

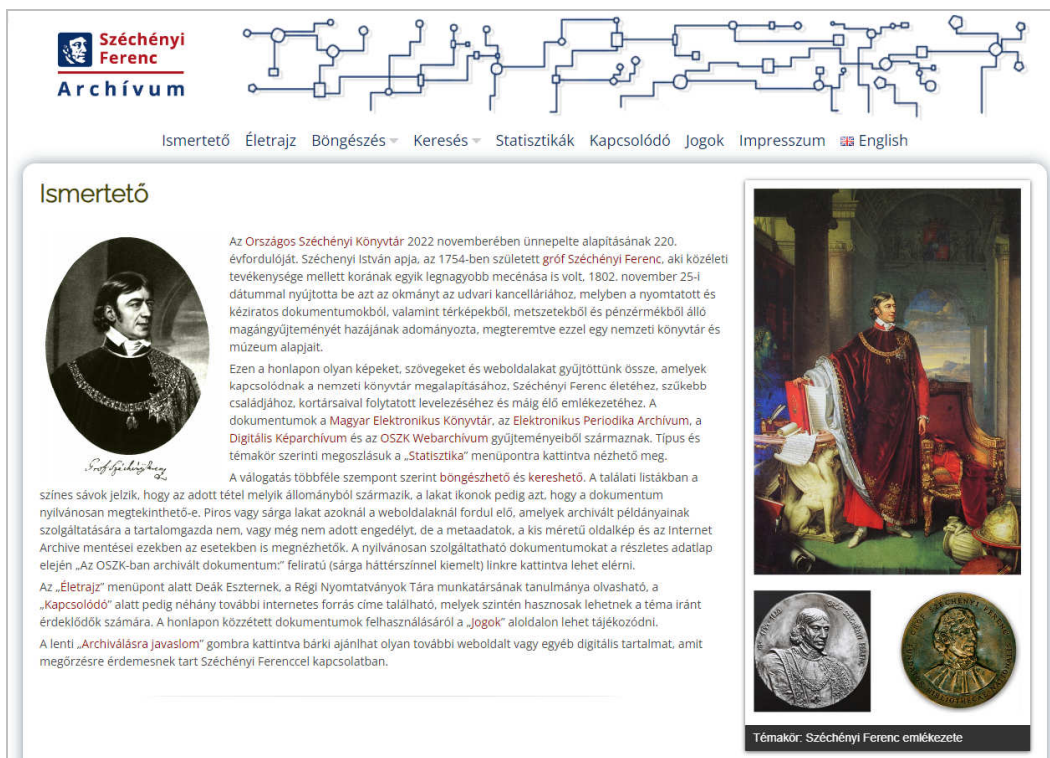
## Az OSZK Webarchívum 2023 júliusi hírei

### Archiválás, tesztelés

A 143 terabájtnyi régi tárhely átköltöztetése csak július végén fejeződik be, ezért a tematikus mentések ebben a hónapban is szüneteltek és a webtér szintű aratást is el kellett tolni augusztusra. A nyilvános szerveren levő Web Curator Tool rendszerben rendben lefutottak az ütemezett feladatok és folytattuk a podcast csatornák adásainak egyedi mentését is egy böngészőkiegészítő segítségével. Ebben a hónapban 161 csatornáról 5338 hangfájlt sikerült letölteni 274,4 gigabájt össz méretben.

Mivel a jövőben – egyebek mellett – szeretnénk megújítani az archiválási munkafolyamatokat is, ezért az informatikus kollégák beüzemeltek egy teszt szervert, amin a WCT-ben kísérletezni tudunk majd a jelenleginél jóval nagyobb számú aratási feladatokkal. Ehhez a keretrendszer alaposabb ismerete szükséges, ezért a DBeaver program segítségével megnéztük a WCT adatbázisának tábláit és a közöttük levő relációs kapcsolatokat.

