



Az Országos Széchényi Könyvtár Webarchívumának 2020-as újdonságai

DRÓTOS László

Ez az írás a „404 Not Found – Ki őrzi meg az internetet?” online workshopon 2020. december 2-án elhangzott előadás szerkesztett változata és a Könyvtári Figyelő 2020/1. számában megjelent Az OSZK webarchívumának újdonságai című cikk folytatása az azóta eltelt egy év eredményeiről.

Bevezetés

Ha a webarchívum egy termelő egység lenne, akkor jelenlegi fejlettségi szintjét, működési struktúráját egy olyan gyárhoz hasonlíthatnánk, ahol már zajlik a tömegtermelés, de az egyik sarokban még nagy erővel folyik a kutatás-fejlesztés is, a különböző gépek és technológiák tesztelése. Ebben a tanulmányban egy mezőgazdasági üzem metaforájaként jelenítem meg, amely most már (viszonylag) rendszeresen „aratja” a magyar webtér mind nagyobb részét. A projekt az elmúlt évben sokat haladt előre a járványhelyzet okozta nehézségek ellenére is, hiszen a munkafolyamatok többsége távmunkában is megfelelő színvonalon elvégezhető. Elsőként lássuk a környezeti változásokat!

„Vidékfejlesztés”

Az egyik fontos fejlemény, hogy az Országos Széchényi Könyvtár (OSZK) szervezeti megújulásának részeként az Információ- és Tartalomszolgáltatási Főosztályon belül létrejött egy önálló Webarchiválási Osztály, egyelőre ugyanazzal a három főállású munkatárssal, akik korábban az E-könyvtári Szolgáltatások Osztály keretében végezték ezt a munkát. A szoros kapcsolat természetesen továbbra is megmaradt az elektronikus könyvtári gyűjteményekkel, hiszen a webarchívum azokat az online forrásokat is menti, amelyekből részben a Magyar Elektronikus Könyvtár (MEK), az Elektronikus Periodika Adatbázis (EPA) és a Digitális Képarchívum (DKA) is gyarapodik, csak mi egyedül dokumentumok helyett webhelyszinten archiválunk.

A másik nagy előrelépés, amely sokat jelent kultúránk digitálisan keletkező és terjedő szegmensének hosszú távú megőrzése szempontjából, hogy az elmúlt évben – az Emberi Erőforrások Minisztériumával (EMMI) történt számos egyeztetés után – megszülettek azok a jogszabályok, illetve törvénymódosítások, amelyek a nemzeti könyvtár feladataként határozzák meg a webarchiválást, és ehhez forrást is biztosítanak. Elsőként a kulturális törvény 2020. május 19-én elfogadott módosításai között jelent meg ez a téma, majd ezt követően egy júliusi kormányhatározat költségvetést is rendelt hozzá. Végül a Kormány december 22-i dátummal kiadott rendeletében részletesebben is szabályozta a webarchiválási tevékenységet, az Országos Széchényi Könyvtár, illetve a tartalomszolgáltatók jogait és kötelezettségeit (<https://webarchivum.oszk.hu/a-webarchivalasra-vonatkozozo-jogszabalyok/>). Ezzel kialakult nálunk az a jogi keretrendszer, amelyre egyes országokban még mindig csak várnak az ottani kollégák, holott esetleg már évtizedek óta archiválnak webtartalmakat – például a kötelezpéldányokra vonatkozó szabályozás hatókörét kiterjesztve –, de egyértelmű jogszabály híján mind a begyűjtés, mind pedig a szolgáltatás terén sokkal nehezebb a dolguk. Ami az informatikai infrastruktúrát illeti, az év folyamán kétszer is költöztettük a gyűjteményt: tavasszal a Kormányzati Informatikai Fejlesztési Ügynökségtől (KIFÜ) az OSZK-ba, majd ősszel vissza az OSZK-ból a KIFÜ-be. Ennek az volt az oka, hogy a 2017

elején a KIFÜ-től az első tesztidőszakra kölcsönkaptott szervert – az időközben történt tárhelybővítés ellenére is – 2020 elejére nagyrészt betelt, ezért az addig összegyűjtött állományt áttelepítettük az OSZK egy saját, nagyobb tárhellyel rendelkező szerverére. Majd, miután beüzemelésre került az Országos Könyvtári Rendszer keretében beszerzett nagy kapacitású infrastruktúra, melyet a KIFÜ működtet, ebbe a „felhőbe” költöztünk be. Jelenleg három szervert használunk: a *webadmin* nevűn található a nyilvános gyűjtemény; a 150 terabájtos háttértárral ellátott *webharvest* gépen futnak a nagy méretű aratások és azon tároljuk a zárt állományt; a *webarchivum* szerver pedig a projekt honlapját szolgáltatja.

