

## Áttekintés az EuropeanaTech 2018 konferenciáról

Az idei évben a holland kikötőváros, Rotterdam, szűkebben az ott horgonyzó SS Rotterdam nevű óriási szálloda- és rendezvényhajó adott otthont az Europeana nemzetközi technikai konferenciájának, az EuropeanaTechnek. A közgyűteményi informatika kortárs, aktuális témáit felvonultató szakmai összejövetelen két napon keresztül tanácskoztak a résztvevők. Hazánkat a Petőfi Irodalmi Múzeum és a Magyar Nemzeti Levéltár munkatársai képviselték.

A 2018-as szakmai összejövetel színvonalas programját a szervezők egy Wikidata workshoppal egészítették ki, amelyet a holland nemzeti könyvtárban, a hágai Koninklijke Bibliotheekben, az Europeana Foundation székhelyén tartottak május 14-én. A PIM részvételét ezen a kísérőprogramon erősen indokolta, hogy az intézmény szakinformatikai programjának fontos részét képezi egy, már több hónapja futó projekt, amelynek fő célja a több százezer nevet tartalmazó személynévtér-állomány Wikipédia- illetve Wikidata-hivatkozásokkal történő adatgazdagítása, illetve a PIM névtér URI-jainak elhelyezése a szócikkekben, valamint a Wikidata strukturált adattárában.

A szakmai program gerincét kétszer három, párhuzamos szekcióülés alkotta, amelyekre előzetes jelentkezés alapján osztották be a résztvevőket. Számomra a délelőtti elfoglaltságot a *Strukturált adatok a [Wikimedia] Commons-on: GLAM-ontológiák és szótárak, copy-right-modellezés* című téma tárgyalása jelentette. Ezt az ülést *Sandra Fauconnier*, a Wikimedia Alapítvány közgyűteményi stratégiákkal foglalkozó munkatársa vezette. Az általa tartott bevezető előadásban elhangzott: a Wikimedia Commons, amely a hivatalos megfogalmazás szerint nem más, mint *ingyenes és szabadon felhasználható képeket, hangfelvételeket és egyéb multimédia fájlokat tartalmazó katalógus*, 2018-ban már csaknem ötvenmillió dokumentumot tartalmaz, és átlagos éves gyarapodása ötmillió tétel. Ezek a médiaállományok azonban javarészt strukturálatlan metaadatokkal kerülnek a rendszerbe; az alkalmazott kategória-rendszer hierarchiája bonyolult és következetlen, helyenként nehézkes, hosszú megnevezéseket alkalmaz. A Wikimedia Alapítvány három évig tartó *Structured Data on Commons* (SDC) projektje (2017-2019) azonban lehetővé teszi, hogy átgondolják és reorganizálják a

Commons metaadat-rendszerét, valamint elősegítsék az adatok hatékonyabb, eredményesebb megjelenítését, keresését, szerkesztését, és biztosítsák azok újrafelhasználását akár több nyelven is. Az átalakítás legfontosabb eleme, hogy bizonyos adatelemek – a tartalom létrehozásában részt vevő, nevesített közreműködők, az ábrázolt személyek, események, tárgyak, stb., a földrajzi helyszínek, a tartalomhoz kapcsolódó dátumok, valamint a *rightsstatements.org* szótára segítségével formalizált copyright-információk – a Wikidatából emelkednének át a médiatartalmakhoz, és megjelenítésük is azzal együtt történne. A koncepciót egy mintarekord segítségével meg is jelenítették.

Az adatmodell átalakítását – a projekt egyik fő célkitűzését – a Commons munkatársai a Wikidata-közösségek, valamint a közgyűteményi szektor (az ún. GLAM-intézmények) bevonásával tervezik megvalósítani, hogy az új leíró elemkészlet minél jobban megfeleljen a közgyűtemények igényeinek, továbbá biztosítsa az átjárhatóságot az általuk jelenleg alkalmazott megoldásokkal. Előkészítő munkaként a 2017. évben több felmérést is végeztek az intézmények médiakezelési gyakorlatának megismerésére, amelynek nem megfelelő végkövetkeztetése a sokszínűség volt: az intézmények eltérő médiatípusokat gondolnak és eltérő metaadat-készleteket alkalmaznak a feldolgozó munka során. Ebből adódik, hogy ahhoz, hogy a Commons, valamint a Wikidata hatékony kiegészítője, sőt integráns része lehessen a közgyűteményi szolgáltatási platformnak, szükség van arra, hogy a legfontosabb metaadat-szabványokat, metaadat-sémákat megfeleltessék a Wikidata egyes kifejezéseinek.

