

A közösségi média archiválásának nehézségei

Drótos László
Országos Széchényi Könyvtár Webarchiválási Osztály
drotos.laszlo@oszk.hu

The difficulties of archiving social media

Web 2.0 sites designed for user-uploaded textual and audiovisual content began to grow exponentially in the 2000s, and today they have become the most important communication channels for culture and public life, beyond the sharing of private information targeted at a narrow audience. However, these platforms are even more “in present tense” than the traditional web. And this is not just about publicly published and then removed content, but about dozens of services with millions of users that have been discontinued for economic reasons, with gigabytes or terabytes of digital material disappeared overnight. Nowadays most social platforms are more optimized for mobile devices, and not archive-friendly at all, or even explicitly prevent downloading by automatic methods. This paper details the problems of their long-term preservation based on the tests carried out within the framework of the web archiving project in the National Széchényi Library, as well as international examples.

Keywords: web archiving, long-term preservation, social media, web 2.0

1. Öntörvényű világok

A World Wide Web megjelenésével az 1990-es évek elején egy új digitális univerzum született meg, mely azóta is fénysebességgel tágul. Weboldalak milliárdjai jöttek létre és tűntek el nyomtalanul, kivéve azokat, amelyeket sikerült – bár sokszor csak töredékesen – elmenteni az olyan archívumokba, mint az 1996-ban indult Internet Archive¹. A világháló első tíz évében jellemzően fájlrendszerben tárolt, statikus HTML oldalakból állt, szintén fájlként beágyazott médiatartalmakkal, az interaktív funkciókat pedig egyszerű CGI szkriptekkel oldották meg. Ha eltekintünk az olyan, a mai böngészők által már nem támogatott technológiáktól, mint a VRML, a Java-applet, a Microsoft Silverlight, vagy az idén nyugdíjba vonult Adobe Flash, az 1.0-ás web viszonylag robot- és archívumbarát volt, jól menthető automatikus módszerekkel és évtizedek múlva is visszanezhető.

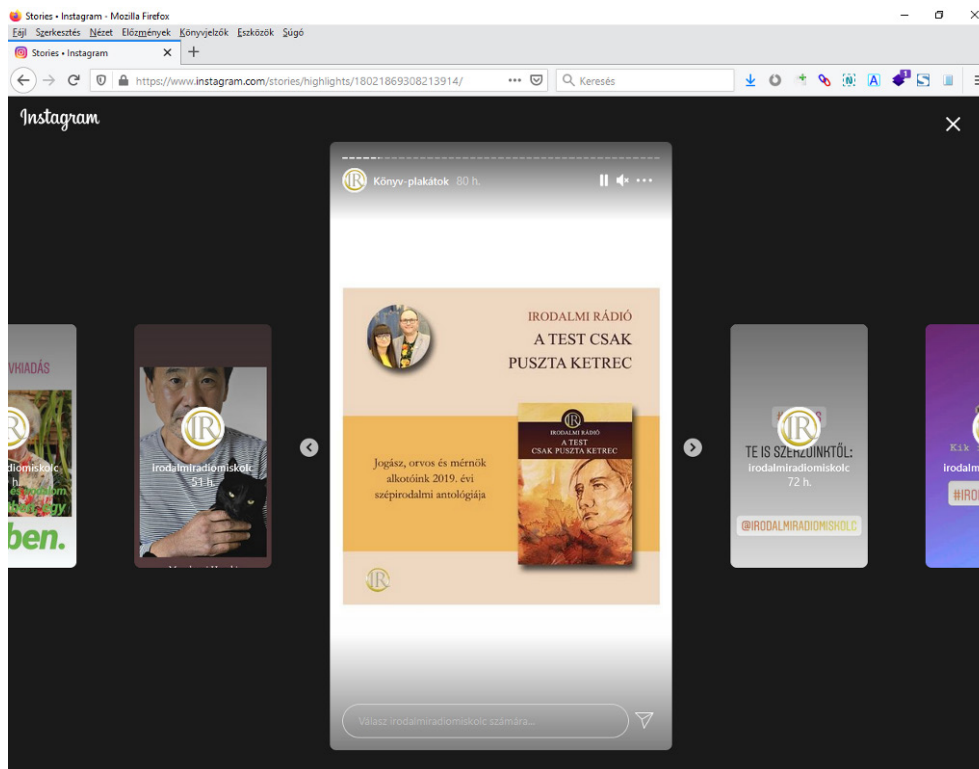
A 21. század internetjét viszont már az adatbázisokból dolgozó programok által generált „webkettes” platformok dominálják, azokon zajlik az igazi élet, ott jelenik meg a legtöbb új tartalom és onnan is tűnik el a leggyorsabban sok minden. Ezek a közösségi terek öntörvényű univerzumokként léteznek a hagyományos web mellett, hasonlóan a már nemcsak sci-fi ötletként, hanem tudományos elméletekben is megjelenő buborék univerzumokhoz. A fizikusok szerint ezekben a feltételezett párhuzamos világegyetemekben eltérő törvények uralkodnak, például más a fény sebessége vagy az elektron tömege, és emiatt teljesen más struktúrák alakulhatnak ki, mint az általunk

