

Drótos László

Az idő fogságában

Ki őrzi meg a közösségi médiát?

A cikk napjaink legfontosabb információs és kommunikációs felületein, az összefoglaló néven közösségi médiának nevezett szolgáltatásokon megjelenő born-digital tartalmak megőrizhetőségével foglalkozik. A szerző felsorolja azokat a problémákat, amelyek miatt ezek a jelen idejű és öntörvényű „buborékuniverzumok” sokkal nagyobb kihívást jelentenek archiválási szempontból, mint a hagyományos web, majd bemutatja a szóba jöhető módszereket és technikákat. Végül ismerteti az Országos Széchényi Könyvtárban eddig elvégzett tesztelések eredményeit és egy rövid nemzetközi körképet is ad erről a speciális területről.

Tárgyszavak: weblap; digitális dokumentum; digitális archívum; internetszolgáltatás; társadalmi kapcsolat; média; kommunikáció

Problémák

Hogy a közösségi média mennyire meghatározó lett – különösen ezekben a karanténos időkben – a közélet, a kultúra, a szórakozás, a hírfogyasztás, az oktatás, a tudományos és áltudományos ismeretterjesztés, a marketing és az online kereskedelem, az emberek és egyes csoportjaik közötti kommunikáció, a különféle társas tevékenységek stb. területén, azt nem kell különösebben bizonygatni. Még aki nem is használja ezeket a szolgáltatásokat, a hagyományos médián keresztül az is értesül róla, hogy milyen aktív virtuális életet élnek emberek százmilliói ezeken a platformokon és milyen mennyiségben jelennek meg rajtuk fontos információk, kreatív alkotások, vagy csak olyan, önmagukban jelentéktelen adatok, melyek nagy tömegben hatalmas értéket képviselnek a legkülönbözőbb kutatások és fejlesztések számára, a jelenben és a jövőben egyaránt.

A digitálisan születő kultúra megőrzésének egyik módja a webarchiválás, és annak egyik speciális formája a webkettes felületeken megjelenő tartalom mentése és szolgáltatása. Már nagyon távol vagyunk az egyszerű HTML fájlokból álló eredeti World-Wide Web koncepciótól, ezért manapság a „hagyományosnak” mondható webhelyek hosszú távú megőrzése sem egyszerű; gondoljunk csak a hírportálokra, melyek mögött tucatnyi különböző

szoftverkomponens fut és állítja össze azt az oldalt, amit a felhasználó éppen meg akar nézni. Hatványozottan nehezebb az összefoglaló néven közösségi médiának nevezett szolgáltatások archiválása, mert ezeknek a platformoknak jó része valójában a webböngésző vagy valami más applikáció és/vagy a szerver által futtatott önálló alkalmazás, s mint ilyen, erősen kötődik az aktuális szoftverkörnyezethez és a szolgáltatóhoz, ezek élettartama pedig legjobb esetben is csak 10-15 év. *Így ennek a 21. századi digitális médiának lényegében csak jelen ideje van.*

E cikkben egyenértékű fogalomként szerepel a „webkettő” és a „közösségi média”, s a felhasználók tömegei által generált/feltöltött tartalomtól álló szolgáltatások, valamint a felhasználói csoportok által közösen használt kommunikációs felületek értendők alattuk. Az 1999-ben született, majd 2004-től elterjedt „web 2.0” kifejezés persze ennél tágabban is értelmezhető: a korábbi statikus, „1.0”-ás honlapok helyett megjelenő, dinamikus generált, interaktív elemeket is tartalmazó oldalakból álló webhelyek új generációját is jelentheti, és akkor ide sorolhatók a hírportálok, a webshopok, az internetes bankok, az árverési oldalak, az online játékok, az e-learning rendszerek és az egyéb modern, webalapú szolgáltatások is.

Bár a blogok, mikroblogok, fórumok, csevegő és üzenetküldő alkalmazások, wikik, kérdezz-felelek oldalak, címkéző/értékelő és véleményező felületek, ismeretségi hálók, könyvjelző- és hírmegosztók, kép/videó/hang/szöveges dokumentum és egyéb típusú fájlmosztók, térképalapú szolgáltatások, csoportmunka-rendszerek, többszemélyes videojátékok, virtuális világok és társaik is a webből nőttek ki, de egyre több közülük már csak egy vékonyodó köldökzsinóron kapcsolódik hozzá. Az olyan platformok, mint a Facebook, a LinkedIn, a Tumblr, a Twitter, az Instagram, a Flickr, a Pinterest, a SlideShare, a Foursquare, a Telegram, a Snapchat, a Discord, a Microsoft Teams, a YouTube, a TikTok, a Twitch, a Second Life, vagy az éppen divatos Clubhouse önálló „buborékuniverzumok”, saját törvényekkel, amelyek megnehezítik, vagy teljesen lehetetlenné teszik a hagyományos webre kidolgozott archiváló technológiák használatát.

Mivel a közösségi médiaszolgáltatóknak az a céljuk, hogy a felhasználók minél több időt töltsenek náluk, hogy minél több adatot lehessen róluk gyűjteni és minél több reklámot eljűk tenni, ezért nem érdekeltek abban, hogy a náluk levő tartalomban könnyen lehessen navigálni vagy annak adott pontjára ugrani, sem abban, hogy tömegesen lehessen lementeni tartalmat offline megtekintés, kutatás, vagy hosszú távú megőrzés céljából. Ezért használnak például az önálló URL-címmel rendelkező weboldalakból álló honlapokkal ellentétben „a végtelenségig” görgethető, egyetlen URL-en megjelenő üzenőfalakat és hírfolyamokat, ami olyan, mintha az oldalszámozott, lapozható könyvektől visszatértünk volna a papirusztekercsekhez. És ezért korlátozzák egyre inkább a szoftverrobotokkal megtalálható és lementhető fájlok, illetve a programozható felületen (API-n) át lekérhető adatok mennyiségét. A tömeges letöltés akadályozásának persze vannak szerzői és személyiségi jogi okai is. Mivel a tartalom nagy részét maguk a felhasználók generálják – átvéve sokszor mások anyagait is –, ezért ezeken a szervereken rengeteg a személyes adat és a védett szellemi termék. Érthető tehát, hogy ezek a „buborékok” csak korlátozottan nyilvánosak: sok esetben az amúgy publikus tartalomból is csak részletek nézhetők meg regisztrálás és bejelentkezés nélkül. Ez persze a közgyűjteményi

célú archiválást is megnehezíti, mert az aratórobotok nem tudnak bejutni ilyen helyekre; a korlátozott nyilvánosság pedig copyright- és GDPR-típusú, illetve etikai kérdéseket is felvet az archiváló intézmény szempontjából, hogy ő milyen szintű hozzáférést adhat ezekhez a lementett tartalmakhoz.

