Webarchiválás projekt
2019. évi beszámoló

Az OSZK-ban 2017-ben indultak el a nemzeti szintű webarchívum előkészítő munkálatai az OKR projekt részeként. A cél a weben levő magyar vagy magyar vonatkozású – elsősorban kulturális, oktatási, tudományos, vagy közéleti témájú – nyilvános tartalom időnkénti mentése és hosszú távú megőrzése kutatási és egyéb felhasználásokhoz.

Az archiválás több szinten zajlik: a webhelytulajdonosokkal való szerződéskötés után egyedi mentések formájában a nyilvános archívum számára; valamint az egyelőre még zárt archívumban téma, műfaj, illetve esemény alapján válogatott részgyűjtemények tömeges aratásával, és évente 1-2 alkalommal a magyar webtér minél nagyobb részére kiterjedő, de kis mélységű „pillanatfelvételek” formájában.

A zárt állomány a KIFÜ által biztosított szerveren van és 2019 végén 12 db tematikus részgyűjteményt, 5 db esemény alapút és 1 db műfaj szerintit (e-periodika) tartalmazott. A válogatott webhelyek száma kb. 25.000, az automatikusan összegyűjtött webhely címek száma kb. 250.000, az összméret 35 terabájt, ami több mint 600 millió lementett fájlt/URL címet jelent. Az OSZK-ban levő nyilvános gyűjteményben 186 db honlap, blog, illetve periodika található részletes metaadatokkal és teljes szövegű keresési lehetőséggel, továbbá 44 db OSZK-s webhely mentései. 2019 végén a KDS pályázat részeként elkészült egy esemény alapú mintaalkalmazás is, mely a Rákóczi-emlékév kapcsán mutatja be a webarchívum integrálásának lehetőségét más digitális gyűjteményekkel.

Az archiválás mellett az ismeretterjesztés terén is jelentős eredményeket értünk el. Összeállítottunk egy e-learning tananyagot a közgyűjteményekben dolgozók számára, melyet 2019-ben a Könyvtári Intézet szervezésében két alkalommal hagyományos tanfolyami formában is oktattunk. Elkészítettük a középiskolásoknak szánt multimédia oktatóanyag kéziratát az internetes tartalmak személyes és az intézményes archiválásának fontosságáról. Több tanulmányt, cikket és összefoglalót publikáltunk magyar és külföldi szaklapokban, továbbá előadásokat tartottunk itthon és néhány más országban a webarchívumok metadatolásáról, a kutatási célú hasznosításukról, az internetes tartalmak megőrzésének fontosságáról, a webarchiválás oktatásáról, valamint az OSZK-s projekt eddigi eredményeiről. November 14-én immár harmadik alkalommal rendeztük meg a „404 Not Found - Ki őrzi meg az internetet?” című workshopot, melyen közel 90-en vettek részt.

Az OSZK 2018-ban csatlakozott a webarchiválással foglalkozó intézményeket összefogó International Internet Preservation Consortium nevű szervezethez. 2019-ben Zágrábban volt az IIPC éves közgyűlése és konferenciája, ahol az OSZK-t Németh Márton képviselte, aki ellátta az egyik szekció vezetését és két előadást is tartott. Ugyancsak ő vesz részt az IIPC oktatással foglalkozó munkacsoportjában, melynek keretében oktató és ismeretterjesztő anyagokat fejlesztenek. Jó személyes, vagy legalább virtuális kapcsolatokat sikerült kiépíteni többek közt szlovák, cseh, osztrák, holland, belga, dán szakemberrel, akiktől nemcsak tanulni tudunk, hanem már mi is képesek vagyunk tesztelési tapasztalatokkal és javaslatokkal segíteni nekik. Felvettük a kapcsolatot a világ legnagyobb webarchívumával, az amerikai Internet Archive-val is, ahonnan árajánlatokat kaptunk a gyűjteményükben található magyar webtartalom visszakeresésére, illetve a magyar domének listájának átvételére.

Folyamatosan egyeztettünk az EMMI Közgyűjteményi Főosztályával a webarchiválásról szóló jogszabály előkészítéséről, mely várhatóan 2020-ban fog megjelenni. Egyeztettünk továbbá az OKR keretében az RDA-alapú metaadatolással foglalkozó munkacsoporttal a webes tartalmak leírásának módszertanáról, valamint a Nemzeti Névtér fejlesztésének vezetőjével a webarchívum és névterek kapcsolatáról.