A *webarchivum.oszk.hu* címen 2020 februárjában indult új honlapon keresni és böngészni lehet a nyilvános archívum anyagában, metaadatszinten kereshetők a zárt gyűjteményben levő webhelyek, és itt érhetőek el a projekt hírei és dokumentumai is. Mivel az internetes tartalmak tömeges megőrzése hazánkban még újdonságnak számít, ezért az oldalt egyben tudásbázisnak is szánjuk, melyhez egy wiki is kapcsolódik (<https://webarchivum.oszk.hu/mia-wiki/>). A honlapon külön menüpontokat hoztunk létre az archívum felhasználói, a tartalomgazdák, a téma iránt érdeklődő közgyűjteményi és egyéb szakemberek, valamint a sajtó munkatársai számára. (1. ábra). A fontosabb információk angol nyelven is olvashatók.



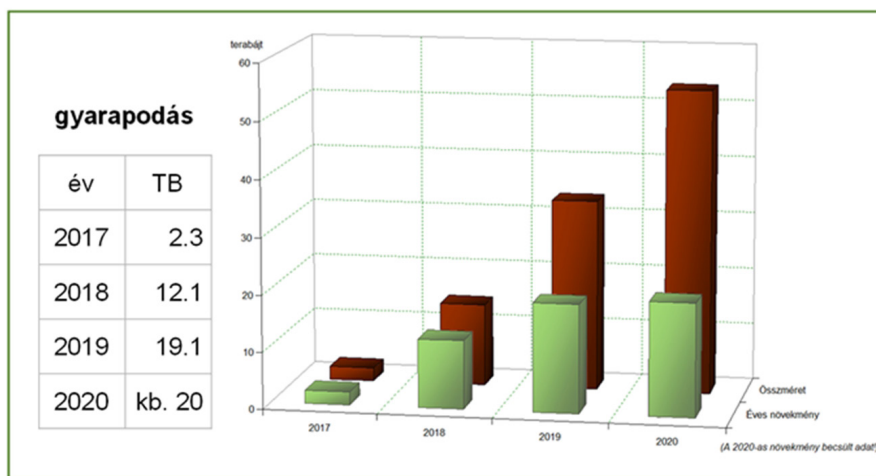
1. ábra
Az OSZK Webarchivumának
2020-ban indult honlapja

Ugyancsak ismeretterjesztési célokat szolgál az a mintaalkalmazás, amely szintén 2020. eleje óta érhető el a rakoczi2019.webarchivum.oszk.hu oldalon. A Közgyűjteményi Digitalizálási Stratégia (KDS) pályázat keretében megvalósított archívum apropóját a 2019-ben kezdődött *II. Rákóczi Ferenc Emlékév* adta. Amellett, hogy lementettünk az emlékévvvel kapcsolatos híreket, valamint a fejedelemmel, a Rákóczi családdal és a szabadságharcral foglalkozó egyéb weboldalakat, honlapokat és blogokat, a MEK, az EPA és a DKA állományából is gyűjtöttük az ezekről a

témákról szóló, digitalizált vagy digitális formában született dokumentumokat: könyveket és könyvrészeket, cikkeket és tanulmányokat, fotókat és egyéb képeket. Ezek a digitális tartalmak együtt kereshetők és a böngészhetők a hatnyelvű honlapon.

„Terméseredmények”

Az aratógépeink által „betakarított” webtartalom növekedése a következő grafikonon látható (2. ábra).



2. ábra

A lementett és eltárolt tartalom éves növekedése, valamint az archívum összmérete

A táblázatban levő számok azt mutatják, hogy az első két év kísérletezése után, a már-már üzemszerűnek tekinthető működés során 20 terabájt körüli az éves gyarapodás. Ennek értelmezéséhez egyrészt tudni kell, hogy általában úgy futtatjuk a robotokat, hogy az előző mentéshez képest csak az új vagy megváltozott fájlokat tárolják el, vagyis 2020-ban is sok, korábban még nem aratott webhely került a gyűjtőkörünkbe, így sikerült közel ugyanannyival növelni az archivált anyag mennyiségét, mint az előző évben. Másrészt persze ezt a 20 terabájtot könnyen meg is lehetne többszörözni, ha másként paramétereznénk az aratószoftvert, vagy ha gyakrabban és hosszabb ideig futtatnánk, de mivel a teljességre törekvés reménytelen egy olyan folyamatosan változó és exponenciálisan növekvő médium esetében, mint a világháló, és a rendelkezésre álló erőforrások is korlátozottak, ezért igyekszünk optimalizálni az aratások mélységét és gyakoriságát.