1 Wayback Machine, Internet Archive. Hozzáférés: 2021.05.10., <http://web.archive.org>



ismert atomok és égitestek. Az új generációs webes szolgáltatásokat üzemeltető cégek is saját világokat alakítottak ki és ezeknél sajnos egyáltalán nem szempont a hosszú távú megőrizhetőség és a kutathatóság. Az alábbi lista csak néhányat sorol fel a sokféle, archiválhatósági szempontból problémás megoldás közül:

- Egyre több helyen a nyilvános tartalom elérése is belépéshez kötött, regisztráció és azonosítás nélkül csak korlátozottan lehet böngészni a publikus felületen. Az ilyen zárt rendszerekbe sem a keresőgépek, sem a webarchívumok által futtatott szoftverrobotok nem tudnak bejutni. Az API-n keresztül való adatkinyerést is mind jobban korlátozzák a szolgáltatók a 2010-es évek közepétől a Cambridge Analytica-ügy és azt követő hasonló botrányok után.
- A nagy webes keresők robotjainak a kizárása vagy erős korlátozása miatt az archiválásra érdemes tartalom csak a platform saját belső keresőjével található meg, az viszont személyre szabott találatokat ad a felhasználó és ismerősei korábbi preferenciái alapján.
- Alternatív felületek léteznek, melyek nem teljesen ugyanazokat a tartalmakat és funkciókat szolgáltatják asztali és mobil böngészőkben, illetve applikációk formájában.
- Egyre gyakoribb, hogy már csak saját app-on keresztül érhető el a szolgáltatás, a honlapján mindössze egy letöltő link található az alkalmazásboltra. Ezek az applikációk persze egyáltalán nem jövőállóak és csak egy-két mobil operációs rendszer adott verzióin működnek.
- Egy oldalon belül tulajdonjogilag kevert a tartalom, mert az algoritmus a felhasználó számára relevánsnak gondolt más híreket és posztokat is mellé tesz, így hiába van engedélye az archiválónak egy fiók mentésére, nem tudja kizárni a más forrásokból odakerült elemeket.
- A bejegyzések és feltöltések mellett sok személyes adat is megjelenik (pl. kommentek, lájkok, nicknevek, avatar képek), melyeknek az archiválása és szolgáltatása GDPR-kérdéseket vet fel.
- A tartalomhoz való hozzáféréshez folyamatos emberi közreműködés szükséges (pl. „vég nélküli” görgetés vagy lapozás) és folyamatos a háttérben zajló adatcsere is a kliens és a szerver között. Ezek a megoldások nagyon megnehezítik a tartalom automatikus aratását, illetve a lementett anyagnak az eredetihez hasonló megjelenítését.
- Az ugyanazon az URL címen a görgetés vagy lapozás hatására megjelenő, illetve a saját URL nélküli, beágyazott vagy felugró tartalmi elemek a fájlok beazonosíthatóságát teszik lehetetlenné az archívumban. (1. ábra) A másik véglet az ugyanahhoz az elemhez véletlenszerűen generált URL-ek tömege, ami miatt az aratószoftver azt hiszi, hogy mindig új tartalmat tölt le, pedig valójában egy robotcsapdában kering.



