

## Engedélykérő levél archivált webhely nyilvános szolgáltatásához

Tisztelt Honlaptulajdonos!

2017 tavaszától az Országos Széchényi Könyvtár (OSZK) egy kísérleti projektbe kezdett, melynek célja, hogy – más országokhoz hasonlóan – létrehozzon egy nemzeti webarchívumot, amiben megőrizhető a jövő számára legalább egy szelete a digitálisan születő kultúránknak. A webarchiválási projektről a <http://webarchivum.oszk.hu> oldalon lehet tájékozódni és ugyanitt érhető el a nyilvános webarchívum (<https://webarchivum.oszk.hu/demo-kezdolap/>), valamint egy wiki is a témával kapcsolatos ismeretekről. Az archiválás a teljes .hu doménre, illetve válogatott fontosabb webhelyekre terjed ki. A 2020 májusában elfogadott kulturális törvény módosítása a nemzeti könyvtár feladatának határozta meg a webarchiválást. A szerzői jogi törvénynek megfelelően az archiválásra kiválasztott honlapok tulajdonosaitól egyedi engedélyeket kérünk arra, hogy a lementett tartalmat az OSZK nyilvánosan is szolgáltatthassa.

Az előzetes tesztek alapján az alábbi címe(ke)n található webhelye(ke)t is beválogattuk az értékes tartalmú és a jelenlegi technológiákkal jól archiválhatóak közé:

(...)

Jelen levelünkkel arra szeretnénk engedélyt kérni, hogy egy ún. crawler (arató) szoftver segítségével – a Google robotjához hasonlóan –, vagy egyéb archiváló eszközzel néhány havonta letöltsük e webhely(ek)nek nyilvánosan elérhető tartalmát, és ezen másolatokat az OSZK szerveréről szolgáltatthassuk egy nyilvános webarchívum formájában.

A nyilvános szolgáltatáshoz kétoldalú szerződésben szeretnénk rögzíteni az archiválás és szolgáltatás feltételeit, mivel a webhelyen található tartalom a szerzői jogi törvény hatálya alá esik, és csak ily módon lehet az azt érintő jogokat átruházni.

Amennyiben a webhely több aldoménre is kiterjed, akkor a szerződésben szereplő engedély ezekre az URL címekre is vonatkozik. De ha esetleg olyan aldomén is tartozik a webhelyhez, amellyel kapcsolatban jogilag nem illetékesek, akkor kérjük, hogy ennek az URL címét a válaszlevélben (tehát nem a szerződésben) küldjék el nekünk. Ha vannak olyan további webhelyek, amelyek archivált verzióinak nyilvános szolgáltatására szintén szeretnénk engedélyt adni, ezek URL címét a szerződésbe írják bele. Munkánkat nagymértékben segítené, ha a nyilvános webhelyeik által használt domén és aldomén címeket, valamint azok alternatív (alias) URL címét is elküldenék nekünk. Szeretnénk azt is kérni, hogy ha új webhelyet (pl. honlap, blog, virtuális kiállítás) hoznak létre a jövőben új domén vagy aldomén alatt, akkor arról értesítsenek bennünket a [mia@mek.oszk.hu](mailto:mia@mek.oszk.hu) címen.

Három mellékletet csatolunk levelünkhöz, megkönnyítve partnereink számára az ügykezelést: a kinyomtatott PDF-formátumot a kézi kitöltéshez, a Microsoft Office (.docx) vagy Open Office (.odt) szövegszerkesztőkben használható változatokat az elektronikus kitöltéshez lehet használni, értelemszerűen csak egyfelét. (A dokumentumok letölthetők a <https://webarchivum.oszk.hu/szolgaltatasi-engedely-urlop/> címről is.) A kitöltött, aláírt és lepecsételt szerződéseket 2 példányban kérjük postázni az Webarchiválási Osztály részére, az Országos Széchényi Könyvtár címére: 1014 Budapest, Szent György tér 4-5-6., levelezési cím: 1827 Budapest, Budavári Palota „F” épület.

A nyilvános szolgáltatáshoz történő hozzájárulás mellett megköszönnénk, ha a weboldalon valahol feltüntetnék, hogy a nemzeti könyvtár archiválja azt. Ehhez készítettünk egy „pecsétet”, amely több változatban is letölthető az alábbi címről:

<https://webarchivum.oszk.hu/archivalast-jelzo-pecsetek/>

Az archivált weboldalakat az Internet Archive „Wayback Machine” (<http://web.archive.org>) szolgáltatásához hasonlóan lehet majd böngészni ebben a rendszerben, és természetesen jól láthatóan feltüntetjük, hogy ez csak egy másolat és hogy az eredeti milyen címen található. A webarchívumból a Google robotját kizárjuk, így nem fordulhat elő, hogy valaki a Google-ban keresve az eredeti honlap helyett az archivált változatra téved.

Az archiváláshoz többféle szoftvert, köztük a legnépszerűbb crawlert, a Heritrix szoftver 3.3-as verzióját használjuk, amely így azonosítja magát a webszerverek felé: Mozilla/5.0 (compatible; OSZKbot/3.3 +<http://mekosztaly.oszk.hu/mia/> on behalf of the National Szechenyi Library)

A webhelyek tulajdonosai a szerver gyökérkönyvtárában elhelyezett robots.txt fájl segítségével tudják szabályozni, hogy mely robotoknak mely alkönyvtárak bejárását engedik vagy tiltják meg, így egyben azt is megszabhatják az OSZKbot azonosítójú (*user-agent*) robot számára, hogy a webhely mely részeit archiválhatja és melyeket nem. A robots.txt felépítéséről bővebben itt lehet olvasni: <https://support.google.com/webmasters/answer/6062596?hl=hu>. Az OSZKbot amellet, hogy tiszteletben tartja a robots.txt fájlban található tiltásokat, ún. *polite* üzemmódban működik, ami azt jelenti, hogy nem terheli túl az eredeti webszervert gyorsan ismétlődő kérésekkel, nem lassítja annak válaszidejét az emberi felhasználók számára. Emiatt egy nagyobb webhely legfelső néhány szintjének letöltése akár napokig is eltart.

Bár a Heritrix a jelenlegi legfejlettebb webarchiváló eszköz, a dinamikus, sok Javascriptet vagy egyéb kódot tartalmazó, illetve a felhasználó közreműködését igénylő weboldalakkal nem mindig tud megbirkózni, ezért előfordulhat, hogy a mentett verzió hiányos lesz: pl. nem jelenik meg az oldal minden eleme, elromlik a külalak, nem követhetők mindenhol a belső linkek, nem működik a belső kereső vagy egyéb beépített program. A nyilvános archívumnak ez is az egyik célja, hogy maguk a honlaptulajdonosok is megnézhessék, hogy a weboldalaik mennyire őrizhetők meg hosszú távon. Szívesen segítünk tanácsokkal és tesztelésekkel abban, hogy milyen változtatásokkal lehetne „archiválhatóbbá” tenni az adott honlapot, ami az akadálymentesítéshez hasonlóan fontos szempont. Illetve abban is tudunk segíteni, ha az intézmény saját webarchívumot szeretne kialakítani (pl. a honlapja korábbi verzióinak és egyéb online tartalmainak megőrzése céljából).

Mielőbbi válaszáat előre is köszönjük!