Az előadás után a résztvevők a workshop szervezői által összeállított kérdéseket válaszolták meg egy közösen szerkesztett dokumentumban. Az első néhány pontban arról esett szó, mennyire ismerik és alkalmazzák az intézmények a Commons-t: van-e saját kategóriájuk, intézményi sablonjuk, történik-e rendszeres, folyamatosan koordinált feltöltés, vagy csupán esetlegesen, önkéntes munkával kerülnek fel állományok. Használ-e az intézmény különféle ellenőrzött szótárakat, ontológiákat a feldolgozás során? Megjelenít-e szerzői jogi információkat? Rákérdeztek arra is, milyen különbségek mutatkoznak ugyanazon médiaállomány saját katalógusbeli, illetve Commons-os megjelenítése között; továbbá arra, hogy használja, vagy tervezi-e használni az intézmény az IIF, valamint a *rightsstatements.org* által nyújtott szolgáltatásokat.

A délutáni szekcióban a tanácskozás – ugyancsak *Sandra Fauconnier* vezetésével – a Wikidata egyre növekvő közgyűteményi authority-szerepének áttekintésével folytatódott. A Wikipédiát is támogató, de attól független tudástár, amelyből a Google-találatok mellett megjelenő, ún. *infoboxok* is táplálkoznak, strukturált adatokat, állításokat tartalmaz a világban létező dolgokról, entitásokról. Ezeket elemeknek (*item*) nevezzük (pl. Petőfi Sándor, az Europeana, vagy a világegyetem is egy ilyen elem), az állításokat pedig ún. tulajdonságok (*propertyk*) segítségével hozzuk létre: Petőfi Sándor *halálozási éve* 1848, az Europeana *székhelye* Hága, a világegyetem *kor* 13.8 milliárd év. Ilyen tulajdonságok segítségével különféle külső azonosítókat, authority-ID-kat is társíthatunk a leírt elemhez, amelyek segítségével ellenőrizhető, szükség esetén növelhető a Wikidatában tárolt adatok hitelessége (Petőfi Sándor *azonosítója a Petőfi Irodalmi Múzeum katalógusában* PIM67655.) Jelenleg a Wikidata 2500-nál is több adatbázis külső azonosítóinak elhelyezhetőségét biztosítja. Ugyanakkor maga a Wikidata is szolgálhat ún. *authority hub*ként, az elemek URI-jai – a tartalmak feldolgozása közben, vagy utólag – beemelhetők más adatbázisokba, katalógusokba, szolgáltatásokba, gazdagítva ezáltal azok adattartalmát.<sup>1</sup> De segítséget

nyújthat például a szerzői jogi kérdések tisztázásában is: a *CopyClear* nevű, igen hasznos segédeszköz a Wikidata adataira támaszkodva gyorsan és hatékonyan ad információkat pl. a gyűjteményben található képzőművészeti alkotások jogi helyzetéről, s így jelentősen meggyorsíthatja azok közzétételét.

A szekcióülések végeztével még egy plenáris előadást hallgathattunk meg, amely egy igen előremutató, összefoglaló jellegű művészettörténeti projekt, a *Sum of All Paintings* bemutatásáról szól. A fő cél a jelentős képzőművészeti alkotásokról szóló tudás bővítése, kiegészítése a Wikidatán. Számos közösségi munkában vehetnek részt az érdeklődő önkéntesek: a hiányos, festményekről szóló Wikidata-elemek hiányzó adatai (alkotók, keletkezési dátumok, technikai részletek, fellelhetőségi információk stb.) pótolhatók, illetve természetesen új elemek is létrehozhatók még nem leírt, jelentős képekről.