1. ábra. Az Instagram *Stories Highlights* funkciója ugyanazon az URL címen jeleníti meg a képeket és videókat, így ezeket sem linkelni, sem robottal bejárni nem lehet.²

- Az idő is egészen sajátosan múlik ezekben az univerzumokban, a jelenidejűségük miatt tipikus az ilyen dátumkijelzés az egyes bejegyzések mellett: „40 perce” vagy „2 hónapja”, aminek természetesen semmi információtartalma nincs, ha egy archivált verziót nézünk vissza. Az pedig még az élő szolgáltatásnál is nagyon zavaró, amikor nem lehet adott dátumra ugrani a hírfolyamban. Az archiválást emiatt nem, vagy csak hosszas görgetés után tudjuk onnan folytatni, ahol korábban abbahagytuk.
- Az önmegsemmisítő üzenetek (pl. Snapchat) és posztok (pl. Instagram story) megítélése már nem ennyire egyértelmű, mert egyrészt érthetően szükség van ilyenekre a magán jellegű kommunikáció védelme érdekében, de amikor egy közszereplő a nyilvánosság számára tesz közzé időzített módon törlődő tartalmat, akkor igen kicsi az esély arra, hogy az bekerül valamilyen közgyűjteményi archívumba. Persze sokszor vannak élelmes felhasználók és újságírók, akik még időben csinálnak róla másolatokat, akár csak azokról a tartalmakról, amiket maga a közlétező töröl le, amikor azok kezdenek kellemetlenné válni számára.
- A platform üzemeltetője vagy az általa futtatott szoftver algoritmusai is folyamatosan és tömegesen töröl tartalmat a nyilvános felületről a közösségi elvek vagy a jogszabályok (pl. a szerzői jog) megsértése miatt. Gyakori az is, hogy felhasználókat tiltanak le és a teljes fiókjukat zárolják akkor is, ha annak csak kis része volt szabálysértő. Így járt az Egyesült Államok elnöke is 2021 elején, akit előbb a Twitter, majd az összes többi nagy közösségi platform kitiltott.

² *irodalmiradiomiskolc* – Az Irodalmi Rádió szerkesztő-házaspárjának könyvélményei, Instagram. Hozzáférés: 2021.02.21., <https://www.instagram.com/irodalmiradiomiskolc/>



- De 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 (pl. iWiW, Hotdog, Myspace, Google+, Panoramio ... és a sor még hosszan folytatható).
- Arra ma már a legtöbb webkettes szolgáltató lehetőséget biztosít, hogy a használói a saját tartalmaikat kimenthessék a rendszerből, így a személyes célú archiválás legalább megoldható. Természetesen ezekből az export csomagokból az eredeti felület és kontextus nem rekonstruálható, csak a korábban feltöltött médiafájlok vannak bennük, meg a szöveges tartalmak JSON, XML, vagy valamilyen más, nem túl felhasználóbarát formában.
- Közgyűjteményi, webkurátori szempontból a legbosszantóbb jellemzője ezeknek az öntörvényű világoknak, hogy még a saját szabályaikat sem tartják be. Mire sikerül kidolgozni egy olyan technológiát és munkafolyamatot, amivel a kulturális vagy tudományos szempontból értékes/érdekes tartalmakat megfelelő mennyiségben és minőségben archiválni lehet, addigra a platform változtat valamint a felületén, szigorít a hozzáféréseken, átalakítja a képek és videók megjelenítő modulját ... és megint el lehet kezdeni az archiváló és visszanező szoftverek hozzáigazítását a megújult rendszerhez. (Ez történt tavaly is, amikor a Facebook dizájnt váltott.)

2. OpenSocial

A 2007-ben elég nagy médiavisszhangot kapott OpenSocial³ a közösségi média szabványosítására tett első komolyabb kísérlet volt. A kezdeményezésben a Google tulajdonú Orkut mellett a Friendster, a LinkedIn, a MySpace és néhány más közösségi platform vett részt, majd 2014-től a W3C is bekapcsolódott a munkába⁴ és az Open Web Platform projektjük keretében létrehoztak egy Social Web Working Group nevű munkacsoportot⁵. Az együttműködés célja egy moduláris, szabványos API-kon keresztül kommunikáló platform, illetve az ehhez szükséges protokollok kidolgozása volt.⁶ Bár elsősorban az egyes rendszerek közötti együttműködést és az azonos funkciók (pl. ismerősök felvétele, kommentálás, képek és videók megosztása) leprogramozását könnyíti meg ez az ajánlás, de ki lehetne bővíteni a hosszú távú megőrzéshez hasznos részekkel is. Sajnos a munkacsoport 2018-ban befejezte a munkáját (2. ábra) és azóta nem igazán hallani arról, hogy a nagy közösségi média cégek a nyílt szabványok irányába fejlesztenék a rendszereiket.

