

Webarchiválási workshopok a nemzeti könyvtárban



Az OSZK-ban 2017-ben elkezdődtek egy nemzeti webarchívum előkészítő munkálatai. Az E-könyvtári Szolgáltatások Osztályon már a kétezres évek elejétől zajlik internetes dokumentumok gyűjtése, feldolgozása és szolgáltatása, de ez eddig jellemzően csak az önálló állományok formájában letölthető könyvekre, folyóiratszámokra és képekre terjedt ki. A webarchiválás lehetővé teszi a nagy mennyiségű fájlból álló honlapok, blogok, portálok és egyéb típusú webhelyek megőrzését és kutathatóvá tételét is. Mivel nálunk még újdonságnak számít ez a tevékenység, a munkacsoport tagjai az archívum kialakítása mellett fontosnak tartják az ismeretterjesztést és a szakmai párbeszéd megindítását is. A memóriaintézmények együttműködése elengedhetetlen ezen a területen, hiszen a magyar webtérben már most sokkal több dokumentum van, mint amennyi korábban hagyományos hordozón megjelent, és ezeknek csak egy kis töredéke kerül be köteles példányként. Ahhoz, hogy a digitálisan születő kultúránk egy reprezentatívnek tekinthető része megmaradjon a jövő számára, a hazai közgyűjteményekben olyan szakemberekre lesz szükség, akik jártasak az internetes tartalmak és azon belül a jelenlegi legfontosabb médium, a web archiválásában. Az ismeretterjesztő előadások, a szaklapokban publikált cikkek, a témával foglalkozó wiki és a Könyvtári Intézet által szervezett tanfolyamok mellett az évente megrendezett *404 Not Found – Ki őrzi meg az internetet?* című workshop is ezt a célt szolgálja. A félnapos rendezvényen alkalmanként 70-90 fő vesz részt, akik a formális program után az előadókkal együtt megvitatják az elhangzottakat. Az előadások felvételei bekerülnek a Videotóriumba, ahol bárki számára elérhetők. A téma itthoni újszerűsége miatt az online és a hagyományos médiában is megjelennek tudósítások ezekről az eseményekről.

Az első ilyen szakmai fórumra 2017. október 13-án került sor. A megnyitó előadást az OSZK informatikai igazgatója, Lendvai Miklós tartotta *A web mint a kulturális gyűjtőkör része* címmel. Ezt követően a magyar

származású Kees Teszelszky mutatta be az általa vezetett holland webarchívumot. Majd két olyan előadás hangzott el, melyek korábbi projekteket ismertettek: Peternák Miklós, a C3 Kulturális és Kommunikációs Központ igazgatója a digitális művészeti alkotások megőrzéséről beszélt, Kampis György, az ELTE Tudománytörténet és Tudományfilozófia Tanszékének egyetemi tanára pedig a Magyar Internet Archívum létrehozására tett első komolyabb kísérletet mutatta be: az akadémiai szféra honlapjairól és néhány hírportálról 2013–2016 között végzett webaratások eredményeit ismertette. Végül Drótos László témafelelős és Németh Márton webkönyvtáros a nemzeti könyvtárban elindult kutatási-fejlesztési munka első néhány hónapját foglalta össze: a szoftverterveztek tapasztalatait, a mentések során észlelt hibajelenségeket, az ismeretterjesztési és kapcsolatépítési törekvéseket. A következő időszakra tervezett feladatokat is felvázolták, köztük egy, a webarchiválás lehetőségeit és korlátait illusztráló nyilvános gyűjtemény kialakítását.

Második alkalommal 2018. november 15-én gyűltek össze a téma iránt érdeklődők az OSZK dísztermében, ahol Tüske László főigazgató üdvözlő szavai után Latorcai Csaba, az Emberi Erőforrások Minisztériuma államtitkára beszélt az internet mint kulturális örökség jövőjéről. Drótos László beszámolt az elmúlt egy évben elért eredményekről, a tematikus és a webtársintű aratásokról, bemutatta a nyilvános gyűjtemény szolgáltatófelületét és a metaadat rekordokat, majd bejelentette, hogy ettől az évtől a magyar nemzeti könyvtár is tagja lett az internetes tartalmak megőrzésével foglalkozó IIPC szervezetnek (International Internet Preservation Consortium). A visszatekintő előadás folytatásaként Moldován István, az E-könyvtári Szolgáltatások Osztályának vezetője egy már üzemszerűen működtethető magyar webarchívum tervét vázolta fel, felsorolva ennek jogi, technikai, szervezeti és egyéb feltételeit. Az OSZK jogásza, Halász Annamária az uniós ajánlások és a külföldi példák ismertetése után a készülő



