

Fejér vármegyei webarchívum szakmai beszámoló az NKA (A2033/3933)

Projektismertető

Napjainkban egész életünket meghatározzák a weboldalak és a különféle internetes tartalmak, azonban kevesen tudják, hogy ezek átlagos élettartama sokszor csupán néhány hét mielőtt örökké eltűnnének, ezért vármegyei hatókörű városi könyvtárként megőrzési gyakorlatunk részévé kívántuk tenni gyűjtőkörünk tekintetében a webarchiválást, vagyis az interneten elérhető weboldalak megőrzését. A projekt során a kezdetektől együttműködve dolgoztunk az Országos Széchényi Könyvtárral.

A Vörösmarty Mihály Könyvtár legújabb szolgáltatásaként megjelenő Fejér Vármegyei Webarchívum (FEWA) célja a mindenkor Fejér megyei tematikájú (Fejér vármegyei településekkel, személyekkel, intézményekkel, helytörténettel stb. foglalkozó) honlapok megőrzése. A munka első fázisaként székesfehérvári vonatkozású honlapok kerülnek archiválásra, majd a folyamatos munka során tovább bővítjük az archívumot a vármegyében levő településekkel kapcsolatos tartalmakkal.

A FEWA célja, hogy hosszú távon megőrizze és rendszerezett gyűjteményként tárja az érdeklődők elé a vármegyei webes tartalmakat, megőrizve azokat az utókornak, akkor is, amikor az eredeti oldal már nem elérhető. Maga a webarchívum általános gyűjtőkörű, tehát a mindenkor Fejér megyével kapcsolatos bármilyen webes tartalom a gyűjtemény részét képezheti. Első körben a székesfehérvári intézmények, szervezetek, turisztikai nevezetességek, illetve a helyi művészet köréből állítottunk össze válogatást. Minden egyes archivált weboldalt gondosan metaadatoltunk és tárgyszavaztunk a minél hatékonyabb kereshetőség érdekében. A weboldalak kereshetők cím és webcím (URL), téma, illetve a weboldalhoz kapcsolódó személyek, intézmények szerint, illetve a részletes keresés alkalmazásával számos egyéb paraméter szerint. A keresést könnyítve, minden a háttéradatbázisban tárolt kereshető metaadattípus a keresőmezőbe beírtaknak megfelelően megjelenik, áttekinthetővé és a keresés során kiválaszthatóvá téve az archívum tartalmait. Fontos megemlíteni, hogy nem minden megőrzésre kiválasztott weboldal archiválása sikeres, ugyanis előfordul, hogy minden erőfeszítésünk ellenére sem sikerül a webaratást végző szoftvereknek maradéktalanul lementeniük az adott honlapot. Ez a jelenség főleg a nem szabványos weboldalak archiválásánál, illetve az aratószoftver számára elérhetetlen médiafájlokat (pl. videókat) tartalmazó oldalaknál, fordul elő és körülbelül a honlapok harmadát érinti.

A projekt az NKA támogatásával (A2033/3933) valósult meg. A Fejér Megyei Webarchívum a Vörösmarty Mihály Könyvtár osztályainak – a munkát koordináló informatikusok, a Helyismereti és Digitalizálási Csoport, az Olvasószolgálat, az Állományalakító és Feldolgozó Csoport – szoros együttműködése révén születhetett meg. Szakmai támogatónk az Országos Széchényi Könyvtár Digitális Bölcsészeti Központjának Digitális Filológiai és Webarchiválási Osztálya.

A projektben érintett támogatási összegek

A projektre kapott támogatási összeg összesen nettó 940 000 forint volt. Ezek a következő táblázat szerint bonthatók kiadásokra.

tételek	nettó	áfa	bruttó
munkadíj (webfejlesztés)	285 000 Ft	- Ft	285 000 Ft
backup NAS (synology DS224+)	122 551 Ft	33 089 Ft	155 640 Ft
backup háttértár (2x12 TB WD)	175 575 Ft	47 405 Ft	222 980 Ft
szerver háttértár (2x 20TB seagate)	241 622 Ft	65 378 Ft	307 000 Ft
webtárhely (1év 1GB)	11 390 Ft	3 075 Ft	14 465 Ft
webtárhely (3év 6GB)	48 538 Ft	13 105 Ft	61 643 Ft
wordpress sablon	24 897 Ft	- Ft	24 897 Ft
wordpress bővítmények	34 320 Ft	- Ft	34 320 Ft
ÖSSZESEN	943 893 Ft	162 052 Ft	1 080 945 Ft