A témák szerint válogatott webcímekeket tartalmazó

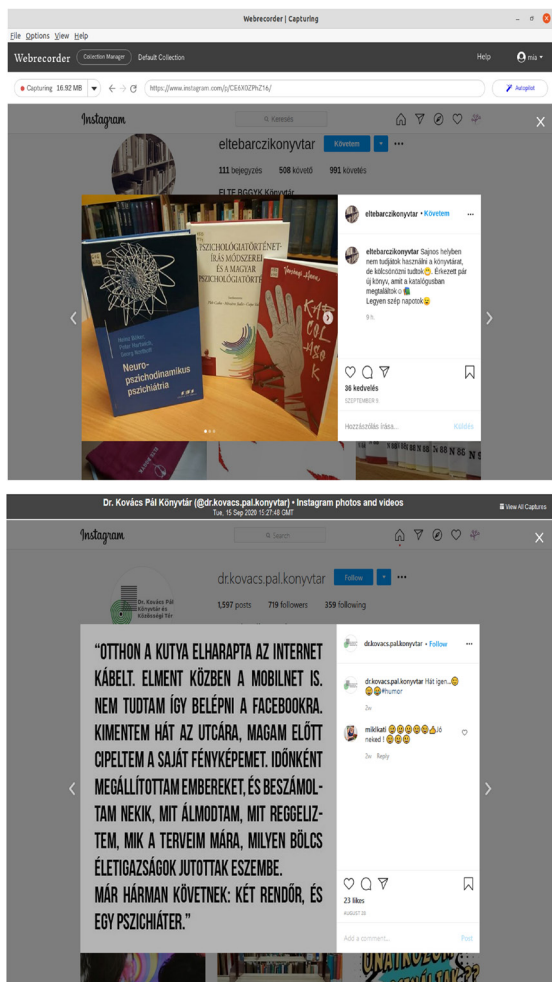
részgyűjtemények köre hárommal bővült a 2020-as évben: *média*, *oktatás* és *sport*, három korábbi résztémát (*könyvtárak*, *levéltárak*, *múzeumok/galériák*) pedig összevontunk *közgyűjtemények* néven, így végül 12-féle tematikus aratást futtattunk a hétfélegre elosztva, negyedéves ismétlődéssel. Az ezekben levő *seed* URL-ek, vagyis az aratórobot számára kiindulópontként megadott webcímek száma az év vége felé már meghaladta a 33 ezret. A járvány miatt otthonról végezhető távmunkában dolgozó OSZK-s kollégák segítségével sikerült ezeket a címekeket az év folyamán legalább egyszer leellenőrizni. A webcímek időnkénti felülvizsgálata azért fontos, mert sok honlap vagy blog nem egyszerűen megszűnik, hanem vagy átköltözik egy másik szolgáltatóhoz és új URL-t kap, vagy a korábban előfizetett domén címét valaki más veszi meg, így az ott található tartalom témája megváltozik, esetleg már nem is tartozik többé a webarchívum gyűjtőkörébe. Szintén először fordult elő tavaly, hogy jelentős mennyiségben kaptunk archiválásra érdemes

címjavaslatokat más forrásokból, főként a KDS pályázatban nyertes könyvtáraktól. Ezért a jelenlegi 12 részgyűjtemény bővítése mellett 8 további témakört is megnyitottunk (pl. *történelem, hely- és családtörténet; turizmus és vendéglátás; ipar és mezőgazdaság*), amelyekhez mi is gyűjtünk még címeket, és ha eléri az ezres nagyságrendet, akkor ezek is bekerülnek majd a rendszeres negyedéves aratások közé. Az egyes részgyűjteményekről XML formátumú és abból HTML-re konvertált „katalóguscédulákat”, vagyis metaadatrecordokat is készítettünk, amelyek megtekinthetők a honlapon.

Az elektronikus periodikumok műfaja kiemelt gyűjtőkörnek számít, ezekről valamivel részletesebb adatokat tartunk nyilván, és a tematikus részgyűjteménynél használt 3 szintű mélység helyett a kezdőlaptól számítva 5 szintig követjük a linkeket az aratórobottal, hogy a többnyire hierarchikus szerkezetű e-folyóiratoknál biztosan eljussunk a cikkekig. Az időszaki kiadványokat tartalmazó webhelyekből vagy webhelyrészekből november végén 5284-et tartottunk nyilván, és ezekből 4996-ot arattunk le. A többi vagy már megszűnt időközben, vagy eleve egy digitális gyűjtemény része, így tárhely-takarékosságból mi nem archiváljuk. Az év során 585 tétellel bővült ez a részgyűjtemény, jelentős arányban települési vagy regionális hírlapokkal, valamint határon túli magyar szervezetek kiadványaival. Ezt a címlistát is negyedévente mentjük újra, de 2020-ban a szerverköltözések miatt két aratás sajnos kimaradt.

