

Visky Ákos László

Országos Széchényi Könyvtár

Digitális Bölcsészeti Központ,
Digitális Filológiai és Webarchiválási Osztály

Rákóczi Emlékév Archívum – egy digitális gyűjtemény mintaalkalmazás

Távcső a történelemre

Aranybulla-webarchívum bemutató és webarchiválási konferencia

Vörösmarty Mihály Könyvtár

2023. április 24.

I. A Rákóczi Emlékév Archívum (REA) létrejötte

- A *II. Rákóczi Ferenc-emlékév* alkalmából, 2019-2020-ban készült a *Közgyűjteményi Digitalizálási Stratégia* pályázat keretében, elsősorban az OSZK akkori *E-könyvtári Szolgáltatások Osztály* munkatársainak közreműködésével.
- Elsődleges célja a webarchiválás technológiájának és egy webarchívum más digitális gyűjteményekkel történő lehetséges összekapcsolásának bemutatása volt egy mintaalkalmazás keretében.
- Az *Emlékévv*el, a *Fejedelemmel*, a *Szabadságharccal* és a *Rákóczi-családdal* kapcsolatos tematika.
- A *Webarchívumból* a weboldalak és webhelyek mentései mellett a *Magyar Elektronikus Könyvtárban (MEK)* levő digitális vagy digitalizált könyvek, az *Elektronikus Periodika Archívum és Adatbázisban (EPA)* található cikkek és a *Digitális Képarchívumból (DKA)* származó fotók, illusztrációk és egyéb képi dokumentumok is összegyűjtésre kerültek és böngészhetők ezen a honlapon.

Drótos L. & Visky Á. L. (2020). *Rákóczi-archívum*. *Könyv, Könyvtár, Könyvtáros* 29(3), 35-48.

https://epa.oszk.hu/01300/01367/00326/pdf/EPA01367_3K_2020_03_035-048.pdf

Miért gondoltuk ezt fontosnak?

- A könyvtárak digitális gyűjteményeinek alapját általában digitalizált dokumentumok alkotják, és ezeket egészíthetik ki a már eleve digitálisan született (*born digital*) különféle dokumentum típusok, adatbázisok.
- Az internetes digitális dokumentumok világa azonban ennél a körnél jóval nagyobb és egyre nagyobb súlyú jelen kultúránkban – gondoljunk csak a honlapokra, blogokra, wikikre, fórumokra, közösségi oldalakra, a kép-, hang- és videómegosztó platformokra stb. –, melyeken milliárdos nagyságrendben vannak a könyvtárak érdeklődési és gyűjtőkörébe tartozó tartalmak.
- Ezek archiválása és megőrzése napjainkban olyan *novum* a közgyűjtemények életében, mint volt húsz-harminc éve a digitalizálás.



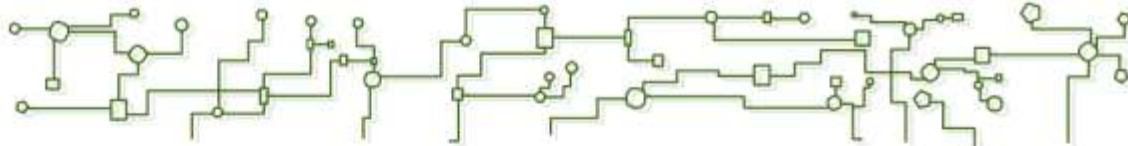
Az emlékév – esemény alapú gyűjtemény

- A *Közgyűjteményi Digitalizálási Stratégia* részeként a megyei és egyetemi könyvtárak számára 2019-ben meghirdetett *KDS-K* pályázatban a digitalizálási célok mellett megjelent a webarchiválás támogatása is.
- Az *OSZK Webarchívuma* négy célt tűzött ki magának, köztük egy mintaalkalmazás készítését szakmai ismeretterjesztéshez és a közoktatásban való felhasználáshoz, amellyel illusztrálható, hogy egy webarchívum anyaga hogyan tudja kiegészíteni a könyvtárak hagyományos digitális gyűjteményeit.
- A mintaalkalmazás apropóját a *II. Rákóczi Ferenc-emlékév* adta, melyet 2019 júliusától, Rákóczi erdélyi fejedelemmé választására emlékezve (1704. július 8.) hirdetett meg az országgyűlés. Az emlékév eseményei egészen 2020. szeptember 17-éig tartottak, mert akkor volt Rákóczi Erdély és Magyarország vezérlő fejedelmévé választásának 315. évfordulója.
- Ha kategorizálni szeretnénk, akkor ez egy *eseményalapú archívum*, de nyilván más mentén is ki lehet alakítani hasonló gyűjteményt – akár személyhez kötve is, amire ma szintén hallunk egy példát.

II. A Rákóczi Emlékév Archívum kialakítása

Gyűjtőkör – gyűjtemények

- A *Rákóczi-webarchívum*ban az emlékévhöz kapcsolódó, folyamatosan keletkező híreket próbáltuk megőrizni a hazai és a határon túli sajtóból, de a gyűjtőkört kiterjesztettük a fejedelemmel, a családjával, a kuruc korról és a szabadságharcra foglalkozó honlapokra, blogokra, hang- és videóanyagokra is.
- Az online újságcikkek mellett az archívumba 2020 őszéig bekerültek intézmények és szervezetek weblapjai, magánemberek blogbejegyzései, *Wikipédia*-szócikkek, képgalériák, elektronikus folyóiratokban megjelent tanulmányok, sulinetes oktatóanyagok, *YouTube*-videók, sőt maga az emlékévről folytatott parlamenti vita jegyzőkönyve is az *Országgyűlés* honlapjáról.
- A webarchívumot kiegészítettük az *OSZK* három további gyűjteményének állományából a Rákóczihoz és korához kötődő dokumentumokkal: a *MEK*-ből könyveket és könyvrészleteket, az *EPA*-ból folyóirat- és újságcikkeket, a *DKA*-ból pedig képeket válogattak össze az ezeket a szolgáltatásokat gondozó munkatársak.
- Ezek között az *OSZK*-ban vagy máshol digitalizált, internetről gyűjtött, illetve szerzők vagy kiadók által beküldött eredeti digitális dokumentumok egyaránt vannak. A *MEK* több tucat könyvvel bővült a *Rákóczi-archívum* miatt, melyek a *Google Books* és az *Internet Archive* szervereiről kerültek begyűjtésre.