A felsorolt tételek alapvetően a szolgáltatási weboldal infrastrukturális kiépítését és fejlesztését, továbbá egy más forrásból beszerzett szervergép háttértárbővítését (nagysebességű SSD-k formájában), illetve a rendszer biztonsági másolását (backup) lehetővé tévő eszközök beszerzését szolgálta. A szolgáltatási weboldal fejlesztésére költött összegek magukban foglalták a külsős webfejlesztő munkadíját, a korszerű oldalhoz szükséges Wordpress kiegészítőket, illetve a webtárhely bérleti összegét. A backuprendszer szükséges óvintézkedés, egy esetleges adatvesztéssel járó esemény során teszi lehetővé a káresetet megelőző adatok helyreállítását. Ehhez egy új NAS (hálózati adattároló egység) üzembe állítása vált szükségessé, két nagykapacitású, gyors háttértárral kibővítve.

A webarchívum célja és gyűjtőköre

A FEWA célja, hogy a mindenkori Fejér megyei webes tartalmakat hosszú távon megőrizze az utókornak. E feladat érvényesítése érdekében a könyvtárakban általánosan alkalmazott helyismereti irányelvek mentén történik a webarchívum gyarapítása, vagyis – a hagyományos dokumentumokhoz hasonlóan – az archiválandó honlapok köre földrajzi és kapcsolódó egyéb tartalmi szempontok szerint alakul ki. Az első esetben a mindenkori Fejér megye (vármegye 1950-ig, megye 2023-ig, illetve a mai vármegye-rendszer) területén lévő települések hivatalos, üzleti, szórakoztató stb. témájú weboldalai, vagy az ezekről a településekről szóló egyéb honlapok képezik az archívum tárgyát (*földrajzi szempont*). A másik gyarapítási szempontot a mindenkori megye területén települt testületek, illetve ott született, vagy tevékenységük okán oda köthető személyek határozzák meg. (*személyi/testületi szempont*). Utóbbiak köre sokszor szakmai dilemma elé állítja a helyismereti könyvtárosokat, mert mindig mérlegelni kell, hogy egy az adott megyében született személy, aki például a fővárosban, vagy máshol tevékenykedett aktívan, annak helye van-e egy helyismereti gyűjteményben. Hasonlóképp kell mérlegelni egy máshol született és ideiglenesen a megyében tevékenykedő személyt, például egy a helyi sportegyesületnél játszó sportoló, vagy a városi szintársulatan néhány évig szereplő színész esetében. A FEWA esetében a személyek két szempont szerint kerülnek a gyűjteménybe. Első esetben az archiválandó weboldal tárgyát képezik (lásd részletesen a metaadatolásnál), vagy a weboldalak kiválasztása során tudatosan a fenti elveknek megfelelően keresünk olyan személyeket, akiknek van saját honlapjuk, vagy egyéb velük foglalkozó tartalom. Az első munkafázis során székesfehérvári művészek weboldalai kerültek ily módon

kiválasztásra. Ahogy a jövőben bővül majd a FEWA-ban archivált weboldalak mennyisége, a gyűjteménybe bekerülő helyi személyek körének meghatározása is minden bizonnyal felveti majd újra ezeket a kérdéseket. A harmadik állományalakítási szempont a mindenkor Fejér megyében történt események és rendezvények köre (*történeti szempont*).

A tartalmak gyűjtése 2024-től indul, de ez nem jelenti kizárólagosan csak új honlapok gyűjtését, jelenleg is számos már nem frissített weboldal került be a FEWA gyűjteményébe. Az induló évszám sokkal inkább jelzi a gyűjtés kezdetét és a mentések idejét. Az archiválás során felmérjük a weboldalak várható változékonyságát és ennek függvényében konfiguráljuk az aratószoftver paramétereit az automatikusan ismétlődő mentésekhez. Értelemszerűen a már nem frissített honlapokat elegendő egyszer lementeni.