A periodikumokon belül külön kategóriát jelentenek az *online hírportálok*. Ezeknél nemcsak az jelent gondot, hogy akár percenként változhat a tartalmuk, tehát a negyedéves aratás messze nem elegendő, hanem az is, hogy sok szoftverkomponenst használó, dinamikusan generált weboldalakkal állnak, amelyekkel a „buta” robotok nem boldogulnak. Ezért a fentiekben említett tömeges aratásokhoz használt *Heritrix* robot helyett elkezdtünk kísérletezni a *Brozzler* nevű szoftverrel, amely, mint a neve (*browser + crawler*) is mutatja, egy böngészőt kombinál egy, a weboldalakban és a javascriptekben talált linkeket követni képes robottal. Ezt az eszközt használja az amerikai Internet Archive *Save Page Now* nevű szolgáltatása is, mellyel bárki archiválthat weboldalakat. Többhetes tesztelés után 2020 őszén elkezdtük egyelőre 18, többféle típusú magyar online hírforrás napi szintű mentését a Brozzlerrel oly módon, hogy csak a címlap és az arról elérhető előző napi dátumú cikket töltjük le minden éjjel. Bár ez a fajta, böngészőn keresztül való archiválás lényegesen jobb eredményt

ad, de ez sem tökéletes. Az egyik nagy hátránya, hogy jóval lassabb, mint a *Heritrix*szel történő aratás, továbbá még így is akadnak megjelenítési gondok a lementett weboldaloknál. Minden próbálkozásunk ellenére a hírekbe beágyazott videókat – amelyek csak főlegesen foglalják a tárhelyet, mivel többnyire nem nézhető vissza – nem sikerült teljesen kiszűrni. A hírportáloknál is nagyobb kihívás a *közösségi média* mentése, amivel a webarchívumok világszerte küzdenek. Az ilyen típusú platformok maximálisan a felhasználó interaktivitására épülnek, és egyre inkább a mobilkészüléken futtatott alkalmazásokra optimalizálják őket, vagyis egyáltalán nem „archívumbarátok”. Többségük bejelentkezés nélkül nem, vagy csak korlátozottan jeleníti meg az amúgy nyilvános tartalmakat; csupán egy „buta” egysoros kereső van bennük, a találati lista pedig az adott felhasználó előzetes preferenciáitól is függ. A régebbi tartalmak csak az oldalak hosszas görgetésével érhetőek el; fontos funkciókhoz (pl. egy képsorozat lapozása) pedig nincsenek gyorsgombok, így a mentés automatizálásához nem lehet billentyűmakrókat használni, vagyis valakinek ott kell ülni a gép mellett, hogy görgessen és kattintgasson. Ráadásul a lementett tartalom az ezeken a weboldalon használt bonyolult, folyamatos szerver–kliens kommunikációra épülő programok miatt sokszor egyáltalán nem nézhető meg, vagy legfeljebb csak töredékesen. 2020 elején a viszonylag egyszerűbb *Instagram* archiválásával kezdtünk el kísérletezni, majd nyár végén a *Facebook*-oldalakkal, decemberben pedig a *Twitter*-csatornákkal. A *Conifer* nevű online szolgáltatást, illetve a *Webrecorder Desktop* programot használtuk ezekhez, melyek korábban még egy rendszert alkottak *Webrecorder* néven. Mindkét eszköz úgy működik, mint egy videomagnó: a böngészőben általunk megnézett oldalakat menti folyamatosan a webarchívumoknál használt szabványos *warc* fájlokba. Bár mindkettőben van egy *autopilot* funkció, amely a felhasználó helyett próbálja végiggörgetni az oldalakat és az ismertebb webkettes platformokon a linkekre is rákattint, illetve elindítja a videókat, de ez a gyakorlatban nem igazán működik, vagy ha mégis, akkor is csak addig, amíg az adott platform felületén nem módosítanak valamit a fejlesztők, ami elég gyakori ezeknél a népszerű szolgáltatásoknál. Ahogy 2017-ben a honlapok és a blogok esetében is történt, 2020-ban is elsőként a közgyűjtemények közösségi oldalaival kezdtük el a kísérletezést. Tavasszal 717 *Instagram*-fiókból töltöttünk le 110 ezer posztot: fotókat és rövid videókat (3. ábra).