Ismertető



Ez a digitális archívum 2019-2020-ban a **II. Rákóczi Ferenc Emlékév** alkalmából készült az Országos Széchényi Könyvtárban, a Közgyűteményi Digitalizálási Stratégia pályázat keretében. Elsődleges célja a webarchiválás technológiájának és a webarchívum más digitális gyűjteményekkel való integrálásának bemutatása egy mintaaikalmazás keretében. Az Emlékévvél, a Fejedelemmel, a Szabadságharcossal és a Rákóczi-családdal kapcsolatos weboldalak és webhelyek mentései mellett a Magyar Elektronikus Könyvtárban levő digitális vagy digitalizált könyvek, az Elektronikus Periodika Archívumban található cikkek és a Digitális Képtárból származó fotók, illusztrációk és egyéb képi dokumentumok is böngészhetők ezen a honlapon. Így egyben a 17. és 18. század magyar törtelmének, társadalom- és hadtörténetének gazdag forrásaként oktatási és kutatási célokra is használható ez az összeállítás.

A böngészés és a keresés a webarchívum mellett a másik három gyűjtemény metaadatai szerint is lehetséges, a találati listákban a színes sávok jelzik, hogy az adott tétel melyik állományból származik, a lakat ikonok pedig azt, hogy a dokumentum nyilvánosan megtekinthető-e. Piros vagy sárga lakat azoknál a weboldalaknál fordul elő, amelyek

archívált példányainak szolgáltatására a tartalomgazda nem, vagy még nem adott engedélyt, de a metaadatok, a kis méretű oldalkép és az Internet Archive mentései ezekben az esetekben is megnézhetők. A teljes szövegű keresés csak a webarchívumra terjed ki és annak is csak a nyilvánosan szolgáltatható részére.

Mivel a web egy komplex médium, ezért a webes tartalmak lementése, hosszú távú megőrzése és megtekinthető állapotban tartása komoly informatikai kihívás. Az OSZK webarchívuma által készített mentések többféle szoftverrel készültek és nézhetők vissza, így összehasonlíthatók az archíváló és a megjelenítő eszközök képességei. Két másik nyilvános részgyűjtemény is elérhető további tanulságos példákkal: az egyik a nemzeti könyvtár saját online szolgáltatásairól készült, a másik pedig egyéb webhelyekből áll.

Az Országos Széchényi Könyvtárban 2017-ben indult webarchiválási projektről a <https://webarchivum.oszk.hu/> honlapon lehet tájékozódni, további információk pedig a webarchivum@oszk.hu levélcímen kérhetők.



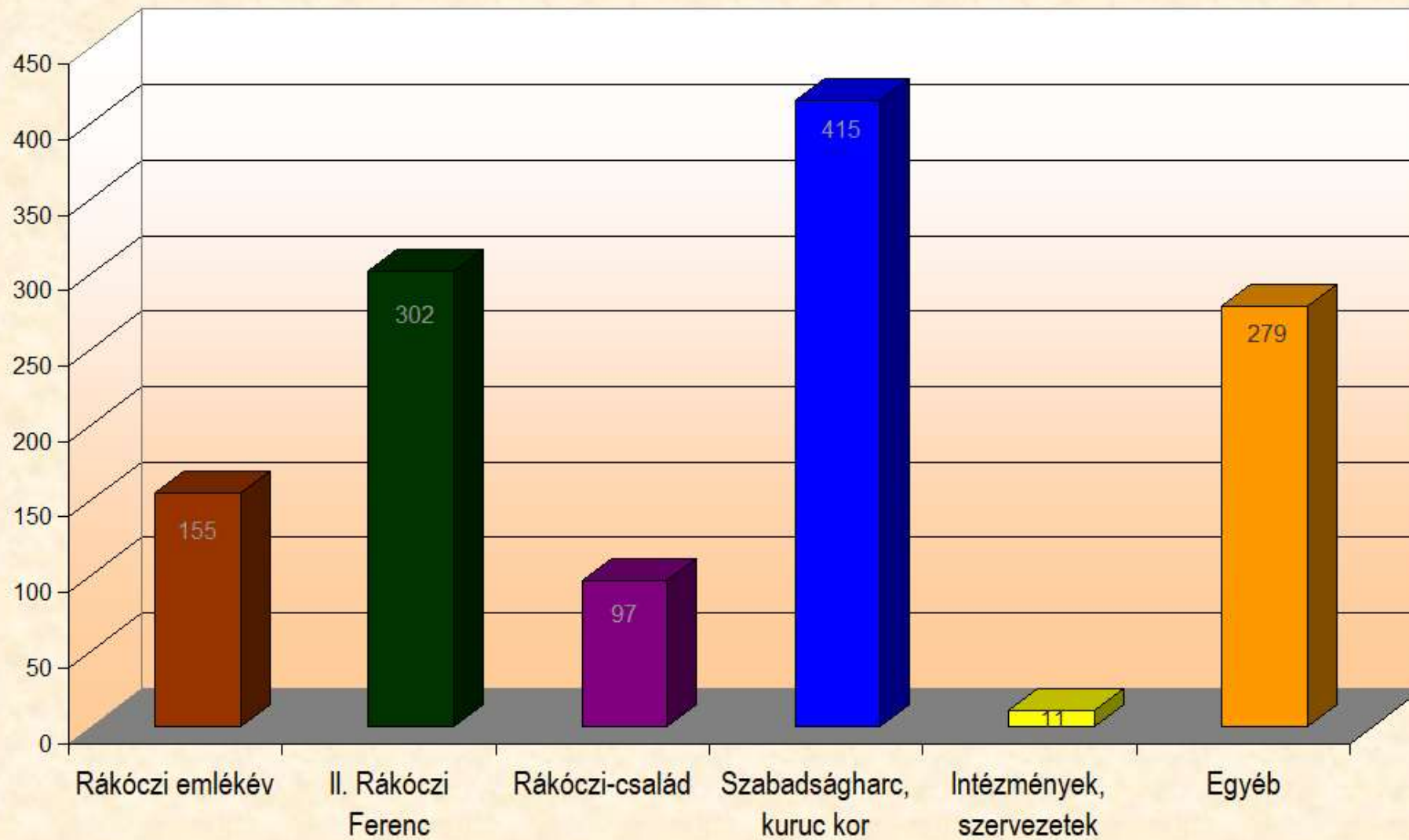
Wikipédia: II. Rákóczi Ferenc emlékévé

Archiválásra javasolom ...