3 OpenSocial, Wikipedia. Hozzáférés: 2021.05.10., <https://en.wikipedia.org/wiki/OpenSocial>

4 OpenSocial Foundation moves standards work to W3C social web activity, W3C Blog. Hozzáférés: 2021.05.10., <https://www.w3.org/blog/2014/12/opensocial-foundation-moves-standards-work-to-w3c-social-web-activity>

5 Socialwg, W3C Wiki. Hozzáférés: 2021.05.10., <https://www.w3.org/Social/WG>

6 Social Web Protocols – W3C Working Group Note 25 December 2017, World Wide Web Consortium. Hozzáférés: 2021.05.10., <https://www.w3.org/TR/2017/NOTE-social-web-protocols-20171225/>

Log in

Page Discussion Read View source View history Search

Socialwg

Welcome to the **Social Web Working Group** (SocialWG) home page (<https://www.w3.org/SocialWG/>)!

The Social Web Working Group was active 2014-07-21 to 2018-02-13.

Please see the [Social Web Incubator Community Group](#) for current and future activity.

Social Web Working Group

- [Announcement 2014-07-21](#)
- [Charter](#) (closed on 2018-02-13)

Chairs for the duration

[Tanlek Çelik](#) (Mozilla)
[Evan Prodromou](#) (e14N)

Staff contacts during and at close

[Amy Guy](#) (W3C/University of Edinburgh)
[Sandro Hawke](#) (W3C/MIT)

Specifications

Specifications that the Social Web WG has produced

Some use cases have been addressed by multiple specifications that provide alternative approaches, so trying to figure out what's what can be a bit confusing. [Social Web Protocols](#), an overview of these protocols, should help with this.

- [Document Status](#)

Recommendations

- [Webmention](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Linked Data Notifications](#) ([Sarven Capadmal](#), [Amy Guy](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Micropub](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Activity Streams Core](#) ([James Snell](#) and [Evan Prodromou](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Activity Streams Vocab](#) ([James Snell](#) and [Evan Prodromou](#)) (Editor's Draft, Issues, Repo)
- [WebSub](#) ([Julien Genestoux](#), [Aaron Parecki](#)) (Issues, Repo) - formerly PubSubHubbub
- [ActivityPub](#) ([Chris Webber](#), [Jessica Tallon](#)) (Editor's Draft, Issues, Repo)

NOTES

- [JF2](#) ([Ben Roberts](#)) (Editor's Draft, Issues, Repo)
- [Post Type Discovery](#) ([Tanlek Çelik](#)) (Editor's Draft, Issues, Repo)
- [Social Web Protocols](#) ([Amy Guy](#)) (Editor's Draft, Issues, **Contains: Guide to all these drafts**)
- [IndieAuth](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo)

Specification Related

- List of github aliases of Social Web WG contributors: [Socialwg/Github](#)
- [Resolutions](#) - e.g. to accept & publish drafts, for controversial issues, etc.

Discussion

The Working Group welcomes participation from everyone. We currently use IRC, email, telcons, and face-to-face meetings for discussion. You have to be a WG member to attend telcons and F2F meetings, but anyone is free to observe these on IRC!

Extensions to Working Group specs are done in the [Social Web Incubator Community Group](#) (SocialCG), which is open for anyone to join as a member.

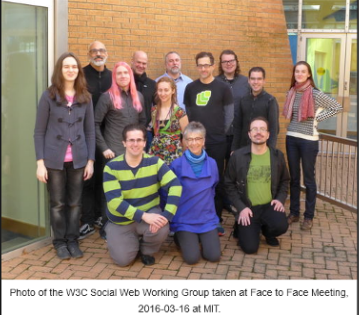


Photo of the W3C Social Web Working Group taken at Face to Face Meeting, 2016-03-16 at MIT.

