokat is.

A nemzeti webarchívumban való keresgéléshez a könyvtár online katalógusa felől kell elindulni. A találati listában a címek mellett látható [electronic resource] bejegyzés tanúskodik arról, hogy archivált weboldalról vagy egyéb online publikált kiadványról van szó. A leírás (1. ábra) alján az „Archived copy” kifejezésre kattintva jutunk el a könyvtár integrált rendszerében (Voyager-ben) őrzött intellektuális egységekhez, voltaképpen a példányokhoz (2. ábra).

The screenshot shows the National Library of New Zealand Catalogue interface. At the top, it says "National Library of New Zealand Catalogue" and "Find items within the National Library". Below this is a search bar with the text "New Zealand" and "Auckland Cricket Association". The search results show "You are connected to the National Library of New Zealand Catalogue" and "Search Request: Keyword Relevance = auckland cricket". The search results display "1 of 11 results". The first result is "Auckland Cricket Association" with a "Link to: Archived copy" button. The result details include: "Reference: ●●●●●", "Title: Welcome to the official web site of Auckland Cricket [electronic resource].", "Vaireur Title: Title in HTML header: Auckland Cricket Association", "Publisher: (Auckland, N.Z.): Auckland Cricket", "Summary: Profile of the Auckland Cricket Association and their responsibilities. Includes news about Auckland cricket, lists Auckland top players and their profiles, player and coach development programmes and their marketing vision.", "Description: Updated irregularly", "Subject: Auckland Cricket Association; Cricket -- New Zealand -- Auckland.", "Link to: Auckland Cricket Association". A red oval highlights the "Link to: Archived copy" button.

1. ábra

Results 1 - 4 of 4



<http://www.aucklandcricket.co.nz/>
[View digital content](#)

Date Harvested: 20-09-2009



<http://club.aucklandcricket.co.nz/>
[View digital content](#)

Date Harvested: 04-10-2008



<http://www.aucklandcricket.co.nz/>
[View digital content](#)

Date Harvested: 20-09-2007



<http://club.aucklandcricket.co.nz/>
[View digital content](#)

Date Harvested: 19-09-2007

2. ábra

Ezek valamelyikére kattintva, mindenekelőtt tájékoztatást (3. ábra) kapunk az aratás időpontjáról és a lehetséges technikai korlátokról (pl. esetenként nem működő linkek).

This archived website contains the public web pages produced by the author.
Normal copyright restrictions apply.

Navigate the website by using the hyperlinks, tables or the site map, as the search boxes may not function. Links to content located on another author's website will not be accessible from this archived website, but may have been captured separately. While we endeavour to make a complete copy of the website, there are times when content may not have been captured due to technical limitations.

Date of Harvest: 20 September 2009 at 08:00 PM

Click the link below to view the archived website:

<http://www.aucklandcricket.co.nz/>

3. ábra

Innen már csak egy kattintás a Rosetta rendszerben hosszú távú megőzésre elraktározott, az NLNZ saját fejlesztésű alkalmazásával emulált módon megjelenített, learatott webtartalom megtekintése (4. ábra).

Zárszó

Ma a nemzeti kulturális örökségünk jelentős része eleve digitális formában (ún. born-digital kiadványként) jön létre, és publikálásának egyedüli módja a világhálón való közzététel. E kulturális javak hosszú távú megőzésének sürgető fontosságát a fejlett országok felelős intézményei – élükön legtöbbször a nemzeti könyvtárakkal – és döntéshozói viszonylag korán felismerték, és megtették a lépéseket a probléma orvoslására. E lépések közül az egyik legfontosabb a jogszabályi háttér megteremtése, amely általában a hatályos kötelezpéldány-rendelet célnak megfelelő módosítását, esetleges újraalkotását foglalja magában. Jóllehet, hazánkban már évekkel ez-



4. ábra

előtt született olyan kezdeményezés, amely egy nemzeti internet archívum létrehozását szorgalmazta, Magyarország továbbra is lépéshátrányban van még a régió né- mely államaival szemben is. (Ausztriában, Csehországban, Horvátországban, Lengyelországban, Szlovéniában már most működik webarchívum, és a legtöbb helyen a jogi szabályozás is megoldott. Szlovákiában többek között a jogi háttér megterem- tése késlelteti a probléma megoldását.) Mindent el kell követnünk, hogy Magyaror- szág ezen a területen is felzárkózzon a többi fejlett országhoz, és hogy kulturális örökségének és a jövő generációk információigényét kielégítő forrásoknak nagy je- lentőséggel bíró hányada ne menjen jóvátehetetlenül veszendőbe.