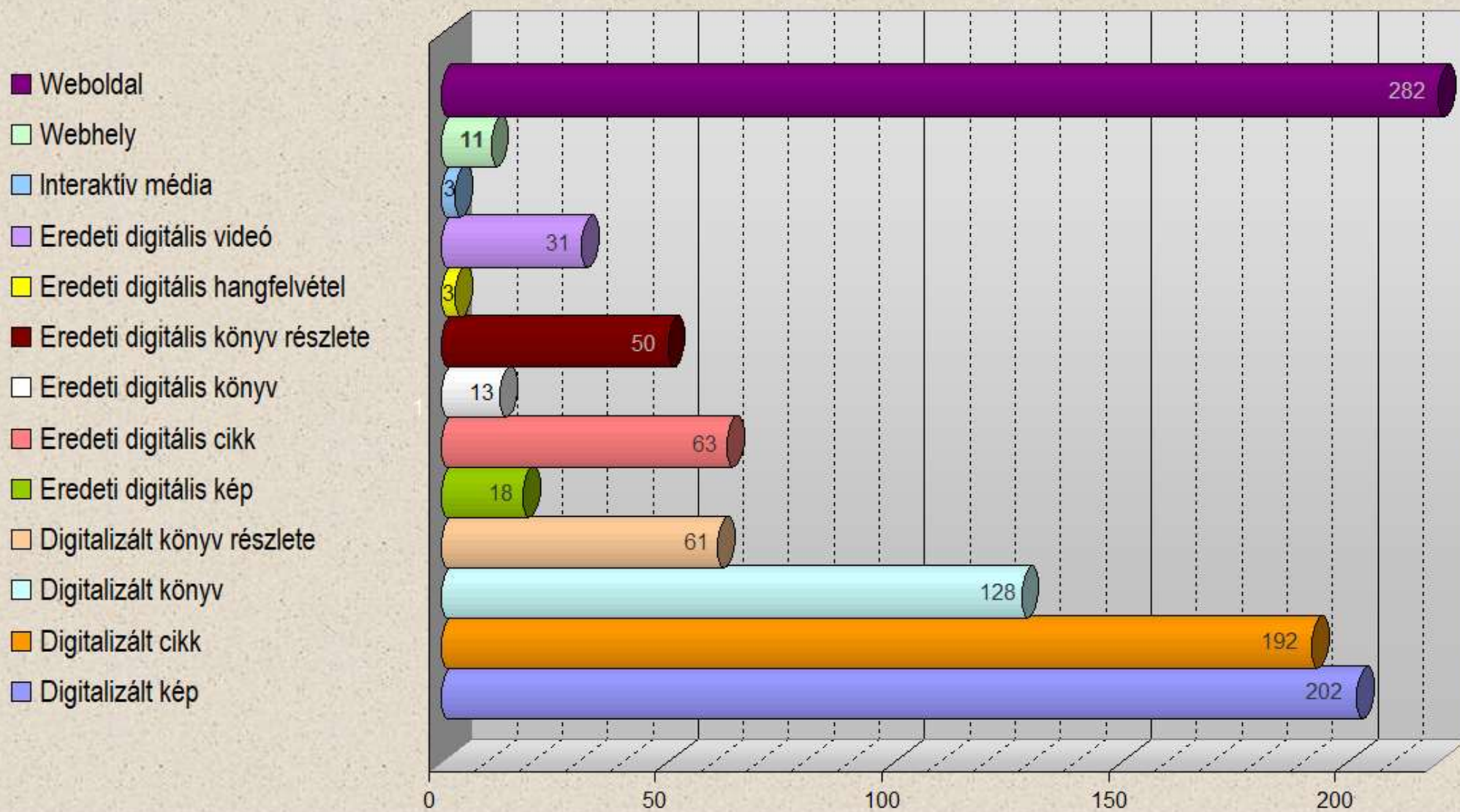


<https://rakoczi2019.webarchivum.oszk.hu/>

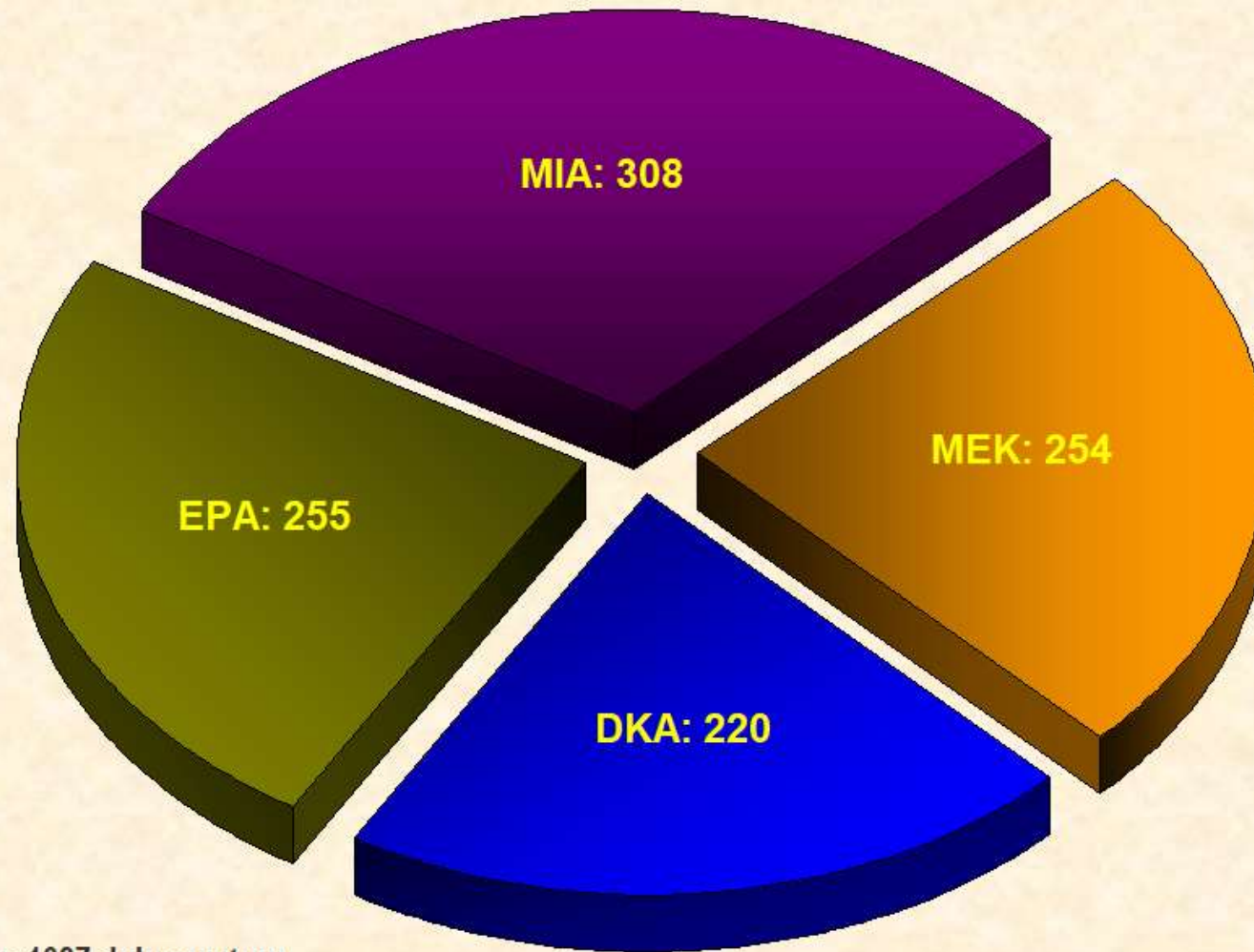
Kategória szerinti megoszlás



Dokumentumtípus szerinti megoszlás



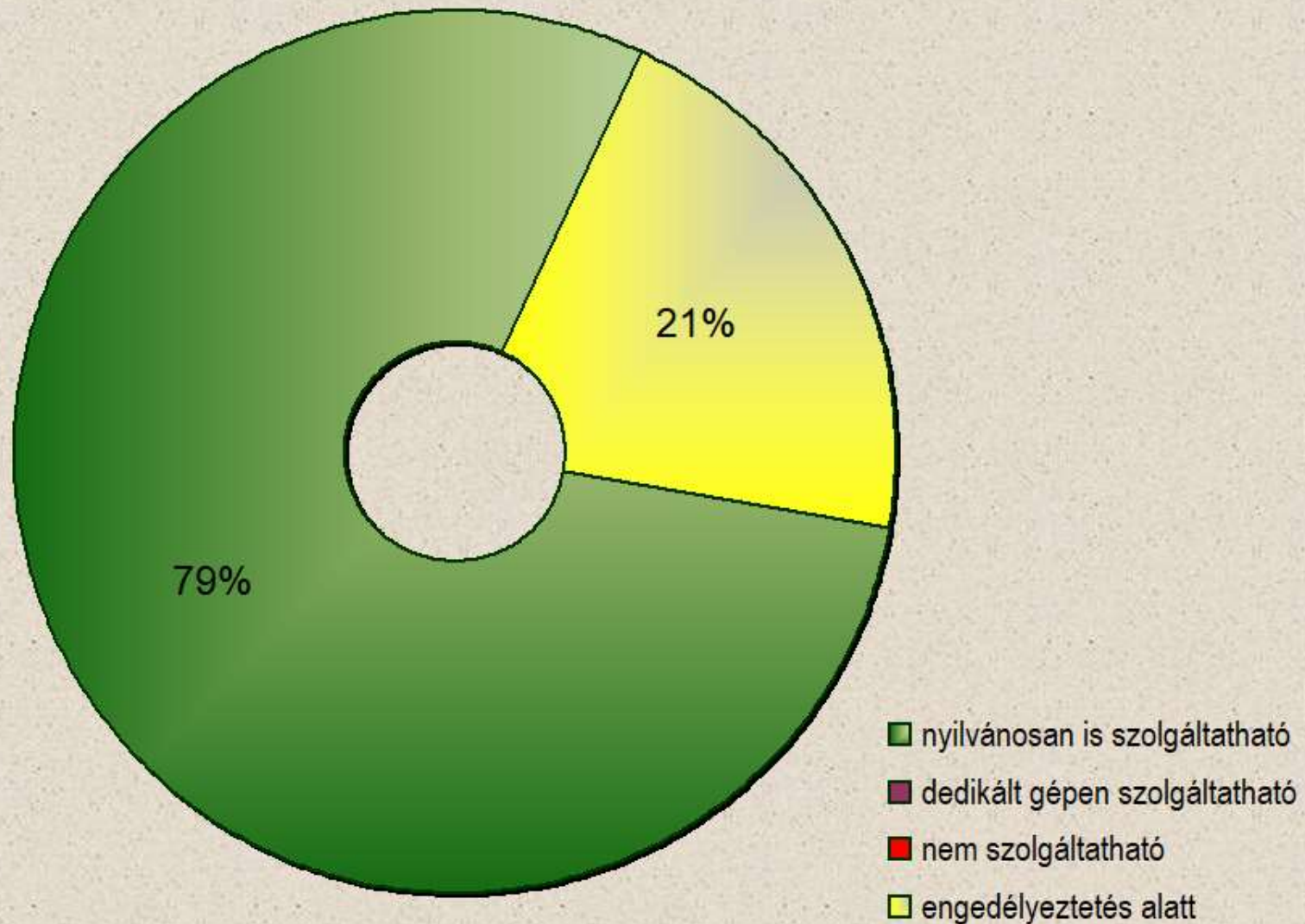
Gyűjtemény szerinti megoszlás



Összesen: 1037 dokumentum

- Össességében közel ezer tételt tartalmaz ez a gyűjtemény, melyek nagyjából azonos arányban származnak a négy digitális szolgáltatásból.

Hozzáférés szerinti megoszlás



Metaadatok 1.

- A webarchívumhoz még 2018-ban kidolgoztunk egy adatstruktúrát, figyelembe véve az amerikai könyvtári szervezet, az *Online Computer Library Center (OCLC)* által létrehozott *Web Archiving Metadata Working Group* az év februári ajánlását.
- Az ajánlás elsősorban a *Dublin Core* adatkészleten alapul, de *MARC21* és *MODS* megfeleltetéseket is tartalmaz. A főként bibliográfiai információk leírására szolgáló mezőket kiegészítettük technikai és adminisztratív adatmezőkkel is, így több mint százféle adatot tudunk rögzíteni a lementett webhelyekről, valamint az azokból kialakított részgyűjteményekről is *XML*-formátumú metaadat rekordokban.
- Probléma: a webarchívumban használt adatszerkezet teljes webhelyekre lett kitalálva, a Rákóczi-gyűjtemény számára viszont nagyrészt egyedi weboldalakat mentettünk le, továbbá a másik három gyűjteményből is át kellett venni a metaadatokat.
- Egyszerűsített leírás kialakítása.