A folyamatos emberi interakciót és kliens-szerver közötti adatcserét igénylő, programmal generált weboldalak, vagy csak adott operációs rendszeren futó mobil applikációkkal elérhető tartalmak és funkciók szintén nem robot- és archívumbarrát megoldások. Mint ahogy a felhasználói felület és API folyamatos változtatása sem az, ami miatt a már bejártott archiváló technológiák egyik napról a másikra használhatatlanná válnak. Az önmegsemmisítő üzenetek (pl. Snapchat) és posztok (pl. Instagram story), a platform moderátorai és a szerveren futó felügyelő szoftver által tömegesen törölt tartalom, a kitiltott felhasználók visszamenőlegesen eltüntetett minden adata és anyaga szintén érdekes problémákat jelentenek archivátori szempontból. (A Twitter például előírja, hogy az élő szolgáltatásból törölt adatokat a letöltött adathalmazokból is törölni kellene. De ez vajon akkor is elvárható az amerikai Kongresszusi Könyvtártól, ha a kitiltott fiók tulajdonosa az Egyesült Államok elnöke, mint ahogy ez 2021 januárjában történt?) Nemcsak egyes posztok vagy fiókok tűnnek el a szervezett megőrzés esélye nélkül, hanem egész platformok zárnak be anélkül, hogy legalább a kutatók hozzáférést kaphatnának valamilyen ellenőrzött módon ahhoz a sok giga- vagy terabájthoz, amit a felhasználók milliói az évek során felhalmoztak: Windows Live Spaces (eredetileg: MSN Spaces), Delicious, Myspace, Yahoo! 360°, Vine, Orkut, Google+, Panoramio, vagy a magyar iWiW, Hotdog, Mommo, Képhost, Noob, ArtistPub, Freeblog, Blogtér ... és a sor még hosszan folytatható.

Megoldások

A közgyűjtemények számára a legjobb megoldás a webkettes tartalmak fenntartható módon való archiválására a szolgáltatóval való együttműködés lenne. Bár, mint fent említettem, a platformok tulajdonosai nem érdekeltek ebben, de azért akad erre is példa. 2010-ben a Library of Congress

megállapodott a Twitterrel, hogy 2006-ig visszamenőleg megkapja valamennyi tweet-et és azok metaadatait, előremenetben pedig hozzáférést kap a Twitter *Firehose* (tűzoltófecske) nevű csatornájához, melyen minden üzenet valós időben megjelenik. Az így keletkező adatmennyiség tárolása és kereshetővé tétele már akkor is komoly technikai kihívást jelentett a könyvtár számára, de amikor a következő években a szolgáltató cég fokozatosan enyhített azon, hogy mi számít bele az eredetileg 140 karakteres korlátba, majd az üzenetek maximális hosszát 2017-ben megduplázta, a Facebookból kiábrándult felhasználók pedig egyre több képet és videót is posztoltak a Twitterre, az üzenetfolyam mérete olyan drámaian megnőtt, hogy a Kongresszusi Könyvtár kénytelen volt feladni a teljes körű megőrzési szándékát és 2018-tól már csak válogatott archiválást végez.

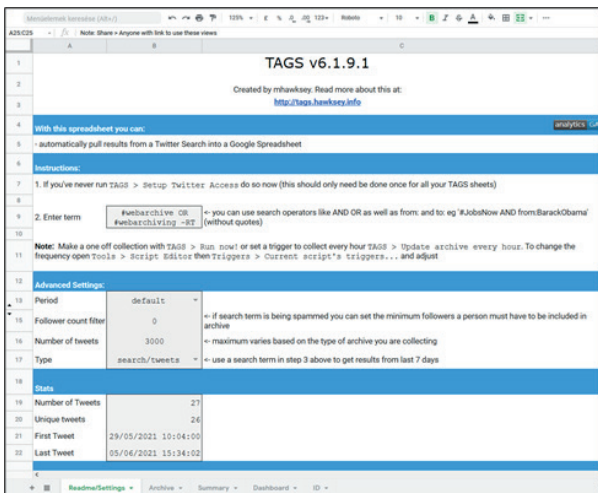
Egy másik kézenfekvőnek tűnő ötlet a fiókok tulajdonosaival való megállapodás, hogy ők maguk mentsék le a saját anyagaikat és küldjék be valamelyik illetékes könyvtárba, levéltárba, múzeumba, vagy audiovizuális archívumba. A saját tartalmak kiexportálására ma már mindegyik népszerű platformon van valamilyen lehetőség. A gond az, hogy ezek a letölthető csomagok csak a felhasználó által korábban feltöltött médiaállományokat, valamint az általa írt szövegeket (bejegyzéseket, kommenteket, üzeneteket) tartalmazzák JSON-, XML- vagy CSV-fájlok formájában. Nincsenek bennük az ezekre másoktól érkezett reakciók (lájkok, értékelések, megjegyzések, válaszüzenetek, megosztások), és természetesen nincs körülöttük az a felület sem, ahol ezek a tartalmak eredetileg megjelentek. Sokszor arra sincs mód, hogy időszeltekben kérjük az exportot, így a teljes anyagot kellene mindig lementeni és újra beküldeni, ami nem a legjobb módja egy archívum frissítésének. Tömeges archiválást ilyen egyedi megállapodásokkal nyilván nem lehet csinálni, de kiemelten fontos személyek, szervezetek és intézmények, illetve a közgyűjtemény saját webkettes fiókjai esetében járható út lehet ez a megoldás is.

Viszonylag egyszerű és akár automatizálható módszer a képernyőfotók vagy teljes oldalképek készítése, bár utóbbi reménytelen próbálkozás egy „vég-

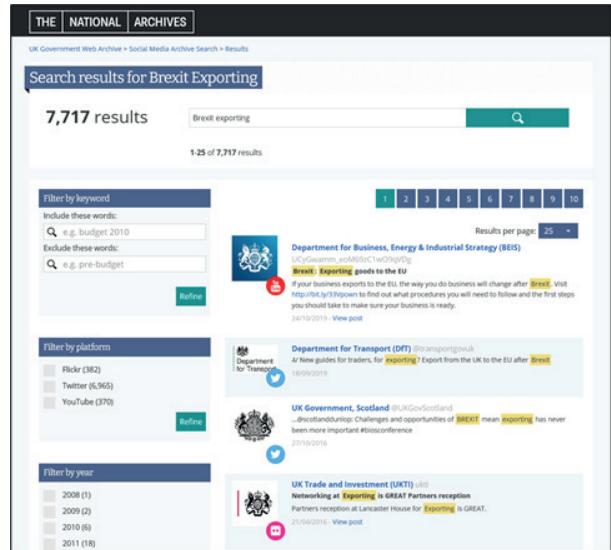
telen” hosszúságú blog-, Facebook- vagy Instagram-oldal esetén. Az olvashatóság megőrzése érdekében lehetőleg veszteségmentes tömörítéssel érdemes ezeket a képeket tárolni, ezért nagy mennyiség esetében a tárhelyigény elég jelentős lehet. Ha fontos a szöveg kereshetősége is, akkor a PDF-formátumba való mentés jöhet szóba. PNG- és/vagy PDF-fájlok készítésére ingyenes és fizetős, külön telepíthető vagy böngészőhöz hozzáadható szoftverek nagy választékban léteznek (pl. *Nimbus Screen Capture*, *GoFullPage*, *FireShot*).