A teszt szerver arra is jó, hogy újabb archiváló és megjelenítő eszközöket próbáljunk ki. Előbbiek közül először a Scrapy programot és az annak menedzselésére szolgáló SpiderKeepert teszteltük. A Scrapy egy Python nyelven írt nyílt forráskódú scraper, melynek első változata 2008-ban jelent meg. Eredetileg keresőrendszerekhez szánták, de a legtöbben weboldalakból szövegek, adatok, linkek és fájlok kigyűjtésére használják. Mi elsőként a Köztérkép nevű, a magyar köztéri alkotások bemutatására 2006-ban indult közösségi szerkesztésű fotómegosztó oldal képanyagát és metaadatait próbáljuk begyűjteni vele. A [kozterkep.hu](http://kozterkep.hu) szerveren már közel 42 ezer műről vannak információk és több mint 477 ezer fotót tartalmaz. Mivel a webhely úgy lett kialakítva, hogy a Heritrix robottal nem lehet rendesen learatni a fényképeket, ezért a Scrapy-vel való „gребlyezés” a legjobb mód ennek a rendkívül értékes anyagnak a megőrzésére. A metaadatok műlapokról való kigyűjtésére már el is készült a Python script, még az alkotói- és a képpoldalak elemzése, majd pedig a hozzájuk tartozó parancsfájlok megírása van hátra.



**Széchenyi Ferenc Archívum**

Ismertető Életrajz Böngészés Keresés Statistikák Kapcsolódó Jogok Impresszum English

### Ismertető

Az Országos Széchenyi Könyvtár 2022 novemberében ünnepelte alapításának 220. évfordulóját. Széchenyi István apja, az 1754-ben született **gróf Széchenyi Ferenc**, aki közelebbi tevékenysége mellett korának egyik legnagyobb mecénása is volt, 1802. november 25-i dátummal nyújtotta be azt az okmányt az udvari kancelláriához, melyben a nyomtatott és kéziratos dokumentumokból, valamint térképekből, metszetekből és pénzérmékből álló magángyűjteményét hazájának adományozta, megteremtve ezzel egy nemzeti könyvtár és múzeum alapjait.

Ezen a honlapon olyan képeket, szövegeket és weboldalakat gyűjtöttünk össze, amelyek kapcsolódnak a nemzeti könyvtár megalapításához, Széchenyi Ferenc életéhez, szűkebb családjához, kortársaival folytatott levelezéséhez és máig élő emlékeztetéhez. A dokumentumok a Magyar Elektronikus Könyvtár, az Elektronikus Periodika Archívum, a Digitális Képtár és az OSZK Webarchívum gyűjteményeiből származnak. Típus és témakör szerinti megoszlásuk a „Statistika” menüpontra kattintva nézhető meg.

A válogatás többféle szempont szerint böngészhető és kereshető. 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 megnézhetőek. A nyilvánosan szolgáltatható dokumentumokat a részletes adatlap elején „Az OSZK-ban archivált dokumentum:” felirattal (sárga háttérrel kiemelt) linke kattintva lehet elérni.

Az „Életrajz” menüpontra kattintva Deák Eszternek, a Régi Nyomtatványok Tára munkatársának tanulmánya olvasható, a „Kapcsolódó” alatt pedig néhány további internetes forrás címe található, melyek szintén hasznosak lehetnek a téma iránt érdeklődők számára. A honlapon közzétett dokumentumok felhasználásáról a „Jogok” aloldalon lehet tájékozódni.

A lenti „Archiválásra javasolt” gombra kattintva bárki ajánlhat olyan további weboldalt vagy egyéb digitális tartalmat, amit megőrzésre érdemesnek tart Széchenyi Ferencsel kapcsolatban.