3. ábra

Instagram oldalak mentése a Webrecorder Desktop programmal, illetve visszanezése a Pywb nevű megjelenítővel

Szeptember-októberben pedig 850 Facebook-oldal nyilvános idővonalát mentettük, melyek közt 147 volt könyvtári témájú, és ezeknél az egyes bejegyzéseket is letöltöttük, összesen kb. 30 ezret. Az idővonalakat igyekszünk legalább egy évre visszamenőleg legörgetni, de az esetek felénél sikerült egészen az oldal létrehozási dátumáig visszamenni anélkül, hogy lefagyott volna a böngésző. A Twitter esetében még kevés tapasztalatunk gyűlt össze, de az Instagram és a Facebook mellett erre is igaz, hogy számos technikai probléma van mind a mentés, mind a lementett fájlok indexelése, mind pedig azok megjelenítése során, nem beszélve a magas élőlomunka-igényről, így ezeknek a tartalmaknak a tömeges mentése a jelenlegi lehetőségek mellett nem tűnik fenntarthatónak.

A tematikus és műfaji részgyűjteményeken túl négy új *eseményalapú archívumot* is létrehoztunk. Ezek az események a már említett Rákóczi Emlékév, a trianoni békeszerződés 100. évfordulója, a Tokióban tervezett nyári olimpia, valamint a koronavírus-járvány. Mindegyiknél mentettünk vagy még jelenleg is mentünk teljes webhelyeket, webhely-részleteket (pl. híroldalak specifikus rovatait), illetve egyedi weboldalakat (pl. egyesével beválogatott vagy releváns címkék alapján kiszűrt híreket), sőt, esetleg hang- vagy videófájlokat és közösségi médiatartalmakat is. Ezeket az aratásokat heti vagy havi rendszerességgel futtatjuk addig, amíg az adott esemény aktuális. Még egy speciális archiválási formát érdemes megemlíteni: a *webtérszintű aratást*. Ez azt jelenti, hogy körülbelül félévente készítünk egy – csak két szint mélységű, de az általunk eddig felderített, illetve a partnerintézmények vagy magánemberek által javasolt valamennyi *seed* címre kiterjedő – „pillanatfelvételt” a magyar és magyar vonatkozású webről. A 2019 végén indított aratás 2020 januárjára is átnyúlt, a következő pedig nyár közepén zajlott le. Ez alkalommal 270 ezer URL-t adtunk a robotnak kiindulásként, amely azután öt nap alatt a linkeket követve további 46 milliót talált, melyekből 22 millió volt az új, korábban még nem letöltött fájl, 2,4 terabájt össz méretben. E cikk írásakor fut egy újabb webtérszintű aratás, ami ugyan csak 251 ezer URL-ről indult, de ezek között jóval kevesebb a már megszűnt vagy duplum cím, mert előtte automatikus módszerekkel és emberi munkával valamennyire kitisztítottuk a listát.

„Mezőgazdasági gépek”

Az elmúlt évben nemcsak a webarchívum, hanem az aratásokhoz és a begyűjtött „termés” szolgáltatásához, megjelenítéséhez használt szoftveres gépparkunk is jelentősen bővült, illetve továbbfejlődött. A már említett Webrecorder/Conifer páros és a Brozzler mellett szintén új eszközként kezdtük el tesztelni a *ReplayWeb.page* nevű, online szolgáltatásként és feltelepíthető programként is használható nézegetőt, amellyel egy warc fájl tartalmát lehet kilisztázni és böngészni. Mivel ezzel egyszerre csak egyetlen ilyen archív állomány nyitható meg, a sok ezer warc fájlból álló teljes gyűjteményhez más megoldásra van szükség. A korábban elsődleges megjelenítő eszközként beállított *OpenWayback* helyett a *Python* nyelven írt *Pywb*-re tértünk át, mivel az *OWB*-t már nem fejlesztik tovább. A *Pywb* esetében a warc fájlok lassú indexelése volt a fő gond, de úgy tűnik, hogy sikerült

erre is megoldást találunk, az *OutbackCDX* nevű indexelőre való átállással. A nyilvános archívumban

a teljes szövegű keresésre szolgáló *SolrWayback* is sokat fejlődött az elmúlt hónapokban (4. ábra).