Az aratószoftver konfigurációja

Mivel az Országos Széchényi Könyvtár fáradhatatlan munkával már kitaposta a webarchiválás gyakorlatának ösvényét, nekünk “csak” implementálnunk kellett a jógyakorlatokat. Aratáshoz az általánosan használt *Web Curator Tool* (továbbiakban WCT) szoftvert alkalmazzuk, amely a Heritrix3 robottal (crawler) végzi a mentéseket. Ez a robot egy beállított seedlistát (többé-kevésbé az archiválandó oldal url-je) kiindulási pontként használva járja be a lementendő oldal előre beállított paraméterek mentén. E robotot vezérlő paraméterek konfigurálása szolgál a WCT keretrendszer. Ezek az ún. profilok, amelyekkel különböző típusú webtartalmak lementése is testreszabható, mert például más beállítások szükségesek egyetlen dokumentum, vagy egy hírportálon feltüntetett egyetlen hír lementéséhez, mint egy teljes weboldaléhoz. Az alapértelmezett profil a mi esetünkben 1 GB-ban korlátozza a mentés adatmennyiségét, de ez felülírható nagy méretű honlapok (pl. önkormányzati portálok, intézményi honlapok) aratásánál. Az ilyen esetekben 5 GB-ra növeltük a lementendő adatmennyiséget. Az első aratásoknál törekszünk az archivált weboldalak teljes lementésére, addig ismételve az 1-5 GB-os mentéseket, míg az összes adattartalom lementésre nem kerül. Ezt követően az ún. auto-prune funkcióval az aratószoftver már csak az új tartalmakat fogja lementeni az ismétlődő mentési munkamenetek alkalmával. A pilotfázis során kizárólag weboldalakat mentettünk, így az eltérő típusú online dokumentumok mentésére szolgáló profilok kidolgozása még jövőbeni feladatunk.

Közös munkafelület

Munkavégzés szempontjából a FEWA két szempont szerint jellemezhető: egyrészt informatikai feladat, amely a hardverkomponensek és az arató/megjelenítőszoftverek telepítését és konfigurációját, az adatbázis és weboldal felépítését és működtetését takarja, illetve a könyvtári feldolgozó munkának tekinthető metaadatolást. Utóbbi elősegítése érdekében egy olyan munkaplatformra volt szükségünk, amely lehetővé teszi a közös szerkesztést. Mivel a Vörösmarty Mihály Könyvtár nem sokkal ezt megelőzően tért át a Microsoft O365 felhőalapú munkakörnyezetre, a közösen szerkeszthető online Excel táblázatkezelő minden különösebb gond nélkül lehetővé tette a csoportos munkát. A munkafolyamat koncepciója a következő: az OSZK gyakorlata alapján meghatározott metaadatokat oszlopokként táblázatban rögzítjük, majd az adatok véglegesítését követően a webarchívum háttéradatbázisába exportáljuk. Így megkerülhető az a problémaforrás, hogy

laikusoknak adatbáziskezelőben kelljen dolgoznia, a munkafolyamatban mindenki a saját szakterületére összpontosíthat: a feldolgozó és helyismereti könyvtárosok végzik a tartalmi feltárást, a webarchívum webkurátorai gondoskodnak az egyéb metaadatokról, illetve koordinálják a folyamatot, az informatikai csapat pedig a működtetést és az IT-támogatást szavatolják. Természetesen előfordul olyan, hogy egy szakember több területen is jártas, így feldolgozóként, webkurátorként, illetve bizonyos feladatokban akár informatikusként is helyt állhat.

Az online Excelben minden sor egy weboldal adatait tartalmazza. A közös szerkesztés révén a feldolgozó munka könnyen felosztható a FEWA-stáb tagjai között mind a weboldalak, mind az adattípusok tekintetében. Az Excelbe számos segédtablázat is beilleszthető (példaként lásd a tartalmi feltárást részletes tárgyalásánál), a bevitt adatok CSV formátumban könnyedén exportálhatók, lehetővé téve az adatbázisexportot.