Contents [hide]

- 1 Social Web Working Group
- 2 Specifications
 - 2.1 Recommendations
 - 2.2 NOTES
 - 2.3 Specification Related
- 3 Discussion
 - 3.1 IRC
 - 3.2 Email
 - 3.2.1 Past Meetings
- 4 Documentation
 - 4.1 Naming
 - 4.2 Historical
 - 4.3 Use Cases
 - 4.4 Social Data Syntax
 - 4.5 Social Vocabulary
 - 4.6 Social API
 - 4.6.1 API research
 - 4.6.2 API candidates
 - 4.7 Federation
 - 4.7.1 Federation candidates
 - 4.8 Inputs into SocialWG
 - 4.8.1 Activity Streams
 - 4.8.2 IndieWeb
 - 4.9 Dependent Work

2. ábra. A 2014 és 2018 között működő Social Web Working Group szócikke a W3C wikijében.

3. BESOCIAL

2020-ban a Belga Királyi Könyvtár három ottani egyetemmel együttműködve BESOCIAL néven indított egy két éves projektet⁷ a közösségi média archiválhatóságával kapcsolatban. A céljaikat az International Internet Preservation Consortium (IIPC) által szervezett Research Speaker Series című webináriumon ismertették.⁸ Megpróbálnak egy fenntartható stratégiát kidolgozni ennek a komplex és gyorsan változó digitális tartalomnak a válogatására, begyűjtésére, megőrzésére és hozzáférhetővé tételére. Első lépésben a Belgiumban és más országokban már létező közösségi média archívumokat, az általuk használt technikai megoldásokat és a jogi környezetet mérték fel. Ennek a fázisnak a kutatási jelentése jelen cikk írásakor még nem volt elérhető, de várhatóan rövidesen megjelenik a könyvtár honlapján. Ezt azután hat további szakasz követi majd, köztük egy pilot projekt, amely során a tartalom arathatóságát, a lementett anyag minőségét és hosszú távú megőrizhetőségét fogják megvizsgálni.

⁷ BESOCIAL, Koninklijke Bibliotheek van België. Hozzáférés: 2021.05.10., <https://www.kbr.be/en/projects/besocial/>

⁸ BESOCIAL research project @KBR: towards a sustainable social media archiving strategy for Belgium, YouTube. Hozzáférés: 2021.05.10., <https://www.youtube.com/watch?v=bX7A5pDoMmQ>



4. Külföldi példák

Európában több nemzeti könyvtár és levéltár is foglalkozik webkettes tartalmak archiválásával, vagy a közeli jövőben tervezi ezt. Többnyire a Facebook, a Twitter, a YouTube, a Flickr és az Instagram oldalairól válogatnak fiókok és/vagy témák szerint a közszereplők (pl. politikusok, újságírók, sportolók, más hírességek) és a szervezetek (pl. hírügynökségek, médiacégek, pártok, kormányzati szervek, kulturális intézmények) által közzétett tartalmakból. Ebből a szempontból kiemelkedő a dán és a brit webarchívum, utóbbinak kiváló nyilvános felülete is van.⁹

Érdekes még megemlíteni az amerikai Library of Congress esetét a Twitterrel, amely a közösségi média megőrzésének egy további nehézségére világít rá. A Kongresszusi Könyvtár 2010-ben jelentette be, hogy 2006-ig visszamenőleg megkapja a Twitter teljes anyagát és előremenetben pedig minden publikus tweetet. 2017 végén viszont már egy olyan közleményt adtak ki¹⁰, mely szerint a jövőben már csak válogatni fognak a Twitter csatornák közül, mert a teljes üzenetfolyam mérete olyan drámaian nő, hogy annak kezelése meghaladja a könyvtár lehetőségeit.

5. Módszerek és eszközök

Mint a fentiekből kiderült, a webkettes tartalmak archiválása komoly technikai kihívás. Szerencsére egyre több olyan szolgáltatás, illetve szoftver jelenik meg, amelyek erre a célra (is) alkalmasak. Itt most csak felsorolásszerűen teszünk ezekről említést, de többségükről külön szócikk van az OSZK Webarchívum honlapjáról¹¹ elérhető wikiben.¹² A csillaggal jelletteket mi is használjuk a nemzeti könyvtárban.