The screenshot shows the SolrWayback search interface. At the top, the search bar contains the text "budai vár". Below the search bar, there are options for "GROUPED SEARCH", "IMAGE SEARCH", and "URL SEARCH". The search results are displayed in a list format. The first result is titled "VÁR, BUDAI VÁR ARCHEOLOGIA" and has a score of 146.5014. The result details include the type "html, web page @ archeologia.hu", the date "20/05-2019", and the URL "https://archeologia.hu/tema/v%C3%A1r/budai%20v%C3%A1r/". The highlighted content is a snippet of text: "pincék a budai vár alatt, Dunaióváros: rejtélyes építmények a Duna medrében Visegrád: római őrtornyot". Below the text, there are three image thumbnails: a stone structure, an interior view, and gold coins. The interface also shows facets on the left side, including domain, content_type_norm, type, and crawl_year.

4. ábra

Keresés a nyilvános archívum teljes szövegében a SolrWayback programmal

Ezt a rendszert Dánia nemzeti könyvtárában fejlesztik, és a jó munkakapcsolatnak köszönhetően elsőként mi kapjuk meg a legújabb verziókat, nálunk zajlik az élesben történő tesztelés, az új funkciók demonstrálása. A SolrWayback újdonságai közül érdemes kiemelni a korábbinál részletesebb találati listát, a szebben megoldott linkgráfot és térképre vetítést, az új statisztikai, adatelemző és -vizualizáló funkciókat (pl. linkgráfok *Gephi* exportja, *Ngram* grafikonok rajzolása, szövefelhők generálása).

A saját fejlesztéseink részben a rendszeres munkafolyamatok automatizálására irányultak. Ilyen például a *Kaptafa* nevű webes felület, amivel az „egy kaptafára készülő” tömeges aratásokat paraméterezzük a *Heritrix* számára; valamint a *Puppeteer* programkönyvtárra épülő szkript, mellyel a webhelyek kezdőlapjairól készítünk jó minőségű oldalképeket, azok külalakjának megőrzése céljából. Ezeken kívül a szolgáltatófelületen is számottevő fejlesztéseket végeztünk 2020-ban: különböző kereső és böngésző funkciókat építettünk be a honlapunkba. Keresni

lehet az archívum nyilvános részét alkotó webhelyek részletes, XML formátumú metaadatai, valamint a zárt gyűjtemény egyelőre táblázatokban nyilvántartott alapadatai között is. Utóbbi esetben az adatok mellett csak a kis méretű oldalképek nézhetőek meg, de az Internet Archive-ban levő, nyilvánosan elérhető mentésekre és természetesen az eredeti, „élő” webhelyekre mutató linkeket is megadtuk.

Szintén új eszköz az a virtuális gép, melyet tesztelési és oktatási célokra állítottunk össze az év második felében. Ezzel egy átlagos *Windows*-os PC-n is tudunk futtatni olyan linuxos programokat, mint a *Brozler* és a *Pywb*, illetve a Linux alatt sokkal gyorsabb *Webrecorder Desktop*. A 2020-as „404 Not Found” workshop délutáni, tutoriál jellegű részében a személyes webarchiváláshoz ajánlott *HTTrack* és *WAIL* programok mellett bemutattuk ezt a virtuális gépet is, és a következő héten nagy örömeinkre már jelentkezett is néhány könyvtár, hogy elkérje a letöltéshez szükséges linket és a rendszergazdai jelszót.

„Agrármarketing”

Az archiválás és a kutatás-fejlesztés mellett a téma „reklámozását” is fontosnak tartjuk, mert csak úgy lesz fenntartható a nemzeti könyvtár webarchívuma és majd minden további más, az internetes tartalmak megőrzésére indított hazai projekt, ha a memóriain-tézményekben dolgozó szakemberek, az illetékes döntéshozók, a tartalomszolgáltatók és a szélesebb nagyközönség is megérti, hogy kultúránk digitálisan születő részének hosszú távú megőrzésére és kutat-hatóvá tételére érdemes erőforrásokat fordítani. Ezért indítottuk el a Könyvtári Intézet segítségével azt az akkreditált tanfolyamot, melyet legutóbb 2020 elején tartottunk meg. Az őszi alkalom a korlátozások miatt elmaradt ugyan, de 2021 februárjában és márciusá-ban online formában hirdetik meg, a tananyag pedig e-learning változatban is elérhető. Emellett – szintén a KDS pályázat keretében – elkészítettük egy, első-sorban a középiskolás korosztálynak szánt ismeret-terjesztő anyag kéziratát is, amely az internetes tar-talmak személyes és az intézményi archiválásának fontosságáról és módszereiről szól.