Metaadatolás kialakítása

Webarchívumként a FEWA legfőbb egyedisége a tradicionális könyvtári feldolgozó munkán alapuló részletes tartalmi feltárást. Ennek lényege, hogy az archívumba bekerülő weboldalak a katalogizált könyvtári dokumentumokéhoz mérhető formai és tartalmi feltárást részesülnek. A rögzítendő adatok kiválasztásában is az OSZK-MIA gyakorlatát követtük, a fő különbség a két webarchívum metaadatolása között, hogy a FEWA-ban nem gyűjteményként, hanem egyedileg dolgozzuk fel a weboldalakat.

Entitásközpontú adatmodell (IFLA LRM alapelvek)

A feldolgozás során az entitásközpontú feldolgozásra törekedtünk. Ennek lényege, hogy a feldolgozás során bizonyos adattípusokat önálló entitásként tárolunk és kezelünk, túltekintve a webarchívum metaadatrétegén, szem előtt tartva nem csupán a megyei könyvtár, hanem a teljes könyvtári-közgyűjteményi rendszer, információs közegét. Az IFLA könyvtári referenciamodellje (IFLA LRM) a világot leíró különböző entitásokat a következő módon határozza meg:

E vezérelv miatt a metaadat-típusainkat az ebbe a közegbe illeszkedő egyedi azonosítóval látjuk el, illetve amit lehet, megfeleltetünk az országos, globális rendszereknek. Az általunk használt entitások és kapcsolatrendszerük részletes ismertetése a háttéradatbázis című fejezetben olvasható.

Formai feltárást kialakítása

A formai feltárást során az egyes weboldalak azonosíthatóságához szükséges, illetve az előállításra vonatkozó, adatokat rögzítjük. Egy még gyűjteménybe nem rendezett weboldal azonosíthatóságát a különböző címadatok teszik lehetővé. Webes környezetben az egyik legfontosabb címadat a webcím, vagyis az oldal URL-je. Meg kell azonban említeni, hogy egy weboldal életciklusa során többször is változhat az URL (például domainváltásnál), és persze a honlap megszűnésekor a webcím is eltűnik, sőt egy másik weboldalnak is kioszthatják. Azon ritka esetekben, amikor a webhely webcíme permanens URL, akkor már valódi azonosítóról, tehát URI-ról beszélhetünk. Az ilyen webcímek jellemzően a webes információs rendszerek (pl. különböző wikik, közgyűjteményi információs rendszerek stb.) sajátosságai.

A FEWA esetében az archivált oldalak webcímét az *Eredeti URL* adatmezőben rögzítjük. Az URL mellett a weboldalak is rendelkeznek hagyományos értelemben vett címadatokkal, azonban a klasszikus könyvtári dokumentumokkal ellentétben (ahol, például könyvek esetében kell lennie dedikált főcímnak, esetleg alcímnak és egyéb címadatoknak) kevésbé explicit és egyértelmű módon vannak feltüntetve. Webhelyek címadatainak meghatározása esetén az oldal HTML-kódjában a fejlécben (head) a *title* tagek között megadott címadatokat (ámbar előfordul, hogy ezt üresen hagyják), illetve magában a honlaptörzsben (body) címsorként (*h1*, *h2*, stb.) megjelenő karakterláncokat vehetjük figyelembe. Némi nehézséget jelent azonban, hogy ezek jellemzően nem hagyományos értelemben vett címek, címszerűségek, leginkább a honlap elsődleges témájának a tartalomtól vizuálisan jól elkülöníthető feltüntetett megjelenítései. Tehát az a testület, személy, rendezvény vagy specifikus téma, amiről a weboldal szól, amit népszerűsít, vagy ismertet. Tehát karakterek szintjén a webhely címszerűsége megegyezik a tartalmi feltárás során megadott entitásával. A FEWA esetében mind a *title* mind a *h1* címsor esetében a feltüntetett cím a *Fejér Vármegyei Webarchívum*, ugyanakkor a témája is a VMK azonos nevű szolgáltatása, amelyet érdemes testületként önálló entitásként is rögzíteni. A könyvtár információs környezetében ilyen esetben tehát van két jelentősen eltérő hozzáférési pont (egy nomen és egy ágens), amely egy jól kidolgozott rendszerben különböző azonosítókkal bír, tehát rendszerszinten elkülönülnek egymástól, azonban karakterlánc szintjén mégis megegyeznek. Annak érdekében, hogy ez a karakteregyezőség ne zavarja össze a felhasználókat, az *Egységesített cím* adatmezőben a honlapra vonatkozó dedikált címváltozatot készítettünk az “XY honlapja” mintát követve. A tágabb bibliográfiai univerzumban is ez az egyik elsődleges hozzáférési pont a webarchívumban feldolgozott weboldalakhoz. Az egységesített cím mellett a honlapon (vagyis a HTML-kódban) a fenti módon feltüntetett címadatot is rögzítjük az *Eredeti főcím* adatmezőben. Amennyiben további címszerűség található a honlapon – például alcím, vagy valamiféle címszerű kiegészítő adat – akkor azok az *Alcím/egyéb* címadat mezőbe kerülnek.