Ha csak képeket, videókat, insta sztorikat, podcastokat szeretnénk lementeni a különféle fájlmegosztó oldalokról, magának a felületnek és a média-tartalomhoz tartozó szöveges információknak meg egyéb metaadatoknak az archiválása nélkül, akkor erre is vannak programok, melyek között egyaránt található a hagyományos weben is működő szoftverek, illetve kifejezetten a webkettes szolgáltatásokra optimalizált céleszközök (pl. *youtube-dl*, *Video DownloadHelper*, *DownThemAll!*, *FastVid*, *Story saver*, *Social Downloader*).

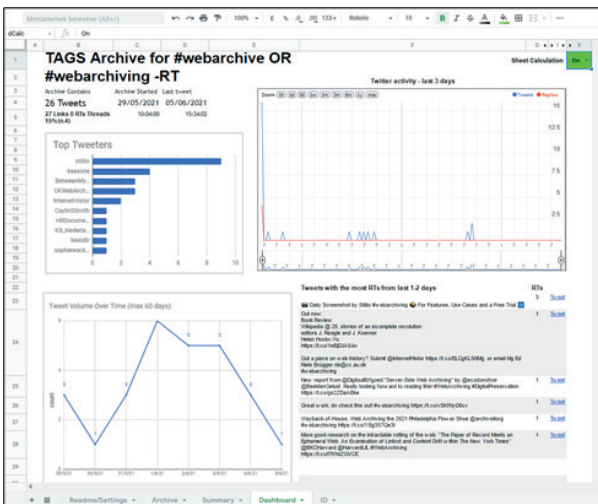
Egy következő szint, amikor a tartalom mellett a metaadatokat is (vagy csak azokat) akarjuk lementeni, szintén az eredeti felület és kontextus megőrzése nélkül. Erre a célra leggyakrabban a platform által biztosított alkalmazásprogramozási felületet, a már korábban emlegetett API-t használják. Ez egy olyan szolgáltatás, melynek kéréseket lehet küldeni egy programból és válaszként elküldi a kért fájlokat és/vagy metaadatokat. Vannak nyilvános API-k is, de a visszaélések miatt egyre gyakoribb, hogy a használatukhoz regisztrálni kell és igazolni, hogy milyen (például kutatási vagy közgyűjteményi célokra) kívánjuk igénybe venni ezt a szolgáltatást. E mellett mennyiségi korlátok is vannak általában: időben, darabszámban, vagy más módon is meg van határozva a letölthető tartalom. A Twitter például már az utolsó 7 napra korlátozza a lekérhető üzeneteket, és a Cambridge Analytica botrány után a Facebook is egyre jobban szigorítja az API-ja használatának szabályait. Néhány példa az API-n keresztül archiváló szoftverekre: *Social Feed Manager*, *Twarc*, *F(b)arc*, *TAGS (1a. és 1b. ábra)*, *Munin*, *Instamancer*. (Utóbbi határeset, mert ez nem közvetlenül az



1a. ábra A TAGS nevű GoogleDocs kiegészítő paraméterezése a webarchiválás témájú tweet-ek letöltése céljából



2. ábra A brit nemzeti levéltár közösségi médiaarchívumából származó Flickr, Twitter és YouTube találatok



1b. ábra A begyűjtött Twitter üzenetek metaadataiból készült grafikonok

API-hoz csatlakozik, hanem egy böngészőt futtat és az azzal lekért fájlokat és metaadatokat menti el a többihez hasonló módon.) Ezek a szoftverek az adatokat és egyéb szöveges tartalmakat TXT, JSON, XML, CSV, XLSX, vagy más, könnyen áttekinthető, feldolgozható és vizualizálható formátumba teszik, a médiatartalmak pedig önálló fájlokban, de a metaadatokkal összekapcsolva kerülnek eltárolásra. Így megoldható, hogy ha nem is az eredeti felületen, de egy, az archívum számára fejlesztett megjelenítő interfészen keresztül lehessen ezeket böngészni és lejátszani, ahogy ez a brit nemzeti levéltár által működtetett UK Government Web Archive esetében is van. (2. ábra)

Az eredeti böngészési élmény és környezet csak úgy őrizhető meg – többé-kevésbé hitelesen –, ha eltároljuk a weboldalt alkotó összes elemet: a HTML- és XML-állományokat, a CSS-stílusfájlokat, a betűkészleteket és a dizájnhoz szükséges minden képet és animációt, a Javascript és egyéb programkódokat, a külső szerverekről beágyazott vagy belinkelt tartalmakat, a többi felhasználó aktivitását (kommentek, lájkok, megosztások), a szolgáltató rendszere által ajánlott további bejegyzéseket, a reklámokat ... és mindezek kapcsolatait, hogy rekonstruálható legyen az oldal eredeti felépítése és külalakja. Ilyen fajta archiváláshoz olyan szoftvereket fejlesztettek ki, melyek egy böngészőt (többnyire Chrome-ot vagy Chromium-ot) futtatnak, vagy ténylegesen megnyitva azt, vagy annak csak a weboldalak „kirajzolását” végző modulját használva (ún. *headless* módban). Egy részükben van robotfunkció, vagyis automatikusan követik a HTML- és a Javascript kódokban talált linkeket, másik részük viszont csak azokat az oldalakat menti el, amelyeket az archiváló személy megnéz, és legfeljebb egy „robotpilóta” (*autopilot*) kapcsolható be, ami elvéggez néhány alaptevékenységet: végiggörgeti az oldalt, megnyit rajta minden bejegyzést és elindítja a videókat. Ide tartoznak például a következők: *Brozler*, *Browsertrix*, *WAIL* (ún. Page üzemmódban), *Webrecorder Desktop*, *ArchiveWeb.page*, *Conifer*, *Crocoite*. Ezek az esz-

közök általában az Internet Archive által kifejlesztett és 2009-ben ISO 28500 néven szabványosított WARC-formátumba mentik a fent említett sokféle fájlt és technikai metaadatot. Sajnos a szabványt nem egyformán értelmezik, így a tapasztalatok szerint gyakran vannak problémák ezeknek a webkettes oldalakról készített WARC-állományoknak az indexelésével és megjelenítésével.

Az egyes WARC-megjelenítő programok képességei ezért különösen fontosak, amikor ilyen közösségi médiatartalmakat szeretnénk visszaneézni. A legjobb eredményt az archiváló eszközhöz tartozó, ugyanazon fejlesztő által készített megjelenítőtől várhatjuk, de még ez sem mindig igaz. Az általunk eddig tesztelt programok közt vannak Windows és/vagy Linux alá telepíthető, illetve online szolgáltatás formájában (is) elérhető megoldások: *ReplayWeb.Page*, *Webrecorder Player*, *Conifer*, *PyWb*, *OpenWayback*, *SolrWayback*.