Az elmúlt évben is sikerült több publikációt meg-jelentetnünk hazai és külföldi szaklapokban, illetve előadást tartani különféle rendezvényeken. Az ősze-re halasztott és videókonferencia formájában megrende-zett Networkshopon *Moldován István a Rákóczi Em-lékév Archivumot* mutatta be, *Németh Márton* pedig az online dokumentumokba beágyazott mikroadatokat felhasználási lehetőségeiről beszélt a webarchiválás különböző fázisaiban (letöltés, megjelenítés, megör-zés, kutatás), és egy tutoriált is tartott. A publikáció-ink közül az OSZK saját kiadványában, a *Mercuri-usban* megjelenés előtt álló cikket szeretném külön is kiemelni, mely az eddigi 404-es workshopokat ismer-teti. Említést érdemel még az angol nyelvű, *A blended learning-based curriculum on web archiving in the National Széchényi Library* (<https://doi.org/10.1108/DLP-03-2019-0012>) című 2019-es tanulmányunk, mivel 2020-ban bekerült az *Emerald Kiadó* által ki-emelten fontosnak ítélt publikációk közé és megkapta a *Literati Award* elismerést.

A szakmai kommunikáción túl tavaly több alkalom-mal sikerült a média érdeklődését is felkeltenünk az OSZK webarchívuma iránt; megjelentek rólunk sajtó-hírek (hirado.hu, profitline.hu), szerepeltünk tévében és rádióban (M1, Kossuth Rádió), így szélesebb kör-ben is fel tudtuk hívni a figyelmet erre a területre. Ez azért is fontos, mert az „internet lementése” korábban sokak részéről értetlenséget váltott ki.

„Külkapcsolatok”

A nemzetközi együttműködések fő katalizátora az International Internet Preservation Consortium (IIPC) nevű szervezet, melynek 2018 óta az Országos Széchényi Könyvtár is tagja. Az IIPC blogjában 2020 júniusában jelent meg az OSZK Webarchívumának és új honlapjának bemutatása (*From pilot to portal: a year of web archiving in Hungary* <https://netpreserveblog.wordpress.com/2020/06/>). Az internetes tartalmak megőrzésével foglalkozó intézményeket és projek-teket összefogó konzorcium 2020-as közgyűlését szintén online formában tartotta meg, ami számunkra szerencsés volt, mert az eredetileg tervezett helyszín-re, Montrealba nem lett volna lehetőségünk elutazni, így viszont mi is figyelemmel tudtuk kísérni az ese-ményt. Az IIPC munkacsoportjai közül az oktatási témájában veszünk részt, amely az év folyamán több angol nyelvű tananyagot és ismeretterjesztő prezen-tációt is közzétett. A tagintézményektől időről időre érkeznek hozzánk körlevelek, melyekben vagy egy felméréshez kérnek adatokat, vagy valamilyen speciális témájú archívumhoz helyi információforrások seed-listáit. Előbbire példaként a belga *BESOCIAL* kutatást emelem ki, amely azt mérte fel, hogy a nemzeti könyvtárak hogyan próbálják a közösségi média-tartalmakat megőrizni. Utóbbival kapcsolatban pedig a COVID-járvánnyal foglalkozó fontosabb online források összegyűjtésében segítettünk magyar rész-ről. Az általunk gondozott koronavírus-témájú rész-gyűjteményről egy külön interjú is készült és jelent meg a *WARCnet Papers* sorozatban. A *WARCNet* a 2020 és 2022 közötti időszakra kiterjedő nemzetközi projekt, központja a dán Aarhus Universtet, a célja pedig a nemzeti és nemzetközi domének történeté-nek kutatása a webarchívumokban megőrzött digi-tális tartalomra alapozva. Reményeink szerint mi is be tudunk majd kerülni az egyik munkacsoportba és publikálhatunk a kiadványaikban.

„Vetéstervek”

Végezetül csak felsorolásszerűen még néhány terv, melyek közül remélhetőleg minél több „szárba szök-ken” 2021-ben:

- Új munkatársak felvétele a megnövekedett webkurátori és -könyvtárosi feladatokhoz.
- Új részgyűjtemények összeállítása és rendszeres aratása.
- A gyűjtőkör pontosítása, belső szabályzat kidol-gozása az egyes munkafázisokhoz és a szolgál-tatáshoz.

- Az engedélyeztetések újraindítása a nyilvános szolgáltatáshoz szükséges szerződések megkötése céljából.
- A metaadat-nyilvántartás átköltöztetése egy adatbázis-kezelő rendszerbe.
- Kapcsolódás az Országos Könyvtári Platform különböző moduljaihoz.
- A zárt archívumhoz való hozzáférés lehetőségének megteremtése dedikált gépekről az OSZK-ban és később más intézményekben is.
- Előfizetés az Internet Archive-ban tárolt magyar anyag kereshető adatbázisára.
- Együttműködés kezdeményezése a közép-kelet-európai webarchívumok között.
- A hazai intézményekkel való munkakapcsolatok erősítése és kiszélesítése.