A *Leírás forrása* mezőben az aktív weboldal metaadatolásának forrásait, jellemzően magát az aktív weboldalt, írjuk le. A *Leírás dátuma* a metaadatolás időpontját, és egyúttal az aktív weboldal e dátum szerinti állapotát adja meg. Amennyiben ismert, az e célra fenntartott adatmezőkben, rögzítjük a weboldal, vagy az első bejegyzés, létrejöttének, illetve a honlap megszűntének, vagy az utolsó frissítés (amennyiben az legalább 2023-as) dátumát. Entitáskapcsolatként rögzítjük a weboldalt előállító és a tartalom létrehozásában közreműködő személyeket és testületeket, illetve a weboldalak tartalmi jogtulajdonosait. Az *Archiváló intézmény* minden esetben a Vörösmarty Mihály Könyvtár. A honlapok nyelvét, több ízben nyelveit, is entitáskapcsolatként regisztráljuk. Ehhez kontrollált szótárként ISO nyelvi kódokat használunk, hozzácsatolt természetes nyelvű megnevezésekkel. A weboldalak változékonyságát saját elemkészletű szótár segítségével határozzuk meg (lásd a háttéradatbázisnál). A sztenderd adatmezőkbe nem megadható további információk rögzítésére megjegyzésmezőket tartunk fent.

Tartalmi feltárás, tárgyszórendszer kialakítása

A FEWA esetében cél volt az igényes, tartalom szerinti kereshetőség megteremtése. A tartalmi feltárásnál alkalmazott adatmezők alapját is az OSZK gyakorlata képezte, azonban jóval

szabványosabb, kontrollált módon, célirányosan úgy tervezve, hogy illeszkedhessen egy nagyobb szabású, szabványos információs környezetbe.

Tartalmi feltárás elképzelhetetlen kontrollált szótár, tárgyszójegyzék nélkül. Magyar nyelvben mindössze kettő ilyen rendszer jöhet számításba: a magyar közkönyvtári használatra tervezett és folyamatosan aktualizált, ámbár digitális környezetben nagyon nehezen használható és számos helyen inkonzisztens Köztaurusz, vagy a Library of Congress tárgyszórendszerének (LCSH) magyar fordítása. Előbbi mellett szól, hogy szélesebb körben elterjedt és a magyar kulturális viszonyokra jobban alkalmazható, rugalmasabban bővíthető, és nem utolsósorban, a VMK-ban is eleve ezt használják tárgyszavazásra. Az LCSH jobban integrálható jelen állapotában digitális környezetbe, viszont az angolszász kultúrkört veszi alapul, és szabványos módon, hazai igények mentén nem bővíthető.

