

Feladatok a KDS-K-projektben a webarchiválás támogatására – kritériumok, követelmények

Készítette: Visky Ákos László

Utolsó módosítás: 2020-04-03

I. Általános bevezető

Az elsődleges feladata a projektben résztvevő könyvtáraknak, hogy – előre megadott tematika szerint – a helyi vagy regionális vonatkozású webhelyek címeit és bizonyos adatait összegyűjtsék a nemzeti könyvtár számára, kiegészítendő az OSZK webarchívumában már összegyűjtött állományt; valamint a gyűjtőkör szempontjából kiemelkedően fontos webhelyek kiválogatása és szolgáltatásra ajánlása a nyilvános webarchívum számára. (Távlatilag ez a gyűjtés alapja lehet egy helyi vonatkozású webarchívumnak is.) További lehetséges feladatok a rendelkezésre álló idő függvényében: az archivált webtartalom minőségellenőrzése és metaadatolása (ezek részleteire most nem tér ki ez a leírás).

Érdekes kísérlet lenne mozgosítani, bevonni az olvasókat a régió webhelyeinek feltárásába, összegyűjtésébe, egyfajta “crowdsourcing” formájában. A könyvtárak honlapján el lehetne helyezni egy figyelemfelhívást, miszerint archiválási céllal össze szeretnék gyűjteni a régió webhelyeit, a javaslatokat lehetne akár űrlapon keresztül fogadni, vagy az olvasószolgálatnál gyűjteni. (A figyelemfelhíváshoz biztosítani tudjuk a webarchívum különböző információs anyagait.)

A fentiek mellett szeretnénk felkérni a projektben résztvevő könyvtárakat és a közreműködő kollégákat, hogy – kihasználva a személyes kapcsolatokat, valamint a “tűzközeliséget” – a felhasználók, az olvasóközönség felé népszerűsítsék a webarchiválást, hívják fel a figyelmet annak fontosságára. Továbbá ajánlják a webhely-tulajdonosok figyelmébe a nyilvános webarchívum szolgáltatásait és segítsék a kapcsolatfelvételt velük a szolgáltatási szerződések megkötéséhez. Mivel a nyilvános webarchívum szabadon elérhető, az abban archivált – és az idő múlásával egyre értékesebb – helyi vonatkozású tartalom kiegészítheti a könyvtár saját digitális állományát.

II. Útmutató a Google-táblázat kitöltéséhez

1. Gyűjtőkör

1.1 Területi lehatárolás

Mivel elsősorban megyei könyvtárak vesznek részt a projektben, ezért kézenfekvő, hogy a megyei vonatkozású webhelyeket gyűjtsék össze. Természetesen nemcsak fel lehet venni, hanem kívánatos is beválogatni megyehatáron túli webhelyeket is a tágabb régióból, mivel sajnos nem az összes ilyen szintű könyvtár vesz részt a projektben. Érdekeselek lehetnek határon túli címek is, ha vannak kapcsolataik külföldi magyar intézményekkel, érdemes felhívni ezen partnerek figyelmét is erre a gyűjtésre. (Érdeklődés esetén az OSZK hivatalosan is felvonná velük a kapcsolatot, de előtte segíthet az informális érdeklődés, mert így már tudjuk, kihez kell fordulni.) Ebben a munkában célszerű támaszkodni a helyismereti-helytörténeti gyűjteményekre, és érdemes bevonni a munkába az ezzel foglalkozó kollégákat is.

1.2 Tematikus lehatárolás

Bár elsődlegesen kulturális, tudományos és oktatási tartalmakat gyűjt az OSZK webarchiválás projektje, de évről-évre bővül újabb és újabb témakörökkel, hiszen a nemzeti könyvtár feladata minden magyar vonatkozású anyag gyűjtése.

Létrehoztunk és minden partnerünkkel megosztottunk egy Google-táblázatot, mindegyik könyvtárnak a neve rövidítésével jelzett külön munkalappal, amiben rögzíteni tudják az adatokat. Előre felvettünk bizonyos tematikus címkéket, amilyen témakörökben várjuk az URL címeket: egyéb (más kategóriába nem sorolható), helytörténet-helyismeret, idegenforgalom-vendéglátás, irodalom (elsősorban szépirodalom), könyvkiadás-kereskedelem, kutatás (tudományos intézetek, kutatók), média (nem periodika, nem írott sajtó), művelődés (színház, művelődési ház, stb.), művészet (nem irodalom), oktatás-képzés (köz- és felsőoktatás), kormányzat és önkormányzat, köz- és magángyűjtemény (könyvtár, levéltár, múzeum, galéria), néprajz, periodika (híroldal, magazin, folyóirat, újság, hírlevél stb.), sport, természettudomány (kivéve intézetek és kutatók), történelem, vallás-egyház-hitközösség. Ezek nagyjából lefedik az eddig létrehozott tematikus gyűjteményeinket, de természetesen fel lehet venni a létező kategóriákba be nem sorolható webhelyeket is, erre szolgál az Egyéb témakör.