Felhasznált irodalom

„404 Not Found – Ki őrzi meg az internetet?” workshop, 2020. december 2. [online]. Budapest: Országos Széchényi Könyvtár, 2020. Hozzáférhető: <https://webarchivum.oszk.hu/404-workshop-2020-december-2/> [Megtekintve: 2021.01.04.]

DRÓTOS László. Az OSZK Webarchívumának újdonságai. == Könyvtári Figyelő, Ú.f. 30. = 66. (2020) 1., p. 67–73. ISSN 0023-3773. Hozzáférhető: <http://ki2.oszk.hu/kf/2020/04/az-oszk-webarchivumanak-ujdonsagai/> [Megtekintve: 2021.01.04.]

DRÓTOS László – MOLDOVÁN István. Az OSZK web-archíváló kísérleti (pilot) projektjének eredményei és egy üzemszerűen működő magyar webarchívum terve. == Könyvtári Figyelő, Ú.f. 29. = 65. (2019) 1., p. 38–51. ISSN 0023-3773. Hozzáférhető: <http://ki2.oszk.hu/kf/2019/04/az-oszk-webarchivalo-kiserleti-pilot-projektjenek-eredmenyei-es-egy-uzemszeruen-mukodo-magyar-webarchivum-terve/> [Megtekintve: 2021.01.04.]

DRÓTOS László – NÉMETH Márton. Az OSZK-ban folyó kísérleti webarchiválási projekt első évének tapasztalatai. == Tudományos és Műszaki Tájékoztatás, 65. (2018) 7/8., p. 389–400. ISSN 1586-2984. Hozzáférhető: <https://tmt.omikk.bme.hu/tmt/article/view/7153> [Megtekintve: 2021.01.04.]

NÉMETH Márton. Az OSZK Webarchívum új honlapjának felépítése és szolgáltatásai. == Könyv, Könyvtár, Könyvtáros, 29. (2020) 6., p. 16–26. ISSN 1216-6804. Hozzáférhető: <http://ki2.oszk.hu/3k/2020/08/az-oszk-webarchivum-uj-honlapjanak-felepitese-es-szolgaltatasai/> [Megtekintve: 2021.01.04.]

VISKY Ákos László. Gyorsmérleg az OSZK Webarchívum és a KDS-K pályázat nyerteseinek együttműködéséről. == Könyv, Könyvtár, Könyvtáros, 29. (2020) 7/8., p. 24–38. ISSN 1216-6804. Hozzáférhető: <http://ki2.oszk.hu/3k/2020/10/gyorsmerleg-az-oszk-webarchivum-es-a-kds-k-palyazat-nyerteseinek-egyuttmukodeserol/> [Megtekintve: 2021.01.04.]

Beérkezett: 2021. január 5.

Hangtár – Hatvan évnyi gramfonlemez hallgathat meg online

Több mint 2600 digitalizált hangfelvétel vált szabadon hozzáférhetővé a nagyközönség és a kutatók számára a legújabb, Hangtár elnevezésű digitális adatbázisunkban. A 2021. január 22-én, A magyar kultúra napján elindult tartalomszolgáltatással több mint 150 órányi archív hanganyag vált közkincsé a Színháztörténeti és Zeneműtár gramfonlemez-gyűjteményének anyagából.

Az összeállítás tartalma számos műfajt ölel fel: legnagyobb részben szórakoztató zenét (katonazenét, operettet, filmdalokat, tánczenét, korai dzsesszt), kisebb részben magyar nótát, népzenei vonatkozású felvételeket – köztük a Patria néprajzi gramfonfelvételek sorozatának néhány lemezét – és klasszikus zenei hanganyagokat tartalmaz, de prózai hangdokumentumok is helyet kaptak a válogatásban.

Az eddig csupán a könyvtárban kutatható dokumentumok hat évtized magyar vonatkozású hanglemez történetét ölelik fel: magyar szerzők műveit, illetve előadók felvételeit tartalmazzák, amelyek az 1904 és 1962 közötti időszakban jelentek meg hazai gyártású vagy külföldi lemezeken.

Maguk a felvételek 1902 és az 1950-es évek vége között készültek, de nem csak Magyarországon, hiszen a magyar előadóművészek régen is rendszeresen készítettek hanganyagokat külföldön, például a két világháború között Németországban vagy 1950 után Csehszlovákiában.

(Forrás: <http://www.oszk.hu/hirek/hangtar-uj-szolgaltatasunk> 2021. január 25.)