A választás tehát a Köztauruszra esett, amelynek adatbázisba integrálása komoly előkészítő munkát igényelt. A Köztaurusz sajátossága, hogy bár teljes mértékben elérhető, sőt szerkeszthető elektronikusan a Relexen keresztül, ám a tezausz maga sajnos nem lépett túl a tezauszmfaj papíralapú korlátaiban. Ez gyakorlatilag azt jelenti, hogy a tezausz lexikai egységeit egyedi azonosító helyett maga a deszkriptor, vagyis az adott fogalmat leíró tezauszcikk előnyben részesített névalakja, szolgál a fogalom azonosítója gyanánt, és ezen deszkriptorok, egy régebbi kísérleti - már elavult Köztaurusz-változatot alapul vevő - OSZK-s SKOS-projekt kivételével nem rendelkeznek URI-val, vagy bármilyen korszerű információs ökoszisztémában használható azonosítórendszerrel. Emiatt nem használhattuk közvetlen módon a Relexet a tartalmi feltárásra, a rendszerszintű szabványosítás követelményei igénylik a lexikai egységek számítógépes környezetben való azonosíthatóságát, tehát szükség van egy korszerű, azonosítóalapú kontrollált szótárra a Köztaurusz bázisán. Ezért a Köztaurusz lexikai egységeit és a közöttük lévő kapcsolatrendszert adoptálva egyfajta sajátos osztályozási és tárgyszórendszert, egyfajta “alkalmazásprofil” készítettünk a közkönyvtári tezauszból. Ennek lényege, hogy a Köztaurusz több mint 80 ezres fogalomkészletéből kiválogattuk a helyismereti munkához hasznos lexikai egységeket, és azok részleges kapcsolatrendszerét. A cél nem a Köztaurusz kiváltása más rendszerrel, hanem a tezausz tárgyszavainak lokális alkalmazása kiterjesztett adatkapcsolati lehetőségekkel, előre betervezve egy tágabb kapcsolattal-környezetbe (linked data) való becsatlakozás lehetőségét. Ez utóbbi miatt a kiválasztott Köztaurusz-tárgyszavak mindegyike kapott egy a VMK saját online ökoszisztémáját jelölő egyedi azonosítót, illetve a háttéradatbázisban rögzített lexikai egységek mindegyék explicit módon megfeleltették a Köztaurusz-deszkriptornak, még abban az esetben is, ha a VMK-s tárgyszóalkalmazás és a deszkriptor névalakja megegyezik. Így szavatoljuk a Köztaurusszal való teljes kompatibilitást, ugyanakkor időnként rugalmas módon változtathatunk magán a jobban érthetőség kedvéért a tárgyszó névalakján, illetve bizonyos esetekben egyszerűsíthetjük a fogalmak közötti kapcsolatrendszert. Például a koncert KT-deszkriptora a *hangverseny*, de a köznapi használatban ide jobban illeszkedik előnyben részesített névalakként a *koncert*, különösen, hogy a könnyűzenei koncert ennek alárendelt fogalma, és hát, valljuk be, elég furcsa a hangverseny alá besorolni mondjuk egy rockkoncertet.

A teljes tartalmi feltárásra alkalmas saját tárgyszórendszer apropóját az OSZK gyakorlata adta. Ők a gyűjtemények tartalmi feltárását illetően a felső szintű rendezési elveket (vagyis a