A periodikák kivételt képeznek, ezeknek a besorolásánál a műfaj a döntő, nem a témakör, tehát minden periodikát ezen címszó alá kell felvenni.

1.3 Műfaji lehatárolás

Műfajtól függetlenül, bármilyen nyilvános webhely címét fel lehet venni a táblázatba. Bár jelenleg a közösségi média (Facebook, Instagram, stb.) oldalainak archiválása csak egyedi mentésekkel oldható meg, tekintettel az egyre nagyobb súlyukra, összegyűjtésük mindenképpen fontos. Ugyanígy fontos olyan jellegű webhelyek nyilvántartása is, amelyek a jelenlegi eszközökkel nem archiválhatók, mint pl. adatbázisok. Tehát mindenféle műfajú webhelyet lehet és kell gyűjteni, legyen az akár általános honlap,

elektronikus periodika, blog, mikroblog, fórum, közösségi oldal, ismerettár/wiki, hirdetési oldal, adatbázis, digitális gyűjtemény vagy virtuális kiállítás. Az adatfelvételnél műfaj szerint nem kell elkülöníteni a különböző típusú webhelyeket, tematikus jellegük szerint kell beírni őket a táblázatba.

1.4 Típus szerinti lehatárolás

A webarchívum gyűjteményében a legkülönbélebb webhely típusok is megtalálhatók, legyen szó akár kormányzati, intézményi, szervezeti, üzleti, magán vagy közösségi tartalomról. Mivel egy gyűjtéssel nem lehet feltárni minden aktuálisan létező webhelyet, és a rendelkezésre álló erőforrások is végesek, érdemes rangsorolni a gyűjtésbe vont tartalmat. Ezért most elsődlegesen közérdekű tartalmakkal rendelkező webhelyeket szeretnénk összegyűjteni, a szigorúan magán jellegűeket nem, illetve az üzleti jellegűeket is csak válogatva (pl. ami illeszkedik valamelyik megadott kategóriába). De itt is érvényes az a tétel, hogy a nem elsődlegesen gyűjtött címeknek is van helye valahol a nyilvántartásban...

Fontos szempont, hogy ismert vagy közéleti embereknek a magán megnyilvánulásai is érdekesek, ezek ugyanolyan kordokumentumok, mint a száz évvel ezelőtti naplóbejegyzések. Ezért a közéleti emberek vagy művészek személyes honlapjainak címét is be kell gyűjteni, az ezeknek helyet adó helyeket (pl. Facebook-oldal) is fel kell venni a gyűjteménybe, de magánember esetében csak akkor, ha közérdekű tartalom is van rajta.

Az adatfelvételnél típus szerint sem kell elkülöníteni a különböző webhelyeket, tematikus jellegük szerint kell beírni őket a táblázatba.

2. Keretek

2.1 Időkeret

A KDS-K-projekt vége 2020 májusáig lett halasztva, eddig szól ez az együttműködés is a kiírt feladat teljesítésére. Ettől függetlenül örülnénk, ha nem szakadna meg a munkakapcsolat és ezt követően is segítenénk egymás munkáját (ez történhetne önkéntes munkával és formalizált kapcsolat keretében is, utóbbi részleteinek kidolgozása folyamatban van).

Szerencsés volna, ha nemcsak a projekt végén, hanem folyamatosan kerülnének feltöltésre az adatok a táblázatba, így lehetőség nyílna pontosításra, egyeztetésre is. Szintén szerencsés lenne, ha jutna idő más feladatok megismerésére és kipróbálására is, mint a minőségellenőrzés vagy a metaadatolás. Mivel nem ismerjük, mennyi emberi erőforrás áll a könyvtárak rendelkezésére, illetve mekkora a térségre jellemző webtartalom, nem tudjuk előre felmérni, hogy mennyi idő szükséges a webhelyek feltérképezéséhez és összegyűjtéséhez, ezért nem tudunk konkrét időpontot megadni a feladat teljesítésére.

Természetesen a projekt lezárulta után is örömmel és köszönettel fogadunk archiválásra javasolt webhelyeket.