Meg kell még említeni, hogy az eddig felsorolt, teljesen vagy korlátozottan ingyenes eszközökön túl léteznek előfizetéses rendszerek is, melyeket külföldön közgyűjtemények és kutatóintézetek is használnak kisebb-nagyobb archiválási feladatok kiszervezésére. Egy részük általános webarchiváló szolgáltatás, de képes a közösségi média mentésére is a fent említett módszerek valamelyikével, és vannak köztük kifejezetten a webkettes tartalmak és adatok gyűjtésére specializálódott cégek is (pl. *MirrorWeb*, *ArchiveSocial*, *PageFreezer*, *Smarsh*, *Intradyn*, *Jatheon*, *Archive-It*).

A különböző megőrzési módokról, az archiváló eszközökről és a gyakoribb használati esetekről az amszterdami székhelyű International Institute of Social History (IISH) 2021 márciusában tett közzé egy „An overview of social media archiving tools” című jelentést, kiegészítve azt egy Gyakran Ismételt Kérdések részzel is.¹ A kanadai Ryerson University Social Media Lab nevű részlege „Social media research toolkit” néven gondoz egy már több mint 80 tételből álló nyilvántartást² a közösségi médiatartalmak vagy metaadatok begyűjtésére és elemzésére használható szoftverekről. A később még részletesebben is bemutatott belga BESOCIAL projekt keretében szintén összeállított-

tak egy összehasonlító táblázatot néhány letesztelt archiváló eszközről „Social media archiving tools” címmel.³ A University of North Carolina at Chapel Hill politikai kommunikációval foglalkozó professzora, *Deen Freelon* pedig egy „Social media data collection tools” nevű wikit szerkeszt.⁴

Saját tapasztalatok

Az Országos Széchényi Könyvtárban 2017-ben indult webarchiválási projekt első szakaszában a hagyományos web aratására koncentráltunk, de azért 2019 elején már megpróbálkoztunk az OSZK saját online szolgáltatásai, köztük a webkettes oldalai archiválásával is, majd ezeket a teszteket két évvel később megismételtük. Többféle szoftverrel mentettünk wikiket, blogokat, valamint Facebook, Google+, Indafotó, Instagram, Pinterest, Tumblr, Twitter és YouTube oldalakat, majd ezeket különböző megjelenítő felületeken tettük visszaneézhetővé. A cél nem annyira a tartalom megőrzése volt, hanem inkább csak tapasztalatgyűjtés arról, hogy az egyes platformok milyen eszközzel és mennyire archiválhatóak. Az eredmények elég kiábrándítóak lettek, a webarchívum honlapjáról elérhető demóban⁵ a rengeteg piros gomb azt jelzi, hogy sok esetben egyáltalán nem sikerült lementeni vagy megjeleníteni semmit.

Az első archívumépítési célú gyűjtésre 2020 tavaszán került sor: február végétől május közepéig kb. 700 publikus magyar Instagram fiókból minden képet és videót letöltöttünk (összesen több mint 100 ezer bejegyzést), majd ősszel frissítettük is ezt a részgyűjteményt az időközben közzétett anyagokkal. 2021 májusában húsz hazai és határon túli magyar hírportál insta posztjait (fotót, képsorozatot és rövid videót) mentettük le 15,5 GB összméretben; a legrégebbi ezek közül 2013. júliusi volt. Az Instagram viszonylag könnyen archiválható, de egy lebegő ablakban egyenként meg kell nyitni az egyes bejegyzéseket, hogy a nagyobb méretű képek, valamint a hozzájuk kapcsolódó szöveges információk, metaadatok és esetleges kommentek megjelenjenek, illetve a videók is csak ilyenkor indulnak el, s a képsorozatokot is csak így lehet átlapozni. Ha vannak kiemelt minigyűjtemények (*stories highlights*) az oldal tetején, akkor azokra

is egyenként rá kell kattintani és a jobbra nyíl gombbal továbbléptetni őket, vagy megvárni, amíg végigmegy rajtuk az automatikus diavetítés, ami percekig is eltarthat. Ugyancsak percekig kell várni néha, amikor nem sikerül betölteni egy képet, és ilyenkor továbblapozni sem lehet, mert a szerver nem válaszol addig, amíg le nem telik valamilyen időkorlát. Szintén nagyon kellemetlen tulajdonsága az Instagramnak, hogy a webes felületen a bélyegképlistán az egyes bejegyzések dátuma nem látszik, így itt nem lehet kideríteni, hogy egy félbeszakadt archiválást vagy egy későbbi időpontban történő frissítést honnan kellene folytatni.

Az Instagram mentéséhez a Webrecorder asztali gépekre telepíthető és online szolgáltatásként elérhető változatait, valamint az ArchiveWeb.page nevű szoftvert és annak Chrome böngészőbe beépülő modulját is használtuk. Mindegyiknél van *autopilot* funkció az Instagramhoz, ami megcsinálja helyettünk az oldal görgetését, a posztok megnyitását, a képsorozatok lapozását, a videók elindítását és a kommenteken is végigmegy, de sajnos ezek a beépített műveletek könnyen összezavarodnak. Ha a szerver nem válaszol időben vagy nem megfelelő adat érkezik, akkor a robotpilóta lefagy és lehet előlről kezdeni az oldal mentését, a hibás WARC-fájlt pedig törölni kell. Ezért inkább az AutoHotkey (AHK) nevű programmal futtatható billentyűmakrókkal vezéreltük az archiválási folyamatot, de ezzel a képsorozatok végiglapozását nem lehetett megoldani, mert ehhez a funkcióhoz az Instagram nem rendelt gyorsbillentyűt, ami akadálymentesítési szempontból is súlyos hiba. A Webrecorder Desktop esetében a Windows verzió nagyon lassan működött, ezért előbb az azóta Conifer névre átkeresztelt online változatával próbálkoztunk, de ott egyrészt csak 5 GB az ingyenes tárhely (igaz, kérésünkre ezt tavaly nagylelkűen megduplázták), másrészt egy idő után az Instagram letiltotta az onnan való belépést, mert gyanús volt neki, hogy egy magyar fiókba miért egy amerikai szerveren keresztül jelentkeznek be. Az első fejezetben említett szigorítások miatt ugyanis minden esetben be kellett lépünk egy erre a célra létrehozott Instagram fiókba ahhoz, hogy hozzáférjünk a lementendő nyilvános tartalmakhoz. Ezután a Webrecorder linuxos változatát használtuk egy

Windows alatt futtatott virtuális gépen, ami elég gyors, és mivel megőrzi a bejelentkezési adatokat, ezért elég csak egyszer belépni vele az Instagramra. Ennek a megoldásnak csak annyi hátránya volt, hogy az AHK-makrók nem működnek Linux alatt, így kézzel kellett megnyitni minden bejegyzést. A valódi megoldást a Webrecorder fejlesztőjének új projektje, az ArchiveWeb.page jelentette, mely 2021 februárjában jelent meg, és bár még júniusban is csak a 0.6.8-as verzióval tart, de sok szempontból már most jobb, mint az elődje. A telepíthető változata valamivel több funkciót tartalmaz, de a Chrome modullal is nagyon jól és gyorsan lehet archiválni bármilyen weboldalt, beleértve a legtöbb webkettes felületet is. A WARC formátumba való exportálás elég lassú ugyan, de ez a munkafázis már nem igényel emberi felügyeletet. Visszanéző eszköznek a *PyWb* vált be a legjobban (3. ábra), bár ennél is előfordul, hogy nem minden kép jelenik meg és nem minden videó indul el.