Metaadatok 2.

- Csak olyan bibliográfiai adatokat vettünk fel vagy át, amelyek egyaránt értelmezhetők a könyvek, könyvrészletek, cikkek, képek, videók, weboldalak és egyéb online műfajok esetében: a dokumentum fő- és alcíme, az azt tartalmazó kiadvány összefoglaló címe, az eredeti fájl(ok) származási helye, a szerzők és közreműködők neve, a kiadó és annak honlapja, a dokumentum műfaja és tematikus kategóriája, az *OSZK Köztauruszból* vett tárgyszavak, a földrajzi és személynevek, valamint ezek névtér-azonosítói, és végül a nyelvkód.
- Ezeket az adatmezőket még néhány adminisztratív információval egészítettük ki, például a feldolgozó neve, az archívumba kerülés dátuma, a hozzáférhetőség státusza, az *OSZK*-ban és az *Internet Archive*-ban levő mentett verziók *URL*-je, a címlapképet vagy oldalképet tartalmazó fájl neve. Utóbbiak közül nem mindegyik jelenik meg a honlapon, mert egy részük csak nyilvántartási célokat szolgál.
- A metaadatok melletti kép a *MEK*-es könyvek és könyvrészletek esetében a címlapot ábrázolja, az *EPA*-ból származó cikkeknel és tanulmányoknál a periodika valamelyik számának a borítóját, a *DKA*-ban levő képi dokumentumoknál pedig magának a képnek a kicsinyített verzióját jelenítjük meg. A weboldalakról is készült egy nagyobb (1280 pont széles) és egy kisebb (300 pont széles és maximum 600 pont magas) oldalkép. Azoknál az eseteknél, amelyeknél nem kaptunk engedélyt a nyilvános szolgáltatásra, csak ez a kis bélyegkép tekinthető meg.

Azonosító: R19-200211

Típus: Eredeti digitális könyv

Az OSZK-ban archivált dokumentum:

<http://mek.oszk.hu/20100/20106/>

Az Internet Archive mentései:

https://web.archive.org/web/*http://kmmi.org.ua/konyvtar/konyvek

A dokumentum eredeti forrása:

<http://kmmi.org.ua/konyvtar/konyvek>

Főcím: Rákóczi turulmadara

Alcim/egyéb cím: A Nagyságos Fejedelem és kora a nép emlékezetében: történeti mondák, 1

Alcim/egyéb cím: KMMI-füzetek 24.

Egyéb közreműködő: Dupka György (összeáll.)

Egyéb közreműködő: Zubánics T. Ferenc (összeáll.)

Kategória: Egyéb

Tárgyszó: monda

Tárgyszó: legenda

Tárgyszó: emlékhely

Tárgyszó: kárpátaljai magyar

Tárgyszó: honismeret

Tárgyszó: helytörténet

Személynév: Rákóczi Ferenc

Személynév azonosító: 12072

Földrajzi név: Kárpátalja

Földrajzi név azonosító: 687

Nyelv kód: hun

Archiválás dátuma: 2020-01-

Hozzáférés státusza: nyilván

Hozzáférhetőség kezdő dátum

II. RÁKÓCZI FERENC

Azonosító: R19-300039

Típus: Eredeti digitális kép

Az OSZK-ban archivált dokumentum:

<https://dka.oszk.hu/061800/061841>

Az Internet Archive mentései:

https://web.archive.org/web/*http://www.agt.bme.hu/varga

A dokumentum eredeti forrása:

<http://www.agt.bme.hu/varga/foto/ajanlat.html>

Főcím: II. Rákóczi Ferenc

Alcim/egyéb cím: Tarpa

Tartalmazó kiadvány címe: Varga József fotói, Beregtől S.

Létrehozó személy: Varga József

Egyéb közreműködő: Bíró Lajos

Kiadó neve: Varga József

Kiadó honlapja: <http://www.agt.bme.hu/varga/foto/ajanlat.html>

Kategória: II. Rákóczi Ferenc

Tárgyszó: emberábrázolás

Tárgyszó: lovaskép

Tárgyszó: uralkodó

Személynév: Rákóczi Ferenc (Erdély: fejedelem), II. (1676-1735)

Személynév azonosító: 120726363

Földrajzi név: Tarpa

Földrajzi név azonosító: 714938

Archiválás dátuma: 2014-09-15

Hozzáférés státusza: nyilvánosan is szolgáltatható

Hozzáférhetőség kezdő dátuma: 2014-09-15

Azonosító: R19-400226

Típus: Eredeti digitális cikk

Az OSZK-ban archivált dokumentum:

https://epa.oszk.hu/03400/03468/00031/pdf/EPA03468_kisgrafika_2012_2_006-008.pdf

Az Internet Archive mentései:

http://web.archive.org/web/*http://www.kisgrafika.hu/

A dokumentum eredeti forrása:

<http://www.kisgrafika.hu/>

Főcím: A Rákóczi szabadságharc

Tartalmazó kiadvány címe: F

Létrehozó személy: Tóth Korn

Kiadó neve: Kisgrafika Barát

Kiadó honlapja: <http://www.k>

Kategória: Egyéb

Kategória: Szabadságharc, ku

Tárgyszó: képzőművészet

Tárgyszó: kisgrafika

Tárgyszó: ex libris

Tárgyszó: rézkarc

Személynév: Rákóczi Ferenc (

Személynév azonosító: 12072

Nyelv kód: hun

Archiválás dátuma: 2018-08-

Hozzáférés státusza: nyilván

Hozzáférhetőség kezdő dátum

A RÁKÓCZI-INDULÓ

Azonosító: R19-100212

Típus: Webhely

Az OSZK-ban archivált dokumentum:

http://webarchivum.oszk.hu/pywb/*?url=http://rakoczi-indulo.oszk.hu

Az OSZK-ban archivált dokumentum:

http://webarchivum.oszk.hu/owb/wayback/*http://rakoczi-indulo.oszk.hu

Az Internet Archive mentései:

http://web.archive.org/web/*http://rakoczi-indulo.oszk.hu

A dokumentum eredeti forrása:

<http://rakoczi-indulo.oszk.hu>

Főcím: A Rákóczi-induló

Létrehozó szervezet: Országos Széchényi Könyvtár

Létrehozó szervezeti egység: Webtartalom Koordinációs Osztály

Kiadó neve: Országos Széchényi Könyvtár

Kiadó honlapja: <http://www.oszk.hu/>

Kategória: Egyéb

Tárgyszó: katonainduló

Tárgyszó: kurucnóta

Tárgyszó: zenemű

Tárgyszó: kotta

Tárgyszó: zenetörténet

Tárgyszó: Rákóczi-induló

Nyelv kód: hun

Nyelv kód: eng

Archiválás dátuma: 2019-02-09

Hozzáférés státusza: nyilvánosan is szolgáltatható

Hozzáférhetőség kezdő dátuma: 2019-10-30



Webarchiválás

- Az archiválásra kiválasztott oldalról először az oldalképet készítjük el, mert így lehet a legpontosabban dokumentálni, hogy hogyan néz ki az a jelenleg használatos böngészőben.
- A *Rákóczi-archívumban* sok egyedi weboldal van, amiket az engedélyezettetés miatt egyesével mentettünk le, hogy önálló *WARC*-formátumú fájlalba kerüljenek az egyes tételek.
- Ennél a projektnél nem alkalmazhattuk a tömeges aratásra használt technikát, ahol vegyes tartalmú *WARC*-ok keletkeznek, hanem egy *PC*-n, *Windows* alatt futó szoftverekkel készültek a mentések.
- A legtöbb mentés a parancsmódban használható *linux-os PythonWayback* programmal készült a *record* üzemmódjában (minden oldalt elment, amit megnyitunk a böngészőnkben).
- Ez elég időigényes, viszont általában elég jó minőségben tudja lementeni még a bonyolult hírportálokat és a videók többségét is, amiket szintén el kell indítanunk ahhoz, hogy letöltse őket.
- Amivel a *PyWb* nem boldogult, azt a hasonló elven működő *Webrecorder* nevű eszközzel mentettük le, illetve még a *WAIL*-rendszer is használtuk, amiben kétféle robotfunkció (*Brozzler*, *Heritrix*) is van.
- A *WARC*-fájlok visszanezésére elsősorban a *PyWb* és az *Open Wayback (OWB)* szolgál, de belinkeltük az amerikai *Internet Archive* saját mentéseit is (ha még nem volt mentés az adott weboldalról, akkor a *Save page now* szolgáltatásuk segítségével készítettünk egyet).
- A weboldalak szövegében való keresésre a *SolrWayback* szoftvert használjuk, amivel a részletes technikai metaadatok is megnézhetők minden egyes tételnél, a *Toolbar* panel bekapcsolásával pedig különböző statisztikák és grafikonok is generálhatók.

Archiválási problémák 1.

Mivel egyedi és sokszor nem teljes webhelyekre kiterjedő mentés történt, újszerű problémákkal szembesültünk:

- Az egyedi weboldalak esetében el kell dönteni, meddig terjed egy webes dokumentum?
- Például egy hírportálon megjelent cikkhez hozzátartoznak-e a mellette megjelenő reklámok, az ajánlott további hasonló vagy éppen egész más témájú, de aktuálisan népszerű hírek, a külön megnyitható képgaléria, a beágyazott videók, a más szerverekről belinkelt *PDF*- és egyéb fájlok, az olvasói vélemények?
- A *Wikipédia* esetében külön mentsünk és írjunk le minden egyes szócikket, vagy tekintsük egy könyvtári egységnek az oldalakat és az abban hivatkozott – például az egyes hadjáratokat, csatákat, hadvezéreket részletesen leíró – szócikkekkel együtt? Hozzávegyük a *Wikimédia Commons* tematikus médiagyűjteményét is, amelyből az illusztrációk be vannak ágyazva a lexikon szöveges részébe? És az egyes nyelvi verziókat, melyek néha csak fordítások, máskor viszont teljesen újraírt szócikkek? Archiváljuk-e a szócikkek alakulását dokumentáló laptörténet és vitalap aloldalakat is?

Archiválási problémák 2.

A problémák másik részét az archiváláskor keletkező technikai nehézségek adják.

- Volt olyan weblap, amelynél már a teljes oldalt kitakaró *cookie* figyelmeztetésen vagy *CAPTCHA* teszten sem tudott túljutni az archiváló szoftver.
- Más esetekben az ékezetes fájlnevek vagy a csak *JavaScript* kódok lefuttatásával keletkező URL-címek akadályozták meg a programot a linkek követésében és a fájlok letöltésében.
- A biztonsági problémák miatt az internetről fokozatosan eltűnő *Flash*-formátum már a mentéseknél is gondot okoz: a böngészőn keresztül való archiváláskor külön engedélyezni kell ezeknek a fájloknak a letöltését és futtatását. (És még így is kérdéses, hogy néhány év múlva egyáltalán lesz-e még olyan szoftver, amivel ezekről a *Flash*-alapú weboldalakról készült mentések visszanezhetők lesznek.)
- Viszonylag gyakori az is, hogy a webservereken levő *robots.txt* fájlokban olyan útvonalak vannak – sokszor nem szándékosan – elzárva a robotok elől, amelyekre a *Google* keresőjének nincs szüksége, de archiválási szempontból fontosak: például a külalakot meghatározó *CSS*-fájlokat, vagy a navigációhoz használt programkódokat tartalmazó alkönyvtárak.

Archiválási problémák 3.

A hibák harmadik csoportja a visszanevezéskor lép fel.

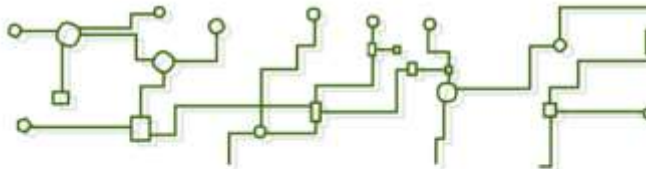
- Hiába van benne a *WARC*-konténerben a weboldalt alkotó összes fájl, ha azok egy részét a megjelenítő szoftver valamiért nem tudja megtalálni vagy megmutatni. Bár sokszor még azt sem könnyű eldönteni, hogy tényleg le lett-e mentve minden szükséges fájl és valóban csak a megjelenítő korlátaiba ütköztünk.
- A *Wikipédiánál* futottunk bele abba a problémába, hogy ott egy úgynevezett *srcset* attribútummal adják meg a szócikkekbe ágyazott képek méretét, ami azt eredményezi, hogy a webszerver a felhasználó képernyőfelbontásához optimalizálva küldi át a képeket. Vagyis, ha böngészőn keresztül mentünk, akkor csak adott méretű képfájlok kerülnek az archívumba, amelyek egy más felbontású monitoron nem jelennek meg.
- További tipikus problémát jelentenek a *JavaScript*-, *Java*- vagy *Flash*-alapú képnézegetők, hang- és videólejátszók. Ezek a megjelenítő felületen sokszor nem indulnak el, még ha maguk a médiafájlok le is lettek töltve.