Másnap reggel vette kezdetét a tulajdonképpeni EuropeanTech-konferencia, amelyet, mint korábban említettem, a város kikötőjében horgonyzó hajón rendeztek. Az impozáns, tágas színházteremben került sor a két napig tartó tanácskozás megnyitójára, majd két vitaindító plenáris előadásra, melyek közül az egyiket *George Oates*, a Flickr korai fejlesztője, a másikat *Ruben Verborgh*, a genti egyetem szemantikustechnológia-professzora tartotta. Az utóbbi előadó a közgyűjteményi aggregáció napjaink decentralizált(ságra törekvő) világhálóján betöltött szerepét vizsgálva arra a következtetésre jutott, hogy alapvető átalakulásra van szükség a szolgáltatói infrastruktúrában. Az aggregátoroknak nem elsősorban összegyűjteniük, hanem továbbítaniuk, lekérdeznüik és megjeleníteniük kell az egyedi gyűjteményekből érkező adatokat, más szóval a központi szerepvállalás helyett elő kell segíteniük a gyűjtemények közötti hálózat kialakulását, s így olyan felületek hozhatók létre, ahol akár minden adatelem más forrásból származik. Ez megkönnyíti az innovatív alkalmazások piaci térnyerését, akik az adatszolgáltatóktól származó, szabad felhasználásra közzétett adatokból részesülve a szolgáltatás minőségétől és nem a birtokolt adatok mennyiségétől függő versenyben vehetnek részt.

A plenáris előadások után a szakmai munka három párhuzamos szekcióban folytatódott, amelyek az adatokkal, a szolgáltatás különböző kérdéseivel, valamint a többnyelvűséggel foglalkoztak. *Frédéric Kaplan*, a lausanne-i Ecole Polytechnique Federale digitális bölcsészeti tanszékének vezetője a *Time Machine* projekt bemutatásának szentelte előadását: a nagy volumenű, 32 ország körülbelül 200 intézményét tömörítő kezdeményezés célja, hogy a ránk maradt kulturális örökség nagyléptékű digitalizálását és elemzését követően bemutathassa az európai városok fejlődésének történetét, valamint a köztük kialakult gazdasági-kulturális hálót. Ennek érdekében igen nagy teljesítményű technikai eszközöket és a legmodernebb MI- és gépi tanulási technológiákat kell bevetni. A bemutatóban szó esett a 2012 óta futó *Venice Time Machine* című projektről is, amely Velence városának történeti dokumentumait dolgozza fel, s mintegy előfutára, kísérleti terepe az egész Európát érintő, gigaméterű vállalkozásnak.

*Cees Snoek* előadása a videófelvetelek automatikus tartalomelemzésének izgalmas kérdéséről szól. A szakember becslése szerint 2022-re körülbelül 45 milliárd kamera fog működni a világban a lehető legkülönbözőbb eszközökbe – autókba, bolti pénztárakba, drónokba, stb. – építve, ezért különösen fontos, hogy a rögzített hatalmas mennyiségű információt értelmezni lehessen, azaz meg lehessen állapítani, hol, mikor és mi történik a felvételeken. Az előadó által bemutatott technológia egy címkézett tanulóhalmazból kiindulva olyan modellt képes alkotni, amely alapján, ha nem is száz százalékos pontossággal,

de lehetségessé válik bizonyos objektumok (pl. hajók) felismerése a videófelveteleken. A periodikus mozgások – mondjuk egy labda pattogatása –, vagy több szereplő cselekvésének együttes értelmezése még további kutatásokat igényel, ismerte el Cees Snoek.

Igen nagy érdeklődés fogadta az ebédszünet után következő újabb plenáris előadást, melyet *Rob Sanderson* tartott a szemantikus weben nagy erővel publikált adathalmazok használhatóságáról. A szemantikus web koncepciója megváltoztatta az adatokról és összekapcsolhatóságukról való gondolkodásunkat, a Tim Berners-Lee által megfogalmazott ajánlások és az új technológiák pedig megváltoztatták azok közzétételének módját. De ez még mindig nem vitt közelebb a célhoz, ugyanis az ötcillagos modell egyik lépcsője sem szól az adatok felhasználásáról. Nem szabad azt gondolnunk, hogy az adatokat közvetlenül a végfelhasználónak készítjük – sokkal inkább fejlesztő szakemberek számára, akik webes alkalmazásaikon keresztül szolgáltatják a közzétett adatokat: nekik pedig használható adatokra van szükségük, hogy ki tudják elégíteni a jelentkező igényeket. Ha azt kívánjuk, hogy az adatainkra minőségi szolgáltatások épüljenek, akkor azokat számukra megfelelő módon kell közzétennünk – API-n keresztül, megfelelően körülhatárolt adattartalommal. Így lesz a Linked Open Datából (LOD) Linked Open Usable Data (LOUD).