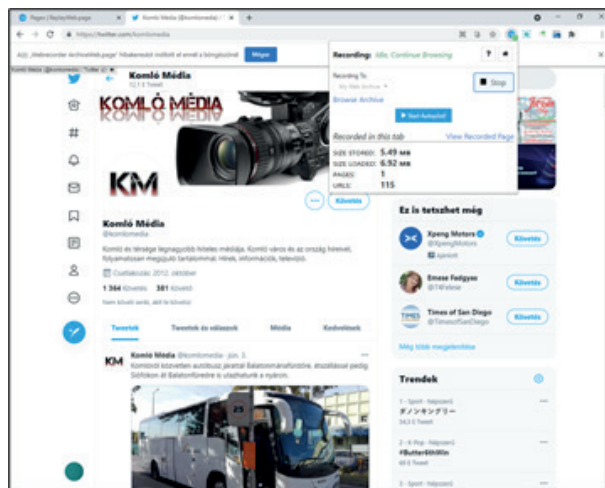
2020 őszén az egyik legnehezebben megőrizhető internetes tartalom, a **Facebook** archiválásával kísérleteztünk a Webrecorder Desktop és a Conifer segítségével. Első lépésben 147 publikus könyvtári Facebook oldalról mentettünk le közel 30 ezer bejegyzést, többnyire képpel, illetve néhol videóval együtt. 35 esetben sikerült a teljes oldalon végigmenni (a legrégebbi letöltött poszt dátuma 2010. április 28. volt), a többinél az utolsó egy-két év bejegyzéseiből mentettünk le átlagosan körülbe-



3. ábra Egy Instagram poszt az OSZK webarchívumból a PyWb megjelenítővel

lül kétszázat. Ezután 1115 db egyetemi Facebook fiók nyilvános idővonalát archiváltuk, körülbelül az oldalak felénél a legelső bejegyzésig visszamenve, egyébként pedig lehetőleg az utóbbi egy évet. A képgalériák és videók idő- és tárhely-takarékoságból nem kerültek letöltésre, de így is több mint 36 GB lett az elmentett tartalom. A fiókok tulajdonosai között hazai és határon túli egyetemek, intézetek, karok, tanszékek és egyéb szervezeti egységek, könyvtárak, valamint kulturális, sport- és diákszervezetek egyaránt találhatók. Az archiválás decemberben művészeti, irodalmi, sport és egyéb témájú oldalakkal folytatódott, majd megszakadt, részben azért, mert a Facebook dizájnváltozása miatt az *autopilot* funkció működésképtelenné vált, részben pedig, mert a cég szigorított a hozzáférési politikáján és belépés nélkül már nem lehet a publikus oldalakat sem végiggörgetni. Bejelentkezve viszont egy idő után nagyon lelassul az archiváló szoftver és a WARC-fájlt sem lehet rendesen visszanézni. További gond, hogy a Conifer esetében az Instagramhoz hasonlóan minden alkalommal újra be kell lépni és ráadásul a fiók neve és jelszava belemertődik az archív fájlba, ami biztonsági kockázat. (A Webrecorderben van erre megoldás a *preview* bekapcsolásával.) Reményeink szerint az ArchiveWeb.page segítségével idén folytatni tudjuk ennek a részgyűjteménynek a bővítését, ha nem is olyan tempóban, ahogyan 2020-ban elkezdtük, az időközben történt korlátozások miatt.

2021 áprilisában elkezdtünk egy **Twitter** gyűjteményt is építeni. Az ArchiveWeb.page Chrome böngészőbe integrálható modulját (4. ábra) és az AHK makróit használva először 89 könyvtár, múzeum és galéria oldalát töltöttük le kb. 600 MB összméretben. 75 esetben tudtunk az első bejegyzésig visszagörgetni, a legrégebbi elmentett tweet 2007. decemberi volt. Májusban pedig közel ezer további Twitter fiókot mentettünk le, ebből 594 esetben sikerült teljes egészében letölteni a tweet listát és az abba beágyazott videókat és képeket (utóbbiakat csak kis méretben). A többinél csupán az utolsó kb. 750 tweetet lehetett megnézni a rendszer korlátai miatt. A tapasztalatok szerint sok intézmény és közszereplő csak a Facebook és/vagy Instagram posztjainak linkjeit teszi ki a Twitterre, vagyis érdemi tartalom nincs a csatornájukon.



4. ábra **Twitter oldal mentése az ArchiveWeb.page nevű Chrome kiegészítővel**

A Twitter amúgy is kevésbé népszerű nálunk, a 2010-es években létrehozott fiókok elég nagy hányada már nem aktív. De azért vannak olyanok is, ahol naponta több, médiatartalommal kiegészített rövid szöveges bejegyzés jelenik meg, vagyis teljes értékű hírforrásként funkcionál a csatorna.

A közösségi médiaplatformokról begyűjtött tartalom szerzői és személyiségi jogok miatt nem érhető el nyilvánosan, ez az anyag elsősorban jövőbeli kutatási célokat szolgál. De ezek a „műfaji” címlisták a tematikus részgyűjteményekkel együtt böngészhetők a honlapunkon.⁶ Bár az OSZK webarchívumának gyűjtőkörébe nem tartozik bele a *stream* vagy *on demand* formában elérhető audio- és videotartalom (ennek megőrzése a Nemzeti Audiovizuális Archívum feladata lehetne), de a jövőben kísérleti jelleggel tervezzük podcastok és YouTube videók lementését is, hogy ezekkel kapcsolatban is tapasztalatokat szerezhessünk.

Összességében az eddigi tapasztalataink nem túl biztatóak. Rengeteg a technikai probléma és nagyon magas az élők munkai igény. Az előbbi talán idővel legalább részben megoldódik, ha az új archiváló és megjelenítő szoftverek okosabbak és kompatibilisebbek lesznek. Utóbbira pedig egy széles körű együttműködés lehet a válasz, melynek kialakításához az első lépéseket a tavalyi KDS-projektben már néhány könyvtárral megtettük és idén pedig tovább szeretnénk bővíteni ezt a kört és a megosztott munkafolyamatokat is. A cél az lenne,

hogy minden együttműködő könyvtár gondozzon egy számára fontos webhelyekből álló részgyűjteményt, és akár archiválja is ebből azokat a webkettes típusú tartalmakat, amelyeket az OSZK nem tud automatikusan lementeni. Csak így lehetne üzemszerűvé és tömegessé tenni napjaink legfontosabb információforrásainak megőrzését a jövő számára.