Témakör: Széchenyi Ferenc emlékezete

### A Széchenyi Ferenc Archívum kezdőlapja

## **Széchenyi Ferenc Archívum**

Július közepén több hírforrás (pl. Papageno Élménykalauz, Kossuth Rádió, Hegyvidék TV, Civil Rádió) is tudósított a Széchenyi Ferenc életét, családját, emlékezetét és a nemzeti könyvtár létrejöttének történetét bemutató új szolgáltatásról ([szechenyi.webarchivum.oszk.hu](http://szechenyi.webarchivum.oszk.hu)), melyben a digitalizált vagy eleve digitálisan született képek és szövegek mellett archivált weboldalak is találhatóak. A gyűjtemény összeállítását 2022 őszén kezdtük el, mert az Országos Széchenyi Könyvtár ekkor ünnepelte alapításának 220. évfordulóját. A honlapról jelenleg 1074 dokumentum érhető el, ezek nagy része a Digitális Képtár anyagából került beavatásra, de vannak könyvek és könyvrészek a Magyar Elektronikus Könyvtárból és tanulmányok az Elektronikus Periodika Archívumból is. Webes tartalomról 69-et archiváltunk, például Wikipédia szócikkeket, híreket, Facebook bejegyzéseket, sőt néhány szobor és emléktábla múlapját is a Köztérképről. A gyűjtemény metaadatai között többféle szempont szerint lehet keresni és böngészni. A magyar és angol nyelvű honlapon egy részletes Széchenyi Ferenc életrajz is található, melyet Deák Eszter, a Régi Nyomtatványok Tára munkatársa írt, valamint van egy linkgyűjtemény további OSZK-s online forrásokra, köztük arra a kapcsolati hálóra, amit szintén a Digitális Bölcsészeti Központ fejlesztett. Ez a vizualizáció azokról a személyekről és intézményekről készült, akiket/amelyeket Széchenyi Ferenc levélben tájékoztatott a nemzeti könyvtár megalapításáról és könyvgyűjteménye katalógusának első kinyomtatott kötetével is megajándékozott.

## **Metaadat nyilvántartás**

A nyári uborkaszezonban van lehetőségünk a korábban összeválogatott webcímek ellenőrzésére, amiből következtetni lehet a webhelyek megszűnésének gyakoriságára is. Július 15. és 18. között a természet- és műszaki tudományok, valamint szakterületek kb. másfél évvel ezelőtt létrehozott nyilvántartását néztük át. A 2057 honlapról és blogból 101 biztosan megszűnt, további 25 pedig szintén nem elérhető, de a Google robotja néhány hete még le tudta indexelni őket, így ezeket egyelőre nem töröltük ki az aratásra kerülő oldalak közül. Hasonló, öt-hat százalék körüli hibaarányt eredményezett a bölcsészeti- és társadalomtudományok, illetve az ezekhez kapcsolódó szakterületek címlistájának átvizsgálása július harmadik hetében. Az 5912 webhelyből 203-at kellett megszűntnek nyilvánítani és 101 volt a talán csak ideiglenesen szünetelő oldal.

A műszaki területek között tartjuk nyilván az internettel foglalkozó webhelyeket is és most ezeknél létrehoztunk egy „linkgyűjtemények” nevű új kategóriát, amibe már 42 olyan forrást vettünk fel, ahol főként magyar oldalakra mutató linkek vannak témakörök szerint rendezve. Bár a Startlap és társai a 2000-es években voltak igazán népszerűek és a címlistáik nagyrészt már elavultak, a kutatóknak mégis értékes információforrások a régi magyar internet történetének felderítéséhez, a még működő linkek között pedig lehetnek olyanok, amelyeket fel tudunk majd használni a webtér nyilvántartásunk bővítésére.

## **Tervezés és újratervezés**

Folyik a webarchiváló csoport munkafolyamatainak és nyilvántartásainak újragondolása, különös tekintettel arra, hogy mit lehetne ezek közül legalább részben automatizálni. Júliusban a fontosabb archiválási formák (webtér aratás, webhelyek/webhelyrészek mentése, esemény alapú és egyéb fókuszált gyűjtések) folyamatábráit készítettük el, megterveztük a metaadatok adatbázisának tábláit, valamint elkezdtünk gondolkodni a hosszú távú megőrzéshez szükséges OAIS modell szerinti beadási és archiválási csomagok összeállításán.