hazai törvényi szabályozásról beszélt. Kokas Károly, a Szegedi Tudományegyetem Klebelsberg Kuno Könyvtárának főigazgató-helyettese egy most formálódó, új szakterületre hívta fel a figyelmet. A webhistoriográfia, vagyis az archivált webtartalmak kutatása ugyanolyan önálló történeti segédtudománnyá kezd válni, mint amilyen a régészet, az oklevéltan vagy a címertan, hiszen a weben a nagy globális folyamatok és a hétköznapi élet lenyomatai egyaránt megtalálhatók. Végezetül Parti Ádám, az Országgyűlési Könyvtár csoportvezetője az állami és kormányzati online tartalmak megőrzésének külföldi gyakorlatát mutatta be, és külön kiemelte, hogy az ausztrál PANDORA archívumban a weboldalak mellett a kormánytagok nyilvános Twitter üzeneteit és a helyi önkormányzatok üléseiről interneten sugárzott videókat is gyűjtik. A rendezvény kerekasztalvitával zárult, melyen a könyvtárosok mellett múzeumi és levéltári szakemberek beszélgettek egymással és a közönséggel a digitálisan születő tartalmak megőrzésének lehetőségeiről.

A harmadik workshopra 2019. november 14-én került sor, melyen Hammerstein Judit, az OSZK megbízott főigazgatója köszöntötte a közönséget és Latorcai Csaba államtitkárt, aki megnyitójában a web történetéről, kulturális szerepéről és a nemzeti könyvtár webarchiválási törekvéseinek fontosságáról beszélt. Az első prezentációt ezúttal is Drótos László tartotta az OSZK webarchívumának újdonságairól. Bevezetésként elmondta, hogy a web már nemcsak egy hipermédia-dokumentum, hanem egyre több benne a szoftverelem, ami megnehezíti

a lementését és a hosszú távú megőrzését. Majd az egyik legfontosabb részfeladatról, az online időszaki kiadványok archiválásáról vetített ki grafikonokat és képernyőfotókat, végül bemutatta a Közgyűjteményi Digitalizálási Stratégia (KDS) pályázat keretében a II. Rákóczi Ferenc-émlékév alkalmából készülő gyűjteményt, melyben együtt böngészhetők lementett weboldalak és egyéb digitális vagy digitalizált dokumentumok. Kokas Károly a helyi érdekű webarchiválás lehetőségeiről, egy regionális „Szegedicum” gyűjtemény ter-

véről beszélt. Az MTA Könyvtár és Információs Központ informatikai főigazgató-helyettese, Holl András *no Webarchive* című előadásában olyan példákat említett, amelyek esetében nem a webaratás a legmegfelelőbb módszer a weben található – elsősorban tudományos – anyagok archiválására. A szünet után még két előadás hangzott el: Indig Balázs számítógépes nyelvész az ELTE Digitális Bölcsészeti Központjában folyó munkáról számolt be, melynek keretében kutatási célokra használható szövegtörzset hoznak létre magyar hírportálok aratásával, Ilácsa Szabina, az OSZK könyvtárosa pedig az aktív és az archivált webhelyek metaadatolási problémáiról beszélt, valamint arról, hogy hogyan lehetne az RDA (Resource Description and Access), illetve az LRM (IFLA Library Reference Model) alapján leírni ezeket. A programot ezúttal is beszélgetés zárta az előadók és a közönség között, főként a KDS keretében való együttműködésről.

A tervek szerint a 2020. évi rendezvényen a környező országokból meghívott előadók is részt vesznek, és fő témája a közép-európai webarchívumok közötti kapcsolatépítés lesz. Az eddigi 404-es workshopok prezentációi, videofelvételei és sajtóhírei a webarchívum megújult honlapján érhetők el: webarchivum.oszk.hu/404_workshop.

Drótos László
drotos.laszlo.at.oszk.hu