2.2 Mennyiség

A projekt keretében begyűjteni kívánt webhelyek számát sem tudjuk meghatározni, nem tudunk előírni egy mennyiséget, amivel teljesülne a pályázati kiírás. Ennek az az oka, hogy nem ismert az adott térség, régió webhelyeinek számszerű gazdagsága, illetve az, hogy ezek közül mi ismert már a nemzeti könyvtár számára (csak olyan címeket kell gyűjteni, amelyek még nem szerepelnek az OSZK gyűjteményében, erről később). Bízunk benne, hogy mindenki lelkiismeretesen, legjobb tudása szerint fogja végezni a munkáját, könyvtáros alaposságával, egy, a valósághoz nem köthető számadat nem segítené a feladat teljesülését.

3. Adatrögzítés

3.1 A táblázat

A megosztott Google-táblázatban kell rögzíteni az összegyűjtött tételeket, minden könyvtárnak a saját – rövidítésével jelzett – munkalapján, a megadott tematikus besorolás alapján. (A *Sablon* munkalap lezárt, csak mintául szolgált, oda adatot bevinni nem lehet.) A táblázatot szerkesztési joggal osztottuk meg, így tovább lehet osztani a projektbe bedolgozó kollégákkal is.

Ötféle adatot lehet rögzíteni a táblázatban, oszlopok szerint elkülönítve. Az oszlopok 1. cellája a rögzítendő adatot nevesíti, a 2. cellában található ennek leírása, a 3. cellájában pedig egy rövid magyarázó szöveg van, segítő a helyes kitöltést.

A Google-táblázat kezelése hasonló a Microsoft Exceléhez, de némileg eltér attól. Mivel most csak adatrögzítésre kerül sor, pár egyszerű alpművelettel könnyen használható és legfőképpen mindenki által elérhető. Amire érdemes odafigyelni: folyamatosan menti önmagát, ami kényelmes használatot jelent, de azzal a veszéllyel jár, hogy egy óvatlanul bevitt adat (pl. valaminek a felülrása) is mentésre kerül. Ilyen esetben a visszalépés nyíllal jutunk vissza az előző állapotba, vagy töröljük és vigyük be újra az adatot. Fontos, hogy ne a korábbi állapotra való visszaállítás lehetőségét használjuk a *Fájl / Verzióelőzmények* menüpont alatt, mert az az egész táblázatra vonatkozik, tehát a másik munkalapon bevitt adatokat is veszélyezteti. Időnként érdemes exportálni a tartalmat biztonsági mentés gyanánt. Az is biztonságot növelő munkamódszer, ha előbb egy saját, külön táblázatban rögzítjük az adatokat, és onnan emeljük át részenként vagy egyszerre őket. Azért célszerű folyamatosan, rendszeresen feltölteni adatokat, hogy lássuk a haladást, illetve jelezni tudjunk esetleges hibákat.

Mivel a tematikus csoportoknál nem készítettünk előre üres helyeket, ezeket az adatbevitelnél kell létrehozni. Fontos, hogy akárcsak az Excelnél, mindig egész sorokat szúrjunk be (vagy töröljük), ne csak cellákat, hogy ne csússzon el a tartalom! Sort kijelölni a sor számára kattintva lehet, több sort pedig a Shift billentyű lenyomása közben újabb kattintással. Ilyenkor annyi új sort lehet beszúrni jobb klikkel, ahány sor ki van jelölve. Az nem baj, ha olyan cellákat is kijelölünk közben, amikben már van tartalom, nem írja felül őket a program.

3.2 Adatmezők

Két kötelezően kitöltendő adatmező van, a webhelyek címe (A-oszlop) és URL-je (B-oszlop), ezek formai követelményeiről később lesz szó.

Ellenben a kapcsolati adatok felvételére szolgáló C-oszlop kitöltése nem kötelező. Ide a webhely tulajdonosának vagy képviselőjének neve, illetve e-mail címe írható be, de csak akkor kell ezt az adatot rögzíteni, ha az adott webhelyen nincs könnyen elérhető és egyértelmű kapcsolati menüpont vagy e-mail cím, viszont a kitöltő rendelkezik a szükséges információkkal.