A hónap végén aktualizáltuk a Jira rendszerben nyilvántartott egyéb feladatainkat és felvettünk néhány újabbat is, például az augusztusi budapesti Atlétikai VB-vel kapcsolatos hírek gyűjtését, illetve az idei „404-es” workshop előkészítését.

## Ismeretterjesztés és ismeretszerzés

Kalcsó Gyula „Building and Processing Corpora from Archived Web Content” címmel tartott előadást július 5-én az MTA Könyvtár és Információs Központ által szervezett LIBER 2023 konferencián. A webarchívumban található ukrain hírek szövegének feldolgozását bemutató prezentáció letölthető a honlapunkról.

A [webarchivum.oszk.hu](http://webarchivum.oszk.hu) honlapon több változtatás is volt az elmúlt hetekben. Egyrészt felkerült rá a Széchényi Ferenc archívum ismertetője és egy rá mutató gomb a kezdőlapra levő gyorslinkek között, másrészt átírtunk minden olyan hivatkozást a projekt keretében készült cikkek és előadások oldalán, amelyek még a régi honlapon levő fájlokra mutattak.

A MIA Wikit ([webarchivum.oszk.hu/mediawiki](http://webarchivum.oszk.hu/mediawiki)) 22 új szócikkkel – főként szoftverek és szolgáltatások ismertetőivel – bővítettük, ami így már 833 bejegyzést tartalmaz. A nyilvános wiki mellett volt egy belső használatra szánt, részletes technikai információkat tartalmazó wikink is az OSZK Redmine rendszerében. Mivel a Redmine frissítése miatt ennek jövőbeli elérése bizonytalanná vált, ezért a szócikkeket kiexportáltuk PDF formátumban.

A Könyvtári Intézet 2023. október 2. és 5. között hirdeti meg „Az internet archiválása mint közgyűjteményi feladat” című tanfolyamunkat, ami valószínűleg utoljára kerül megtartásra ebben a formában, mert utána lejár az akkreditációja.

Július 5-én a Digital Preservation Coalition szervezet a webarchívumok minőségellenőrzéséről rendezett webináriumot „Web Archiving: How to Achieve Effective Quality Assurance” címmel, melyen a British Library, a brit National Archives és az amerikai Library of Congress szakemberei tartottak számunkra is nagyon tanulságos előadásokat az általuk használt eszközökről és módszerekről.

Az International Internet Preservation Consortium májusi konferenciájának felvételei felkerültek a szervezet YouTube csatornájára. A megnyitó és a személyes részvétellel megtartott panelbeszélgetések és előadások mellett visszanezhetők az online nap előre felvett videóik is ([youtube.com/@iipc8855/playlists](https://youtube.com/@iipc8855/playlists)).

## Együtműködések

Július 11-én újabb egyeztetés volt a luxemburgi webarchívum munkatársaival az ősze tervezett kölcsönös tanulmányutakról. A Teams megbeszélésen Gerencsér Judit, az OSZK általános főigazgató helyettese is részt vett.

Július 18-án a Szegedi Tudományegyetem és az Országos Széchényi Könyvtár vezetői historiográfia kutatási, természetes nyelv feldolgozási, webarchiválási, valamint országos könyvtári szolgáltatási feladatokra kiterjedő együtműködési megállapodást írtak alá. A webarchiválás területén az első feladat a Karikó Katalin gyűjtemény webes tartalmakból álló részének közös építése lesz.

Az együtműködés formalizálása előtt is már többször egyeztettünk a SZTE Klebelsberg Kuno Könyvtárának munkatársaival és július 20-án a szegedi egyetemi könyvtár minőségirányítási asszisztensével, Frankó Mátéval személyes találkozóra is sor került az OSZK-ban, akit a webarchívum működésének kívülről kevésbé látható részleteiről tájékoztattunk.

\* \* \*

Az OSZK Webarchívum hírei a <https://webarchivum.oszk.hu/a-projektrol/hirek-esemenyek/> oldalon kísérhetők figyelemmel. Kapcsolati cím: [webarchivum@oszk.hu](mailto:webarchivum@oszk.hu)