Fizetős platformok, szolgáltatások: MirrorWeb, ArchiveSocial, PageFreezer, Smarsh, Intradyn, Jatheon, Archive-It

API-n keresztül archiváló szoftverek: Social Feed Manager, Twarc, F(b)arc, TAGS, Munin Fájletöltők: youtube-dl*, DownThemAll!, FastVid, Story saver, Social Downloader

Böngészőn keresztül archiváló eszközök: Brozler*, Browsertrix, WAIL*, Webrecorder Desktop*, ArchiveWeb.page*, Conifer*, Instamancer, Crocoite

Megjelenítők: ReplayWeb.Page*, WebrecorderPlayer*, Conifer*, PyWb*, OpenWayback*

6. Saját tapasztalatok

Az Országos Széchényi Könyvtárban 2017-ben indult webarchiválási projekt első éveiben a hagyományos web aratására koncentráltunk és bár már 2019 elején megpróbálkoztunk az OSZK saját online szolgáltatásai, köztük a webkettes oldalai

⁹ UK Government Social Media Archive, The National Archives. Hozzáférés: 2021.05.10., <https://webarchive.nationalarchives.gov.uk/social/search/>

¹⁰ Update on the Twitter Archive at the Library of Congress, Library of Congress Blog. Hozzáférés: 2021.05.10., <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/>

¹¹ OSZK Webarchívum honlapja. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu>

¹² MIA Wiki, OSZK Webarchívum. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu/mediawiki/>

mentésével (3. ábra), komolyabban csak tavaly kezdtünk el kísérletezni a közösségi média archiválásával. Február végétől május közepéig kb. 700 publikus magyar Instagram fiókból minden képet és videót letöltöttünk (összesen több mint százezer bejegyzést), majd ősszel frissítettük is ezt a részgyűjteményt az időközben közzétett anyagokkal. Ezután 147 könyvtári Facebook oldalról mentettünk közel 30 ezer bejegyzést, többnyire képpel, illetve néhol videóval együtt. 35 esetben sikerült a teljes oldalon végigmenni (a legrégebbi letöltött poszt dátuma 2010.04.28. volt), a többinél pedig az utolsó egy-két év bejegyzéseiből mentettünk le átlagosan körülbelül kétszázat. Mindkét platform esetében a Webrecorder Desktop programot és annak online változatát, a Conifert használtuk. 2021 tavaszán pedig az ArchiveWeb.page Chrome böngészőbe integrálható modulját teszteltük a Twitteren. Az első alkalommal 89 könyvtár, múzeum és galéria csatornáját töltöttük le 600 MB összméretben. 75 esetben sikerült egészen az első bejegyzésig visszagörgetni, a legrégebbi elmentett tweet 2007 decemberi volt.

Webkettes oldalak

A webhely neve	URL címe	Archiváló szoftver	Megjelenítő szoftver					Oldalkép	IA mentés	Eredeti
			OpenWayback	PjWb	SoiWayback	Conifer	Webszerver			
DKA – Facebook oldal (mobil felület)	m.facebook.com/digitaliskeparchivum	Webrecorder	🔴	🔴	🔴	🔴	—	🔵	🟪	🟠
MEK – Facebook oldal	www.facebook.com/Magyar-Elektronikus-Könyvtár-338675729150	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – Facebook oldal	www.facebook.com/nemzetikonyvtar	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – Google+ oldal	plus.google.com/u/0/114309059631430797547/	Webrecorder	🔴	🟡	🔴	🔴	—	🔵	—	—
OSZK – Indafotó oldal	indafoto.hu/nemzetikonyvtar/	Webrecorder, HTTrack	🟡	🟡	🟡	🟡	🟡	🔵	🟪	🟠
OSZK – Instagram oldal	www.instagram.com/nemzetikonyvtar/	Webrecorder	🔴	🟢	🔴	🟢	—	🔵	🟪	🟠
OSZK – Pinterest oldal	hu.pinterest.com/nemzetikonyvtar/	Webrecorder	🔴	🟡	🔴	🟡	—	🔵	🟪	🟠
OSZK – Tumblr mikroblog	nemzetikonyvtar.tumblr.com	Webrecorder, HTTrack	🟢	🟢	🟢	🟢	🟡	🔵	🟪	🟠
OSZK – Twitter mikroblog	twitter.com/nemzetikonyvtar/	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – YouTube csatorna	www.youtube.com/user/nemzetikonyvtar/	Webrecorder, Conifer	🔴	🟡	🔴	🟡	—	🔵	🟪	🟠

