

Az MNMKK OSZK Webarchívum 2025. novemberi és decemberi hírei

Archiválás és nyilvántartás

Közvetlenül az év végi ünnepek előtt befejeződött a webarchívum által használt szerverek és tárhelyek átmásolása az új infrastruktúrára. Jelenleg a költözés (pl. az IP címek megváltozása és a tűzfalbeállítások) miatt keletkezett problémák felderítése és javítása zajlik. A nyilvános gyűjtemény esetében a legnagyobb gond, hogy hibaüzeneteket ad az OpenWayback megjelenítő, ezért a mentések csak a PyWb és a Solr-Wayback programokkal nézhetők vissza. A zárt archívumnál viszont még ezek sem működnek, valószínűleg újra kell majd indexelni ezt az állományt is.

Az átállás ideje alatt, vagyis október második felétől január elejéig le kellett állítani a tömeges aratásokat és az egyéb ütemezett mentéseket. A tesztszerveren viszont szinte folyamatosan tudtunk archiválni a Browsertrix szoftverrel és naponta letölteni vele az eddig bekonfigurált öt hazai hírportál friss tartalmait. Ezek összmenyisége novemberben és decemberben megközelítette a 350 gigabájtot. November 28. és december 2. között futtattunk egy teljességre törekvő Browsertrix-aratást (116 GB) a szabadeuropa.hu webhelyről, illetve az ArchiveWeb.Page (AWP) programmal csináltunk mentéseket a kiadványhoz kapcsolódó Facebook és Instagram fiókokról is, mert a Szabad Európa magyar szerkesztősége 2025. november 21-től befejezte működését.

November 6-án szintén az AWP-vel letöltöttük a 2024 végén indult „Könyvtári kihívás” pályázat eredményhirdetéséről szóló híreket és közösségi média posztokat, december elején pedig létrehoztuk a KVTkih2025 kódjelű részgyűjteményt és elkezdtük az új pályázattal kapcsolatos webtartalmak heti szintű archiválását. December 28-ig 107 weboldalt, továbbá több mint 600 Facebook, Instagram és X bejegyzést, illetve YouTube és TikTok videót mentettünk le és töltöttünk fel a zárt webarchívumba. Az aktuális URL címlista elérhető a honlapunkon az [„Esemény alapú aratások” aloldaltól](#).

A nyilvános archívumot üzemeltető szerver és a hozzá tartozó tárterület átköltöztetése után december első hetében ismét aktiváltuk a Web Curator Tool keretrendszeren belül a demó gyűjteményben levő webhelyek ütemezett mentéseit, miután ellenőriztük, hogy elérhetőek-e még az élő weben. Hogy a leállás miatti kiesést pótoljuk, 393 honlap/blog esetében egy azonnali aratást is lefuttattunk. A publikus állomány metaadatolása is újraindult az év utolsó hónapjában: 18 részletes XML fájl készült el és került ki a demó archívum oldalára.