A megjegyzéshez (D-oszlop) minden fontosnak gondolt dolog (pl. régi honlap, hamarosan megszűnik, folytatását lásd..., vagy kapcsolódás jelzése, stb.) beírható egyrészt, másrészt három további információ rögzítésére szolgál. Az egyik annak jelzése, hogy az adott webhely jellege szerint több tematikus csoportban is szerepelhetne, de mivel csak a hangsúlyosabb vagy jellemzőbb szerint kell besorolni, itt lehet figyelmeztetni a többes érintettségre (pl. egy főleg irodalommal és mellette képzőművészettel foglalkozó blogot az IRODALOM-hoz kell felvenni, és a megjegyzéshez MŰVÉSzt írni). Továbbá lehet jelezni (pl. sárga szín kiemeléssel), ha az adott webhelyet kiemelten fontosnak gondolják, ezért javasolják felvételét a nyilvános webarchívumba. Szintén a megjegyzésben lehet jelezni, ha az adatok OSZK általi átvétele után új tétel felvételére került sor (akár a projekt ideje alatt, akár a későbbiekben).

Az ötödik, E-oszlop annak rögzítésére szolgál, hogy ki vette fel az adott tételt. Monogramok használata esetén érdemes a táblázat a munkalap elején megadni a feloldásukat is.

3.3 Címellenőrzés

Mivel a projekt célja olyan webhelyek összegyűjtése, amik még nem kerültek be az OSZK gyűjteményébe, csak olyan címeket kell felvenni a táblázatba, amelyek nincsenek nálunk nyilvántartva. A webarchívum részgyűjtemény keresőjén keresztül lehet ellenőrizni a felvenni kívánt tételt: <https://webarchivum.oszk.hu/a-reszgyujtemenyekben/>

Lehet névre is keresni, de biztosabb az URL ellenőrzése. Az URL-keresőben nem az egész, a böngésző címsorából kimásolt URL címre érdemes rákeresni, hanem csak a domén nevére vagy egy jellegzetes betűcsoportra, a felsőbb domén (.hu, .com, stb.) és a http/https, www előtagok nélkül. Tehát ha a <http://www.oszk.hu/> címet szeretném ellenőrizni, akkor elég csak az "oszk" karaktersorra keresni. Így nem okoz téves visszajelzést az, ha a webarchívumban más névalak van felvéve a szinonim címek közül (egy webhelynek sokszor van http és https-es változata is, ahogy www-vel és anélkül is be lehet írni a címet, mert működik az átirányítás). Másik előnye a címtörzs használatának, hogy a találatok között megjelenik az összes ismert cím, amiben előfordul az a karaktersor. Tehát nemcsak a [http://www.oszk.hu/-t](http://www.oszk.hu/) fogom megtalálni, hanem az ismert aldoméneit is (ezekről később), pl. a <http://www.ki.oszk.hu/-t>.

3.4 Adatfelvétel

A táblázat kitöltésénél nyelvtan szerinti kisbetűs írásmódot kérünk, neveknél nagy kezdőbetűvel, még akkor is, ha a webhely csupa nagybetűs írásmódot használ a

címbe. Ezeket kimásolás esetén át kell alakítani kisbetűssé, pl. a Word, a Notepad++ vagy a Textpad használatával, és úgy beilleszteni a táblázatba.

Másik szempont, hogy beillesztésnél ne vigyük be a táblázatba az eredeti formázást, mindig formázás nélküli beillesztést végezzünk (ha nemcsak ráállunk a cellára, hanem bele is kattintunk, akkor eleve ez a beillesztés fog történni). Több cellányi tartalom beillesztésénél a Ctrl+Shift+V billentyűkombinációt érdemes használni, mert akkor sem marad meg az eredeti formázás.

A táblázat A-oszlopába kell beírni a felvenni kívánt webhely nevét/címét. Az első probléma ennek megállapítása, ugyanis többféle névalak is előfordulhat egy-egy webhelyen. Pl. egy intézménynél szerepelhet a teljes hivatalos név, egy közkeletű és a rövidített név is... Problémát okozhat a hivatalos névalak megállapítása, ha a fejlécbe, az impresszumban, a kapcsolat menüpontban vagy a böngésző fülön más-más szerepel.

A táblázatba felvenni a teljes hivatalos névalakot célszerű, utána zárójelben következzen a rövidített vagy mozaik név, majd vesszővel elválasztva a település neve, ahol működik, pl. Országos Széchényi Könyvtár (OSZK), Budapest. Kötőjellel írhatjuk utána, ha alcím is van, ennek felvétele nem kötelező.