Engedélyeztetési problémák

- A közgyűjtemények szabad felhasználás keretében, nem haszonszerzés céljából végezhetnek többszörözést a védett művekről is – ez vonatkozik a webarchiválásra is.
- Ez a szabály csak a megőrzésre és a helyben vagy zárt hálózaton történő szolgáltatásra ad lehetőséget, a nyilvános közzétételre nem. Utóbbi a jogtulajdonos engedélyével történhet, de ehhez kétoldalú szerződést kell kötni.
- A *MEK*, *EPA* és *DKA* állományából válogatott dokumentumoknál nem kellett engedélyeztetéssel bajlódni, mert már vagy rendelkeznek ezekkel vagy eleve szabad felhasználásúak – ahogy az ilyen minőségű webhelyek esetében sem kértünk felhasználási engedélyt.
- Szerzői jogi szempontból a webhelyek védett és nem védett alkotásokat is tartalmazhatnak. Jellegüket tekintve gyűjteményes műnek minősülnek, ezért nem kell külön-külön szerződni minden közreműködő alkotóval, de tulajdonjogi szempontból viszont lehet több tulajdonos is.
- Ugyanazt az engedélyeztetési eljárást alkalmaztuk, mint amit korábban a nyilvános webarchívum esetében kialakítottunk, de most nemcsak teljes webhelyekre kértünk szolgáltatási engedélyt, hanem gyakran csak egy-egy weblapra, mivel főleg cikkek és blogbejegyzések kerültek ebbe a gyűjteménybe.
- Speciális problémák webhelyeknél: sokszor nehéz a tulajdonos kiderítése és a kapcsolatfelvétel; a webhely gazdája nem minden tartalom esetében jogtulajdonos (pl. linkelt anyag); ismétlődő engedélykérések (pl. a folyamatos keletkezés miatt).
- Az archivált weboldalak közel fele nem hozzáférhető.

III. A REA felépítése, szolgáltatásai és funkciói

- Önálló, *WordPress*-alapú honlap – amely egyben prototípusként szolgált az *OSZK Webarchívumának* új honlapjához is –, amihez még különböző kereső- és böngészőfunkciókat is fejlesztettünk.
- Hatnyelvű honlap, a magyar mellett készült angol, német, francia, lengyel és szlovák verzió is Rákóczi és a szabadságharc európai jelentősége miatt.
- Az archívum anyaga kilistázható tematikus kategóriák, dokumentumtípusok, illetve gyűjtemények szerint.
- Keresni lehet egyszerre az összes metaadatban, vagy több mező kombinálásával, a találati listák pedig négyféle szempont szerint rendezhetők.
- A teljes szövegű kereső jogi és technikai okok miatt csak a weboldalakra és azok közül is csak a nyilvánosan szolgáltathatókra terjed ki.
- A böngészés és a keresés a webarchívum mellett a másik három gyűjtemény metaadatai szerint is lehetséges.
- A találati listákban a színes sávok jelzik, hogy az adott tétel melyik állományból származik, a lakat ikonok pedig azt, hogy a dokumentum nyilvánosan megtekinthető-e. Piros vagy sárga lakat azoknál a weboldalaknál fordul elő, amelyek archivált példányainak szolgáltatására a tartalomgazda nem, vagy még nem adott engedélyt, de a metaadatok, a kis méretű oldalkép és az *Internet Archive* mentései ezekben az esetekben is megnézhetők.



Ismertető Böngészés ▾ Keresés ▾ Teljes szövegű keresés



kategória

[kategóriák szerint ...](#) →

[dokumentumtípus szerint ...](#) →

[gyűjtemények szerint ...](#) →

Rákóczi emlékév

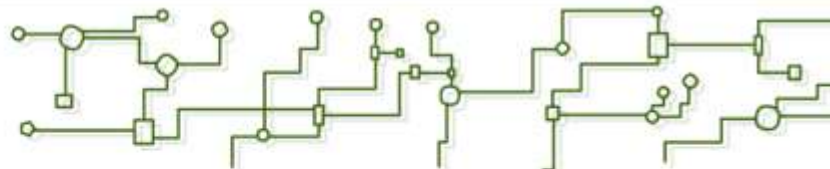
II. Rákóczi Ferenc

Rákóczi-család

Szabadságharc, kuruc kor

Intézmények, szervezetek

Egyéb



Ismertető Böngészés ▾ Keresés ▾ Teljes szövegű keresés [Statisztika](#)



kategória

[kategóriák szerint ...](#) →

[dokumentumtípus szerint ...](#) →

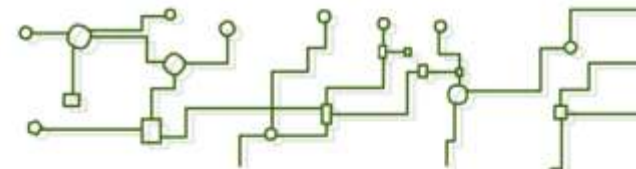
[gyűjtemények szerint ...](#) →

OSZK Webarchívum (MIA)

Magyar Elektronikus Könyvtár (MEK)

Elektronikus Periodika Archivum és Adatbázis (EPA)

Digitális Képtár (DKA)