3. ábra. Az OSZK saját webkettes oldalainak mentései különféle megjelenítőkkal. A piros nyilak a teljesen hibásak, a sárgák a kissé hibásak vagy hiányosak, a zöldek pedig a jók.¹³

Az eddigi tapasztalataink nem túl biztatóak. A webkettes tartalmak üzemserű és tömeges archiválásának egyik nagy akadálya a magas élőmunka igény. Mivel a munkafolyamatok zöme nem automatizálható, ezért valakinek ott kell ülnie a számítógép előtt, bejelentkezni az adott platformra, elindítani az archiváló szoftvert, megnyitni a letöltendő URL címet, végiggörgetni és végigkattintgatni az oldalt, lementeni az archivált fájlt, majd feltölteni a szerverre, végül pedig rögzíteni a fontosabb metaadatokat. Ezen a hosszadalmas folyamaton az archiváló szoftverekbe beépített „robotpilóta” funkciók, illetve az általunk készített billentyű makrók sem gyorsítanak érdemben, ráadásul ezeknél gyakoribb a hibázás vagy a lefagyás esélye, ami miatt újra kell kezdeni az egészet. A problémák másik része az archiváló, az indexelő és a megjelenítő eszközök hiányosságai és inkompatibilitásával magyarázható. Rengeteg ilyenbe futottunk már bele, a magyar ékezeteket tartalmazó URL címeiktől kezdve, a Facebook-ról készített WARC fájlok indexelhetetlenségén át, a lementett, de lejátszhatatlan videókig. Sajnos azt kell mondanunk, hogy a jelenlegi létszámmal és technológiával a közösségi média

¹³ OSZK-s webhelyek archívuma. Hozzáférés: 2021.03.20., <https://webarchivum.oszk.hu/oszk-s-archivum-kezdolap>



archiválásának mostani metodikája nem hatékony, nem robosztus, nem skálázható és nem fenntartható.

7. Útkeresés

A témában az utóbbi években megsaporodott szakcikkek, az említett BESOCIAL felmérés előzetes eredményei, az IIPC konferenciáin és webináriumain elhangzó előadások azt mutatják, hogy más nemzeti könyvtárak és egyéb archiváló szervezetek is küzdenek ezzel a feladattal. Van, ahol még csak a tervezés vagy a kísérletezgetés szintjén tartanak, de akad köztük néhány olyan is, ahol már évtizedes tapasztalat gyűlt össze és van egy bejáratott rendszerük. Utóbbiak esetében az a jellemző, hogy nagyon megsűrik az archivált tartalmat, egy-két platformra koncentrálnak (leggyakrabban a Twitterre), és azokon belül is csak kiemelten fontos fiókokat vagy hashtag-eket mentenek. Az eredeti kezelőfelület megőrizhetőségének problémáját pedig néhol úgy kerülik meg, hogy csak a tényleges tartalmat töltik le API-n keresztül vagy valamilyen célszoftverrel, majd egy saját fejlesztésű megjelenítővel teszik böngészhetővé.

A honlapunk *Szakembereknek*¹⁴ menüpontja alatt elérhető forrásokban további információkat találhatnak az érdeklődők erről a speciális szakterületről is.

Bibliográfia

- Drótos László. „Az OSZK webarchívumának 2020-as újdonságai.” *Könyvtári Figyelő* 67, 1. sz. (2021), 31–38. Hozzáférés: 2021.05.10. <http://ojs.elte.hu/kf/article/view/2296>
- Drótos László, Németh Márton. „Egyedi mentésekre szolgáló webarchiváló szoftverek.” *Könyv, Könyvtár, Könyvtáros* 29, 12. sz. (2020), 3–11. Hozzáférés: 2021.05.10. <http://ki2.oszk.hu/3k/2020/12/egyedi-mentesekre-szolgalo-webarchivalo-szoftverek/>

¹⁴ OSZK Webarchívum honlapja – Szakembereknek menüpont. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu/szakembereknek/>