JEGYZETEK

- 1 Drótos László: Mi a MIA? – Javaslat egy Magyar Internet Archívum létrehozására. = Tudomá- nyos és Műszaki Tájékoztató, 53. (2006) 6. sz. 267–274. p.
Elektronikus változat: http://tmt.omikk.bme.hu/show_news.html?id=4431&issue_id=473
- 2 A litván példáról lásd: Jodelis, Remigijus: Harvesting and archiving of electronic resources in Lithuania: towards virtual library. Ref.: Elektronikus források begyűjtése és archiválása Litvániaiban. Úton egy virtuális könyvtár felé / Birkás Bence. = Tudományos és Műszaki Tá- jékoztató, 51. (2004) 6. sz. 257–260. p.
Elektronikus változat: http://tmt.omikk.bme.hu/show_news.html?id=3640&issue_id=451;
a cseh példáról lásd: Gruber, Lukáš–Sibek, Tomáš–Coufal, Libor: Archivace webových strán-

- nek v českém prostředí aneb Jak funguje WebArchiv. = Čtenář, 61. (2009) 5. sz. Ref.: Cseh weboldalak archiválása avagy Hogyan működik a WebArchiv? / Prókay Margit. – <http://www.kithirlevel.hu/k.php?k=11365&s=1&h=w>
- 3 Vö.: [http://www.vilagtudomany.hu/index.php?data\[mid\]=7&data\[id\]=260&internetarchivum-abritishlibraryiselindította](http://www.vilagtudomany.hu/index.php?data[mid]=7&data[id]=260&internetarchivum-abritishlibraryiselindította)
- 4 Rónai Iván: Kulturális digitalizálás – félüton az Europeana felé. = Tudományos és Műszaki Tájékoztatás, 56. (2009) 8. sz. 371–377. p.
Elektronikus változat: http://tmt.omikk.bme.hu/show_news.html?id=4431&issue_id=473
- 5 Amberg Eszter: Szerzői jog a digitális könyvtárban. = Tudományos és Műszaki Tájékoztatás, 57. (2010) 6. sz. 250–258. p.
Elektronikus változat: http://tmt.omikk.bme.hu/show_news.html?id=5330&issue_id=515
- 6 <http://www.enssib.fr/bibliotheque-numerique/document-1730>
- 7 Lásd erről: <http://dancsz.fxnet.hu/konyvtar/a-google-books-esete-robin-hooddal>
- 8 Lásd: www.onb.ac.at/files/Mediengesetznovelle_BGBI_8_2009.pdf
- 9 A projektről bővebben lásd: Dippold Péter: A nemzeti bibliográfiák gyűjtőköre, avagy elérhető-e a teljesség? = Könyvtári Figyelő, 52. (2006) 2. sz. 261–268. p.
Elektronikus változat: <http://ki.oszk.hu/kf/kfarchiv/2006/2/dippold.html>
- 10 <http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>
- 11 <http://www.sino.uni-heidelberg.de/dachs/>
- 12 <http://www.archipol.nl/english/index.html>
- 13 Rövid ismertetés a kötelezpéldány-szolgáltatásról a könyvtár honlapján: <http://www.natlib.govt.nz/services/legal-deposit-donations/legal-deposit-intro>
- 14 <http://ndha-wiki.natlib.govt.nz/ndha/pages/NDHAWebDepositToolIntro>
- 15 <http://www.natlib.govt.nz/services/about-us/current-initiatives/ndha>
- 16 <http://www.exlibrisgroup.com/category/RosettaOverview>
- 17 <http://www.natlib.govt.nz/services/get-advice/digital-libraries/web-curator-tool>