Ha egy intézményhez vagy egy személyhez több webhely is tartozik, akár önálló címmel, akár aldoménnel, ezeket külön tételként, egymás alá kell felvenni a táblázatba úgy, hogy mindig a teljes névalakból induljunk ki, és utána szerepeljen a megkülönböztető rész. Ebben az esetben a település felvétele nem szükséges, mert már ismert. Pl. Országos Széchényi Könyvtár (OSZK), Könyvtári Intézet. Ha azonos nevet használnak a webhelyeknél és csak műfajilag különböznek, ezeket a névalak mögött kell jelezni, pl. Országos Széchényi Könyvtár (OSZK) Facebook oldala (vagy Instagram fiókja, stb.).

A B-oszlopba az URL-t a böngésző címsorából másoljuk át, így jó eséllyel az alapértelmezett változatot fogjuk felvenni. Fontos, hogy teljes webhely esetében a felsőbb domén utáni fájlnevek vagy alkönyvtárak (pl. /index.html, /start.php, /hun, stb.) nélkül vegyük fel a címeket, mert az aratáshoz nincs rájuk szükség, sőt esetleg problémát okozhatnak. (Ha bizonytalanság van ezzel kapcsolatban, inkább maradjon a címbe, majd átvételnél mi javítjuk.)

Ha egy felvett webhelyhez önálló aldomének is tartoznak, akkor ezeket is fel kell venni külön tételként.

Alkonyvtárra mutató címet csak akkor vegyünk fel, ha valakinek vagy valamely intézménynek a saját weboldala egy gyűjtőoldalon van kialakítva. Ezeket nyilvántartani is érdekes, és esetleg célzottan aratni is. Webhelyek nagyobb részegységeit (pl. digitális gyűjtemény) nem kell kigyűjteni és felvenni, elvileg az archiválás során mentésre kerülnek ezek is.

III. Oktatás

Helyi oktatás

A megbeszéléseken felmerült, hogy tudnánk-e helyben valamilyen képzést, tájékoztatást vagy bemutatót tartani a webarchiválásról. Természetesen örömmel teszünk eleget az ilyen felkéréseknek, hisz nekünk is fontos, hogy terjeszteni tudjuk az ezirányú ismereteket az országban. Igény esetén egyeztessünk, hogy milyen jellegű, terjedelmű előadásra, esetleg gyakorlati bemutatóra (pl. személyes webarchiválás) lenne inkább igény. A lehetőségre érdemes a megyei kollégák figyelmét is felhívni, esetleg összekapcsolni a továbbképzéssel.

Emellett tervbe van véve, hogy a Könyvtári Intézet szervezésében megvalósuló négy napos tanfolyamunk ne csak Budapesten, de nagyobb városokban, régióközpontokban is megtartásra kerüljön (pl. ősszel Szegeden várható egy ilyen kihelyezett tanfolyam).

Tanfolyam

A Könyvtári Intézet szervezésében zajló akkreditált továbbképzésről, tanfolyamról itt lehet tájékozódni:

<http://ki.oszk.hu/tanfolyamok/az-internet-archivalasa-mint-kozgyujtemenyi-feladat>

Az eddigi tanfolyamok ingyenesek voltak, reméljük a továbbiakban is meglesz a szükséges támogatás ehhez. Érdemes figyelni a közzétételeket. Sajnos nem a naptári évre van előre meghirdetve a tanfolyam, mert nem arra szól a támogatás, hanem tanévekre. Várhatóan évente két alkalom lesz megtartva.

IV. Hasznos linkek, elérhetőségek

- E-mail elérhetőség, kapcsolatfelvételre, tájékoztatásra: mia@mek.oszk.hu
- A KDS-hez kapcsolódó hírek, a tájékoztatás fóruma a MIA-L levelezőcsoport, érdemes feliratkozni rá: <http://mekosztaly.oszk.hu/cgi-bin/mailman/listinfo/mia-l>
- A webarchívum fejlesztési fázisban lévő honlapja: <https://webarchivum.oszk.hu/> (Még nem publikus, de már minden lényeges információ elérhető rajta.)
- A részgyűjtemények keresője: <https://webarchivum.oszk.hu/a-reszgyujtemenyekben/>
- Nyilvános demó webarchívum: <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/demo-kezdolap/>
- Szolgáltatási engedély szerződés nyomtatvány: <https://webarchivum.oszk.hu/tartalomgazdaknak/szolgaltatasi-engedely-urlap/>
- A KDS-projekt keretében fejlesztett mintaalkalmazás a Rákóczi emlékévé kapcsán mutatja be a digitális tartalmak és a webarchívum egymást kiegészítő lehetőségeit: <http://rakoczi2019.webarchivum.oszk.hu/>
- A tanfolyam tananyaga megismerhető az alábbi helyen is: <https://webarchivum.oszk.hu/szakembereknek/tanfolyam-es-e-learning/>