A délutáni *Data* szekcióban elhangzó további bemutatók a Wikimedia Commons adatainak strukturálásáról, a DBPediáról, valamint a kereséértékelés különféle metódusairól szóltak. Az első napot *Herbert van de Sompel* zárta, aki a *Memento* webarchiválási projektről, valamint a szakterület felmerülő kihívásairól és azok megoldási lehetőségeiről tartotta izgalmas, gondolatébresztő előadását.

A második napon ugyancsak az adatok előállítása, összekapcsolása és újrafelhasználása volt az előadások vezérfonala. *Ben Vershbow* keynote-jában ismét elhangzottak azok a lényeges gondolatok, melyek már a “nulladik napi” workshopon megfogalmazódtak: a wiki-univerzumnak egyre jelentősebb szerepe lesz a közgyűjteményi területen is, mindazonáltal még nagyon sok munka van előttünk, hogy ezt az együttműködést mindkét oldal számára gyümölcsözően meg lehessen valósítani.

A szekciómunkában – az általam választott témacsoport ismét az adatok, illetve a közösségi részvétel egyes aspektusait járta körül – összesen öt előadást hallgathattunk meg. Az Észtországból érkezett *Raivo Ruusalepp* a *blockchain* nevű informatikai biztonsági technológiáról tartotta bemutatóját, amelyet pl. alternatív fizetőeszközök IT-támogatására használnak, de a kulturális intézmények számára is tartogat lehetőségeket. Hatékony alkalmazása azonban intézményi együttműködést kíván, ezért minél előbb érdemes megismerkedni vele, felfedezni azokat a potenciálokat, amelyeket pl. a Cultural Coin nevű kriptovaluta használata ígér számunkra. *Lynnsey Weissenberger* előadásában a *LITMUS* projektről, pontosabban az annak keretében készült ír népzenei ontológia fejlesztésének folyamatáról és kihívásairól hallhattak az érdeklődők. Julia Beck és Marko Knepper egy eseményalapú, az Europeana Data Modelre építő adatmodellezési megoldást ismertettek, amely a színházi és táncelőadásokon mint eseményeken keresztül kapcsolja össze az előadásról készült film- és hangfelveteleket, az előadáshoz köthető nyomtatott dokumentumokat, műsorfüzeteket, színházjegyeket, stb., illetve az egyes közreműködőket: színészeket, koreográfusokat, jelmeztervezőket.

A konferencia zárására végül ismét a nagyteremben került sor. Az utolsó keynote-ot *Emilie Gordenker*, a hágai Mauritshuis múzeum igazgatója tartotta, felvillantva olyan fest-

mények vizsgálati-restaurálási folyamatát, mint Rembrandt Saul és Dávidja. Az MA-XRF roncsolásmentes képalkotási technológia segítségével a festett műalkotások eddig el nem érhető rétegei is láthatóvá váltak az elemző szem számára. Így lehetett megállapítani, hogy a ma már vita nélkül a holland művészek tulajdonított alkotás nem kevesebb, mint tizenöt vászonra készült, amelyből az egyik egy van Dyck-portré másolatának darabja.

Inspiráló, izgalmas, aktuális – ezzel a három szóval lehetne talán a legjobban összefoglalni a három, Hollandiában töltött napot. Inspirál, mert számos fejlesztési-fejlődési irányt, kutatási kérdést, megvalósítható jó gyakorlatot kínál, amelyek a legfrissebb kutatási eredményeken, trendeken alapulnak. S rendkívül izgalmas, mert látni engedi, hogy a közgyűjtemények világa talán az egyik legnagyobb átalakulás előtt áll, sőt talán már bele is kezdett: a világot átfogó és lefedő szolgáltatási hálózat épül, amelyben Európa és a világ gyűjteményei úgy egyesítik erőiket a kulturális örökség megőrzése és bemutatása területén, hogy kompromisszumok nélkül őrizhetik meg saját különleges, egyedi ismeretjegyeiket is.

## Jegyzet

1. A finn közmédiá-szolgáltató, az YLE médiatartalmak tagelésére használ Wikidata-elemeket, a Laurentian University könyvtárának katalógusa az authority adatokról közöl bővebb információkat a segítségével. Hazánkban pl. a Petőfi Irodalmi Múzeum emelte be személynév-rekordjaiba a Wikidatáról származó azonosítókat.



*A konferencia helyszíne*