Külföldi gyakorlatok

Ha a közösségi média fogalmát nem kötjük a web 2.0-hoz, sőt magához a webhez sem, hanem beleértjük a még az 1980-as években megjelent Usenet-típusú fórumokat, *newsgroup*-okat is, akkor egészen régi projekteket is találni, melyek már ezek megőrzésével foglalkoztak. A fentebb említett IISH például 1995-ben kezdte el az Antenna Foundation és az Association for Progressive Communications közel háromezer hírcsoportjának archiválását. Ezek különböző társadalmi, politikai, kulturális és környezeti kérdésekkel foglalkoztak, egyebek mellett a szétesőben levő Jugoszláviában zajló háborúkkal. Az Occasio Digital Social History Archive 2,3 millió üzenetet tartalmaz, melyek 1988 és 2002 között születtek, és az archívum jelenleg is kutatható.⁷

Az SMA (*social media archiving*) témájú publikációk a 2010-es években kezdtek szaporodni a nemzetközi szakirodalomban, amint a mai értelemben vett közösségi média egyre fontosabbá vált és amint a már akkor is webarchiválással foglalkozó szakemberek rájöttek, hogy az eddig bevált eszközök és munkafolyamatok nem megfelelőek a modern webkettes platformok anyagának lementésére és kutathatóvá tételére. Az érdekesebb cikkekből és jelentésekből az irodalomjegyzékben található egy válogatás. Itt csak egy esettanulmányt⁸ ismertettek röviden, ami a British Library (BL) blogján jelent meg 2020 májusában, mert jól illusztrálja azokat a nehézségeket, amelyekről a jelen cikk is szól.

A 2019. december 12-i parlamenti választásokat követő napon a BL webarchiváló munkacsoportja megkísérelte 17 politikai vezető Twitter, Facebook és Instagram oldalát letölteni. Ez összesen 44 közösségi fiókot jelentett, mert nem mindegyik politikai volt jelen mindhárom platformon. Az archivá-

láshoz 12 irodai PC-t használtak, melyeken a Webrecorder-t futtatták először *autopilot* módban, majd pedig ha az hibára futott, akkor a programot újraindítva emberi közreműködéssel. Az egész folyamat nagyon lassú és nehézkes volt, így a munka egy nap helyett közel egy hétig tartott. (Például *Jeremy Corbyn* tweetjei közül csak az utolsó egy hónapnyi letöltése közel tíz órát igényelt és így is mindössze 93 megabájtnyi tartalmat sikerült leszedni. Ugyanakkor egy másik ugyanolyan gépen *Nigel Farage* Twitter oldaláról valamivel több mint négy óra alatt 506 MB lett letöltve.) Mivel már az elején világossá vált, hogy a tervezett idő alatt nem fognak végezni az archiválással, ezért a helyi hálózaton elérhető további szabad PC-ken is elindították a Webrecorder-t, így egy-egy munkatárs több gépen is dolgozott egyszerre. Ez gyorsított a folyamaton, viszont így meg a közös hálózati tárhelyre mentett, rejtélyes nevű WARC-fájlok ellenőrzése, a hibások törlése és a jók átmozgatása, majd a sikeres mentések adminisztrálása igényelt sok időt. A gyűjtemény az Egyesült Királyság kötelezpéldány jogosultságú könyvtárainak olvasótermeiben férhető hozzá, illetve a UK Web Archive részeként az internetről is elérhetők azok az archivált fiókok, amelyeknek nyilvános szolgáltatására engedélyt kapott a British Library.

Belgium nemzeti szintű könyvtára, a Koninklijke Bibliotheek 2017–2019 között egy PROMISE nevű kutatás keretében megvizsgálta a belga webtér archiválásával kapcsolatos kérdéseket, majd ennek folytatásaként 2020-ban három egyetemmel együttműködve egy újabb kétéves projektet indított BESOCIAL néven⁹ a közösségi web megőrizhetőségének felmérése és egy fenntartható archiválási stratégia kidolgozása céljából. A hét fázisra tervezett munka elején a Belgiumban és más országokban már létező közösségi médiaarchívumokat mérték fel irodalomkutatás, kérdőívek és interjúk segítségével. Az erről készült jelentés végső változata 2021 áprilisában került ki az internetre „WP1 report: an international review of Social Media Archiving initiatives” címmel.¹⁰ Ez csak a nemzetközi gyakorlatot összegzi, a belgiumi helyzetértékelést később publikálják. A felmérés során 12 ország 15 intézményét/archívumát keresték meg, köztük Magyarországról az OSZK-t is. A kérdések köre a váloga-

tásra, a technikai megoldásokra, a hozzáférésre, a hosszú távú megőrzésre és a jogi vonatkozásokra egyaránt kiterjedt.

Az általános tanulság az, hogy az egyes országokban nagyon eltérő a jogi környezet és részben ez, részben pedig a technikai lehetőségek és az emberi erőforrások határozzák meg, hogy mit és hogyan archiválnak, illetve hogy mihez és hogyan adnak hozzáférést a nemzeti könyvtárak, vagy más közgyűjtemények és tudományos intézmények. A webkettes oldalokról begyűjtött tartalmak hosszú távú megőrzésével, esetleges migrálásával pedig még sehol nem foglalkoznak érdemben. Ahogy a mellékelt táblázatból (5. ábra) kiderül – amely a „WP1 report” összesítése alapján készült a magyar helyzet időközbeni aktualizálásával –, a legtöbb helyen archivált platform a Twitter, amit vagy API-n keresztül, vagy a Heritrix robotjával, vagy pedig valamilyen böngésző-alapú megoldással mentenek. Az is látszik az ábrán, hogy vannak országok, ahol még egyáltalán nem, vagy csak minimálisan foglalkoznak a könyvtárak és levéltárak a közösségi média megőrzésével, többnyire a törvényi felhatalmazás hiánya miatt.

Néhány további információ szintén ebből a jelentésből: Az új-zélandi nemzeti könyvtár 2009-ben hozta létre a Twitter archívumát. A British Library

Ország	Intézmény típusa	Facebook	Twitter	YouTube	Instagram	Flickr
Dánia	királyi könyvtár	☹	☺	☺	☺	☹
Egyesült Államok	egyelemi könyvtár	☹	☺	☹	☹	☹
Egyesült Királyság	könyvtári konzorolium	☺	☺	☹	☹	☹
Egyesült Királyság	nemzeti levéltár	☹	☺	☺	☹	☺
Észtország	nemzeti könyvtár	☺	☹	☹	☹	☹
Franciaország	nemzeti könyvtár	☺	☺	☺	☺	☹
Franciaország	nemzeti audiovizuális archívum	☹	☺	☺	☹	☹
Hollandia	nemzeti levéltár	☹	☹	☹	☹	☹
Hollandia	királyi könyvtár	☹	☹	☹	☹	☹
Írország	nemzeti könyvtár	☹	☺	☺	☹	☹
Kanada	nemzeti könyvtár és levéltár	☺	☺	☺	☺	☺
Kanada	regionális könyvtár és levéltár	☺	☺	☹	☹	☹
Luxemburg	nemzeti könyvtár	☺	☺	☺	☹	☹
Magyarország	nemzeti könyvtár	☺	☺	☹	☹	☹
Svájc	nemzeti könyvtár	☹	☹	☹	☹	☹
Új-Zéland	nemzeti könyvtár	☺	☺	☹	☺	☹

5. ábra A BESOCIAL jelentés táblázata alapján készült, némileg frissített összeállítás a projekt keretében megkérdezett intézmények webkettes archívumairól, ☺ = rendszeresen archivált, ☹ = kísérleti jelleggel archivált, ☹ = nem archivált

2010-től kezdett bele a rendszeres webkettes archiválásba, korábban csak korlátozott mennyiségű Twitter, Facebook és YouTube tartalmat mentettek le. A brit levéltárnak egészen 2008-ig visszamenőleg vannak archivált Twitter fiókjai. A francia nemzeti könyvtár a webarchívuma létrehozásakor, már 2006-ban elkezdett foglalkozni a Facebook mentésével is, de a platformon történt változtatások miatt 2010-ben kénytelenek voltak leállni ezzel a munkával. A szintén francia audiovizuális archívum, az Institut national de l’audiovisuel a YouTube mellett a Dailymotion, a Vimeo és a Soundcloud oldalairól is gyűjt video- és hangfelvételeket.