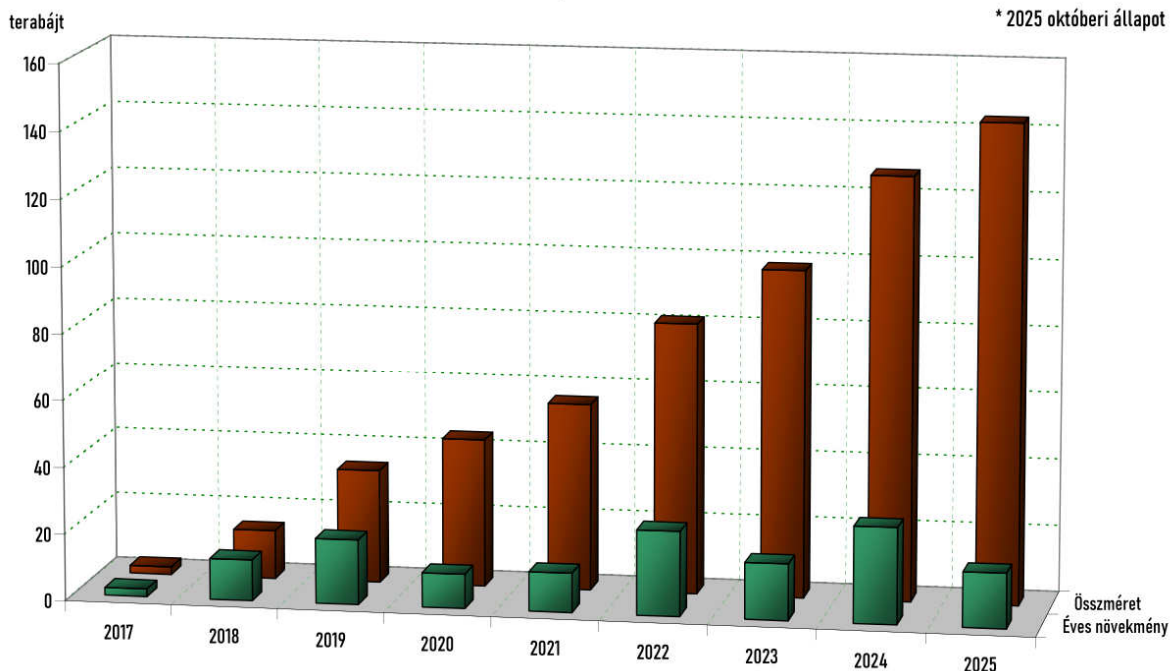
Folytattuk a nyári webtér aratás előtt a WARC fájlokból kigyűjtött, a .hu országhoz tartozó és korábban még külön nem nyilvántartott URL címek átnézését és a gyűjtőkörünkbe tartozó webhelyek besorolását valamelyik tematikus vagy műfaji részgyűjteménybe. Az elsődleges válogatási szempont most az idegenforgalom és vendéglátás volt, így a TURIZMUS kódnevű táblázat az év utolsó két hónapjában kb. 4 ezer címmel bővült, de több száz új URL-t vettünk fel a többi táblába is. A közgyűjtemények nyilvántartása szintén frissült a korábban összegyűjtött, de a megfelelő kategóriába és alkategóriába eddig még be nem sorolt weboldallal.

Rendezvények

November 12-én tartottuk meg a kilencedik „404 Not Found - Ki őrzi meg az internetet?” című konferenciát és workshopot, ami iránt a korábbi évekhez hasonlóan nagy volt az érdeklődés. Több mint százan regisztráltak és nagyjából fele-fele arányban voltak azok, akik személyesen az OSZK-ban, illetve azok, akik az élő YouTube közvetítést követve vettek rajta részt. Az eseményről megjelent sajtóhír a nemzeti könyvtár [honlapján olvasható](#), az előadások kivonatai és prezentációi pedig a [rendezvény weboldalán](#) érhetők el.

A zárt archívumban levő WARC fájlok mérete évenként és összesen *

* 2025 októberi állapot



	2017	2018	2019	2020	2021	2022	2023	2024	2025
Éves növekmény	2,3	12,2	19,4	10,3	11,8	25,4	16,9	28,8	16,5
Összméret	2,3	14,5	33,9	44,2	56,0	81,4	98,3	127,1	143,6

Egy grafikon Drótos László „Adatok az adatokról” című prezentációjából

Megjelent a 2025. május 13–15. között a Széchenyi István Egyetemen megrendezett Workshop 25 lektorált előadását tartalmazó kötet, benne Kalcsó Gyula „A magyar webtér aratásával kapcsolatos kurátori feladatok” című tanulmánya. A könyv az MTA Könyvtárának repozitóriumából [tölthető le](#), az eredeti előadás prezentációja pedig a honlapunkon [található meg](#).

A nemzetközi IIPC konzorcium webinárium sorozatának részeként november 20-án két francia nyelvű bemutatót nézhettek meg az érdeklődők, melyeket a belga és a svájci nemzeti könyvtárak szakemberei tartottak „BelgicaWeb: Preserving Belgian Digital Heritage”, illetve „Access e-Helvetica System Architecture and Automated Web Portal Testing” címmel.

A projekt további hírei a <https://webarchivum.oszk.hu/a-projektrol/hirek-esemenyek/> oldalon kísérhetők figyelemmel. Kapcsolati cím: webarchivum@oszk.hu