Ismertető Böngészés ▾ Keresés ▾ Teljes szövegű keresés



kategória

[kategóriák szerint ...](#) →

[dokumentumtípus szerint ...](#) →

[gyűjtemények szerint ...](#) →

Webhely

Weboldal

Interaktív média

Eredeti digitális videó

Eredeti digitális hangfelvétel

Eredeti digitális könyv

Eredeti digitális könyv részlete

Eredeti digitális cikk

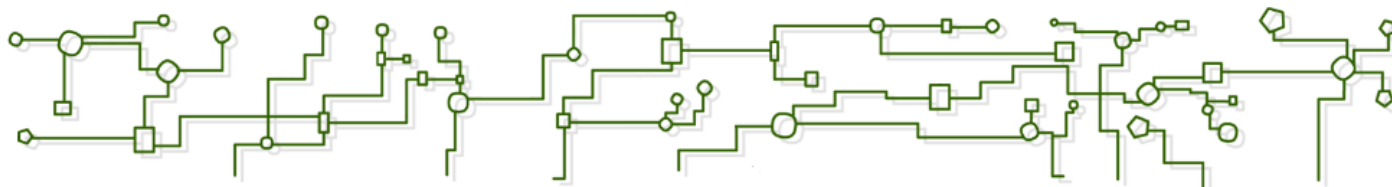
Eredeti digitális kép

Digitalizált könyv

Digitalizált könyv részlete

Digitalizált cikk

Digitalizált kép

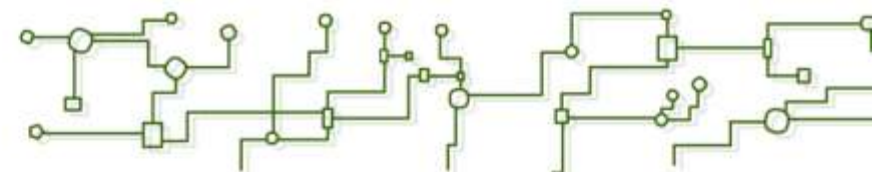


Kategória: II. Rákóczi Ferenc

Főcím ▾ Kiadvány cím ▲ Típus ▲ Kategória ▲

Találatok: 302

- [R19-100227] **Mann Jolán: "Az Isten árvaságra téve bennünket"** (OSZK blog) Kategória: II. Rákóczi Ferenc [Weboldal] 
- [R19-100094] **Kósa Károly: "Dübörgő ércszekeren megjött a nagy Fejedelem..." : II. Rákóczi Ferenc hamvainak hazahozatala** (Kósa Ferenc [Weboldal] 
- [R19-400058] **Tóth Ferenc: "Ex Turcia nulla redemptio". 1717 – II. Rákóczi Ferenc fejedelem török földre lép** (Magyar tudomány) Kategória: *cikk*] 
- [R19-400250] **"Hős Rákóczi népe" : Művelődéstörténeti vetélkedő munkafüzet** (Honismeret) Kategória: Szabadságharc, kuruc kor, II. R
- [R19-400235] **Kincses Katalin Mária (1968): "Minden különő ceremonia nélkül" : A Rákóczi-kultusz és a fejedelem hamvainak haza** Kategória: Egyéb, II. Rákóczi Ferenc [Eredeti digitális cikk] 
- [R19-100025] **Nagy Ádám: 2019 – A Rákóczi-emlékév** (Hajdútánc.hu) Kategória: Rákóczi emlékév, II. Rákóczi Ferenc [Weboldal] 
- [R19-100202] **A Fejedelem : Rock Musical** Kategória: Egyéb, II. Rákóczi Ferenc [Webhely] 
- [R19-300023] **Greguss Imre: A Képzőművészeti Társulat tavaszi kiállításából : 1703. II. Rákóczi Ferencz indulója** (Vasámapi Ujság) Kategória: Szabadságharc, kuruc kor [Digitalizált kép] 
- [R19-100104] **Szabó Ottó: A Nagyságos Fejedelem** (ROVART - A Rovás alkotócsoport netűjságja) Kategória: II. Rákóczi Ferenc, Szabad
- [R19-400027] **Dóczy Jenő: A Rákóczi harang** (Nyugat) Kategória: II. Rákóczi Ferenc [Digitalizált cikk] 
- [R19-200119] **A Rákóczi-emigráció török okmányai, 1717-1803** Kategória: II. Rákóczi Ferenc, Egyéb [Digitalizált könyv] 
- [R19-400046] **Szádeczky Lajos: A Rákóczi-forradalom Erdélyben** (Erdélyi Múzeum) Kategória: II. Rákóczi Ferenc [Digitalizált cikk] 
- [R19-400031] **Éber László: A Rákóczi-freskó** (Művészet) Kategória: II. Rákóczi Ferenc [Digitalizált cikk] 



Részletes kereső

Mezőnév: Főcím ▾

Keresett szavak:

Kapcsolat: és ▾

Mezőnév: Főcím ▾

Keresett szavak:

Kapcsolat: és ▾

- Főcím
- Alcím vagy egyéb cím
- Tartalmazó kiadvány címe
- Létrehozó szervezet
- Létrehozó szervezeti egység
- Létrehozó személy
- Gyűjtemény
- Típus
- Kategória
- Tárgyszó
- Személynév tárgyszó
- Földrajzi név tárgyszó
- Nyelv kód
- Hozzáférés státusza
- Rekord azonosító

IV. Problémák és tanulságok

- A *REA (R19)* forrásai közül a webes gyűjtés tekinthető reprezentatívnak (próbáltunk minden releváns forrást összegyűjteni), a *MEK–EPA–DKA* a saját meglévő állományából válogatott, illetve bizonyos mértékű digitalizálás és új gyűjtés történt, de csak a szolgáltatható esetekben, így nem biztos, hogy a téma szerint minden releváns forrás megtalálható benne.
- A webes mentések az archiválás szempontja szerint történtek, és nem az engedélyezések szerint, így előfordulhat, hogy az egy egységben mentett tartalom egyik részére van engedély, de a vele együtt eltett kapcsolt anyagra (pl. videó) nincs.
- A szabad szavas kereső csak az archivált webes tartalomban működik, mert a *SolrWB* csak a *WARC*-okat tudja megjeleníteni, a többi gyűjteménynél csak a metaadatokban lehetséges a keresés (a keresést a *Solr* végzi, az tudna keresni a *MEK-EPA* állományban is). A *MEK-EPA-DKA* gyűjteményeknek van közös szabad szavas keresése (a *DKA*-ban csak metaadatban, illetve a közös keresőnél az ismertetőben keres), jó lett volna összekapcsolni a webes keresővel, de megjelenítés problémája miatt nem lehetett.
- Az előadásra készülve átnéztük mi közpénzes tartalom a webcímek közül, hogy utólag szabadon elérhetővé tegyük ezeket (a *REA* létrehozásakor még nem élt ez a lehetőség), de technikai okokból még nem tudtuk bővíteni ezekkel a szabadon hozzáférhető tételeket.

Köszönöm a figyelmet!

webarchivum.oszk.hu

webarchivum@oszk.hu
visky.akos.laszlo@oszk.hu