A BESOCIAL felmérés során sok érdekes kérdés is felmerült, melyekre nincsenek egyszerű válaszok. Például, hogy hogyan határolható be a „nemzeti web” az olyan multinacionális platformokon, mint a Facebook és társai? A csak egyetlen nép által használt nyelvek esetében, mint a mienk, a nyelv egy jó szűrő lehet, de persze a magyarok sem feltétlenül magyarul írnak ezeken a felületeken. És bár a Twitter keresőjében lehet ugyan nyelvre is szűrni, de a tapasztalatok szerint az eredmény nagyon „zajos”, mivel a rövid és esetleg ékezet nélküli szövegeken az automatikus nyelvelismerés nem működik megbízhatóan. A másik módszer a hashtag alapján való szűrés lenne, de ez is reménytelen olyan nemzetközi szavaknál, mint a #Covid, az országjelzést (pl. HU) pedig ritkán teszik be az emberek a hashtag-ekbe. További kérdés, hogy meddig terjed egy poszt? Egy Twitter üzenethez hozzátartozik-e például az abban szereplő rövidített URL-en található weboldal, a belinkelt vagy beágyazott kép/videó? Eltegyük-e a „hivatalos” csatornák tartalmához érkezett kommenteket és egyéb felhasználói interakciókat? Ha igen, akkor anonimizáljuk-e ezeket a személyiségi jogok és a jó hírnév védelme érdekében?

Felmerül a reprezentativitás problémája is: például olyan arányban archiváljuk-e a járvány- és oltástagadók posztjait, mint amilyen mennyiségben ezekkel elárasztják a közösségi médiát, vagy ami a feltételezhető tényleges arányuk a társadalomban? Mit kezdünk a tudatlanságból vagy szervezett módon terjesztett fake news tartalmakkal? És a szélsőségekkel? Bevonhatók-e valahogy a közösségi

rendszerek használói és egyes civil szervezetek a megőrzésre érdemes dolgok válogatásába, sőt akár archiválásába is? Vagy vannak-e automatikus megoldások a valóban fontos/érdekes tartalmak kiszűrésére abból a hatalmas „zajból”, ami ezeken a platformokon van? Hogyan lehet archivátorként gyorsan reagálni a váratlan események hatására futótűzszerűen meginduló információáradatra? És az ugyancsak gyorsan eltűnő, a közzététel vagy a szolgáltató által törölt tartalmakra? Biztosítani kell-e a felejtéshez való jogot egy, csak kutatók által hozzáférhető zárt webarchívum esetében is?

Irodalom

Magyar nyelven:

Drótos László: A közösségi média archiválásának nehézségei. In: Networkshop 2021 konferenciakötet (megjelenés alatt!) Az előadás videofelvétele: <https://kifu.videotorium.hu/hu/recordings/42378> [Hozzáférés: 2021. június 7.]

Drótos László. Az OSZK webarchívumának 2020-as újdonságai. In: Könyvtári Figyelő, 2021. (67 évf.), 1. sz. pp. 31-38. <http://ojs.elte.hu/kf/article/view/2296> [Hozzáférés: 2021. június 7.]

Drótos László – Németh Márton: Egyedi mentésekre szolgáló webarchíváló szoftverek. In: Könyv, Könyvtár, Könyvtáros, 2020. (29. évf.), 12. sz. pp. 3-11. <http://ki2.oszk.hu/3k/2020/12/egyedi-mentesekre-szolgalo-webarchivalo-szoftverek/> [Hozzáférés: 2021. június 7.]

Drótos László (ref.): Zimmer, Michael: A Kongresszusi Könyvtár Twitter archívuma. In: Tudományos és Műszaki Tájékoztatás, 2015. (62. évf.), 11-12. sz. pp. 445-447. <https://tmt.omikk.bme.hu/tmt/article/view/610/581> [Hozzáférés: 2021. június 7.]

Angol nyelven:

Acker, Amelia – Kreisberg, Adam: Social media data archives in an API-driven world. In: Archival Science, 2020. (20. évf.), 2. sz. pp. 105-123. <https://link.springer.com/article/10.1007/s10502-019-09325-9> [Hozzáférés: 2021. június 7.]

Acker, Amelia – Kriesberg, Adam: Tweets may be archived: civic engagement, digital preservation and Obama White House social media data. In: Proceedings of the Association for Information Science and Technology, 2017. (54. évf.) 1. sz. pp. 1-9. <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pr2.2017.14505401001> [Hozzáférés: 2021. június 7.]

Ben-David, Anat: Counter-archiving Facebook. In: European Journal of Communication, 2020. (35. évf.) 3. sz. pp. 249-264. <https://doi.org/10.1177/0267323120922069> [Hozzáférés: 2021. június 7.]

Nyilvánvalóan különböző szempontokat kellene itt összeegyeztetni, mert mások az érdekei és céljai a szolgáltatóknak, a tartalomgyártóknak és -fogyasztóknak, a szellemi tulajdont és a személyiségi jogokat védőknek, a kutatóknak és az archiválóknak. Egy biztos csak: a közösségi oldalak napjaink legfontosabb, de legnehezebben megőrizhető információterjesztő és kommunikációs felületei. Memóriaintézményként a háborút már elvesztettük, de még kisebb csatákat megnyerhetünk, hogy a digitális jelenünket ne falja fel teljesen az idő, mint *Stephen King* kisregényében a langolierék.

Breed, Miana: Capturing a Moment: The Practices and Ethics of Social Media Archiving. Chapel Hill : University of North Carolina, 2019. <https://cdr.lib.unc.edu/downloads/b27740677> [Hozzáférés: 2021. június 7.]

Bruns, Axel: After the 'APocalypse': Social media platforms and their fight against critical scholarly research. In: Information, Communication & Society, 2019. (22. évf.) 11. sz. pp. 1544-1566. <https://snurb.info/files/2019/After%20the%20%27APocalypse%27.pdf> [Hozzáférés: 2021. június 7.]

Fondren, Elisabeth – McCune, Meghan Menard. Archiving and preserving social media at the Library of Congress: institutional and cultural challenges to build a Twitter archive. In: Preservation, Digital Technology & Culture, 2018. (47. évf.), 2. sz. pp. 33-44. https://www.academia.edu/download/58436184/Archiving_and_Preserving_Social_Media_at_the_Library_of_Congress_PDTC_December_2018.pdf [Hozzáférés: 2021. június 7.]

Hemphill, Libby – Hedstrom, Margaret L. – Leonard, Susan Hautaniemi: Saving social media data: Understanding data management practices among social media researchers and their implications for archives. In: Journal of the Association for Information Science and Technology, 2021. (72. évf.) 1. sz. pp. 97-109. <https://asistdl.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/asi.24368> [Hozzáférés: 2021. június 7.]

Hockx-Yu, Helen: Archiving social media in the context of non-print legal deposit. IFLA WLIC Libraries, Citizens, Societies: Confluence for Knowledge, Lyon, 2014. <http://library.ifla.org/999/1/107-hockxyu-en.pdf> [Hozzáférés: 2021. június 7.]

Huang, Qunying – Xu, Chen: A data-driven framework for archiving and exploring social media data. In: Annals of GIS, 2014. (20. évf.), 4. sz. pp. 265-277. <https://www.tandfonline.com/doi/pdf/10.1080/19475683.2014.942697> [Hozzáférés: 2021. június 7.]

Jeffrey, Stuart: A new digital dark age? Collaborative web tools, social media and long-term preservation. In: World Archaeology, 2012. (44. évf.) 4. sz. pp. 553-570. <http://radar.gsa.ac.uk/2857/1/00438243.2012.pdf> [Hozzáférés: 2021. június 7.]

Jules, Bergis – Summers, Ed – Mitchell, Vernon Jr.: Documenting The Now white paper : Ethical considerations for archiving social media content generated by contemporary social movements: challenges, opportunities, and recommendations. DocNow, 2018.

<https://www.docnow.io/docs/docnow-whitepaper-2018.pdf> [Hozzáférés: 2021. június 7.]

Kavvadia, Zefi: An overview of social media archiving tools. Amsterdam : IISH, 2020

<https://confluence.socialhistoryservices.org/display/ESMAT/Final+report+-+An+Overview+of+Social+Media+Archiving+Tools> [Hozzáférés: 2021. június 7.]

Kelly, Mat et al.: On the change in archivability of websites over time. In: International Conference on Theory and Practice of Digital Libraries. Berlin, Heidelberg : Springer, 2013. pp. 35-47. <https://arxiv.org/pdf/1307.8067> [Hozzáférés: 2021. június 7.]

Lindström, Lisa. Archiving in the era of online activism: challenges and practices of collecting and providing access to activist social media archives. Lund : Lund University, 2019.

<https://lup.lub.lu.se/student-papers/record/8980793/file/8980795.pdf> [Hozzáférés: 2021. június 7.]

Littman, Justin et al.: API-based social media collecting as a form of web archiving. In: International Journal on Digital Libraries, 2018. (19. évf.), 1. sz. pp. 21-38.

<https://link.springer.com/article/10.1007/s00799-016-0201-7> [Hozzáférés: 2021. június 7.]

Lucas, Christina: Digital ethics framework recommendations for social media archiving applications. Atlanta : Georgia Institute of Technology, 2018.

<https://smartech.gatech.edu/bitstream/handle/1853/59852/LUCAS-THESIS-2018.pdf> [Hozzáférés: 2021. június 7.]

Macnaught, Bill: Social media collecting at the National Library of New Zealand. IFLA WLIC Transform Libraries, Transform Societies. Kuala Lumpur, 2018.

<http://library.ifla.org/2274/1/093-macnaught-en.pdf> [Hozzáférés: 2021. június 7.]

Hivatkozások

[Hozzáférés: 2021. június 7.]

- 1 <https://confluence.socialhistoryservices.org/display/ESMAT/>
- 2 <https://socialmediadata.org/social-media-research-tool-kit/>
- 3 <https://docs.google.com/spreadsheets/d/1nGuTC9Ww5y-WZQ0wSUPPnITMJBF1JyEaDOCO0Ve-O9U/>
- 4 <http://socialmediadata.wikidot.com/>
- 5 <https://webarchivum.oszk.hu/oszk-s-archivum-kezdolap/#webkettes>

Marshall, Catherine C. – Shipman, Frank M.: On the institutional archiving of social media. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. 2012. pp. 1-10. <https://www.microsoft.com/en-us/research/wp-content/uploads/2012/06/jcdl12-loc-final.pdf> [Hozzáférés: 2021. június 7.]

Michel, Alejandra et al.: Towards a sustainable social media archiving strategy for Belgium: BESOCIAL : WP1 report : an international review of Social Media Archiving initiatives. Brüssel : KBR, 2021.

<https://researchportal.unamur.be/en/publications/towards-a-sustainable-social-media-archiving-strategy-for-belgium> [Hozzáférés: 2021. június 7.]

SalahEldeen, Hany M. – Nelson, Michael L.: Losing my revolution: How many resources shared on social media have been lost?. In: International conference on theory and practice of digital libraries. Berlin, Heidelberg : Springer, 2012. pp. 125-137.

<https://arxiv.org/pdf/1209.3026> [Hozzáférés: 2021. június 7.]

Thomson, Sara Day – Kilbride, William: Preserving social media: The problem of access. In: New Review of Information Networking, 2015. (20. évf.) 1-2. sz. pp. 261-275.

<https://eprints.gla.ac.uk/111476/1/111476.pdf> [Hozzáférés: 2021. június 7.]

Thomson, Sara Day: Preserving Social Media: applying principles of digital preservation to social media archiving. 2017.

https://archivedweb.blogs.sas.ac.uk/files/2017/06/RESAW2017-Thomson-applying_principles_of_digital_preservation_to_social_media_archiving.pdf [Hozzáférés: 2021. június 7.]

Velte, Ashlyn: Activist social media archiving: Practices, challenges, and opportunities. Chapel Hill : University of North Carolina, 2016.

<https://cdr.lib.unc.edu/downloads/wp988p676> [Hozzáférés: 2021. június 7.]

Ye, Yunshan et al.: Web-Archiving Chinese Social Media: Final Project Report. In: Journal of East Asian Libraries, 2017. 165. sz. pp. 93-112.

<https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2701&context=jeal> [Hozzáférés: 2021. június 7.]

6 <https://webarchivum.oszk.hu/tematikus-aratasok/>

7 <https://socialhistoryportal.org/news/articles/109834>

8 <https://blogs.bl.uk/webarchive/2020/05/using-webrecorder-to-archive-uk-political-party-leaders-social-media-after-the-uk-general-election-2.html>

9 <https://www.kbr.be/en/projects/besocial/>

10 <https://researchportal.unamur.be/en/publications/towards-a-sustainable-social-media-archiving-strategy-for-belgium>

Beérkezett: 2021. V. 25-én.



Drótos László

könyvtáros
OSZK – Webarchiválási Osztály
E-mail: drotos.laszlo@oszk.hu