tartalmak osztályozását) három szintben (kategóriák vagy főtémacörök, témacörök és altémacörök) formájában érvényesítik, saját megnevezéseket használva. Osztályozási/rendezési alapelvnek mi is ezt választottuk, azonban saját, “házi szótár” helyett mi megfeleltettük az OSZK-MIÁ-s osztályozási kifejezéseket a hierarchiában magasabb szinten megtalálható Köztaurusz ekvivalensekkel, így kontrollált lexikai egységeket használhattunk erre a célra is. Alkalmazásprofilként nincs szükség egzakt csúcsfogalomra és az onnan kiinduló teljes hierarchiára, így az egyes kategóriák lettek a VMK-s tárgyszócsoportok csúcsfogalmai, a webarchívum tartalmi feltárásához szükséges, témacör, altémacör és tárgyszó feladatkörben alkalmazott lexikai egységek ezen kategóriák alá vannak kiválasztva és besorolva, nagyrészt megtartva ezek köztauruszos hierarchikus kapcsolatrendszerét. Miért írjuk, hogy nagyrészt? Leginkább azért, mert a rendelkezésre állónál jóval kevesebb lexikai egységet használunk, így vannak olyan számunkra felesleges kapcsolatok, amelyek a mi alkalmazásunk szempontjából feleslegesek. Ebből következően lazítottunk a Köztaurusz szigorú, háromtengelyű¹ kapcsolatrendszerén, az alkalmazott tárgyszavakat egy a nyugati módszertanban gyakran alkalmazott laza hierarchiára cseréltük. A tárgyszavak relatíve alacsony példányszámon tartása kiemelt cél volt, ezért kényszerültünk a hierarchia alsóbb szintjein szakítani a szigorúsággal, és a nekünk “kellő” tárgyszavakat bármilyen minőségű kapcsolat mentén már hozzákapcsoltuk az egyes kategóriákból kiinduló hierarchiába. A feldolgozó munkatársak munkájának megkönnyítésére a munkára kiválasztott tárgyszavak körét az új laza hierarchia mentén olyan Excel-táblába rendeztük, amely adatérvényesítés alkalmazásával oly módon segítette elő a szabványos tárgyszavazást, hogy a feldolgozó Excel-táblában a tárgyszavakra korlátozta le a bevitelre alkalmas adatok körét. Ezzel kizártuk a nem sztenderd tárgyszavak bevitelének lehetőségét. Amennyiben a weboldalak feldolgozása során akár a Relexből, akár a helyismereti gyakorlatból kiindulva új tárgyszóra volt szükség, első ízben azokat a VMK-s hierarchiába illesztettük be. Ezzel a módszerrel hoztunk létre a gigantikus Köztaurusz alapján egy 7 kategóriából, 39 témacörből, 76 altémacörből és 3153 tárgyszóból² álló olyan könnyebben kezelhető tárgyszórendszert, amely alkalmas egy helyismereti webarchívum tartalmi feltárására oly módon, hogy lexikai egységek és bizonyos kapcsolatok mentén (elismerve hogy a teljeskörű tezaurusz-kapcsolatrendszerből “fel kellett áldozni” valamennyit) csereszabatos vele és meg is tartja a Köztaurusz szabványosságát. Bónuszként ehhez társul hozzá a lexikai egységekhez csatolt egyedi azonosítók rendszere. Itt kell megemlíteni, hogy a tartalmi feltárás a könyvtári feldolgozó munka talán legszubjektívebb szegmense, ezért minél mélyebb a feltárás, annál több inkonzisztencia jelentkezhet a tartalmi feltárás céljait, tehát hogy adott tárgyszó az adott témához kapcsolódó összes dokumentumhoz hozzáférési pontot biztosítson, illetően. Ezt a fajta inkonzisztenciát szintén csökkenti a leszűkített tárgyszókészlet, hiszen egy háromezres tárgyszórendszer ilyen szempontból jóval előnyösebb, mint a 80 ezres Köztaurusz.³

¹ hierarchia (A-F), része-egésze (P-T), előfeltétel - rezultáns (E-R)

² a projekt első fázisának lezárása idején

³ Meg kell említeni, hogy a Köztaurusz azért tartalmaz ennyi lexikai egységet, mert a tárgyszórendszeren túlmenően köznévtéri feladatot is ellát, tehát a magyar nyelv általánosan használt szavaihoz kíván kontrollált szótárt adni.

A jövőben, összhangban a Vörösmarty Mihály Könyvtár készülő új stratégiájának szabványosítási céljaival, ezen “alkalmazásprofilszerű” tárgyszórendszerből készül majd a finn nemzeti könyvtár fejlesztette Skosmos szótárkezelő szoftver segítségével egy minden tekintetben szabványos kontrollált szótár, amelyben a tervek szerint nem csak a saját használatra szánt laza hierarchia lesz szoftveresen rögzítve, hanem a VMK által használt tárgyszavak Köztaurusz-ekvivalenseinek kapcsolatrendszere is a deskriptorok feltüntetésével. A VMK lokális tárgyszórendszerének szabványos, korszerű linked-data kifejlesztése tehát egyúttal a teljes Köztaurusz, a rendszert alkalmazókkal együttműködésben történő, korszerű alapokra hozásának előfutára is lehet.

A tárgyszórendszerrel feltárt *téma* (kategória, témakör, altémakör, tárgyszavak) mezőkön túlmenően a FEWA entitáskapcsolatokkal is bővíti a weboldalak tartalmára vonatkozó információkat, tehát, a könyvtári feldolgozó gyakorlatnak megfelelően, az egyes oldalak tárgyaként *kapcsolódó személyek, testületek, rendezvények és földrajzi helyek* is hozzáférési pontként szolgálnak. Az adatmodellnek megfelelően ezek az entitások olykor eltérő funkcióban is szerepelhetnek, tehát a *Vörösmarty Mihály Könyvtár* testület a leírásban szerepelhet tárgyként, létrehozóként, közreműködőként, jogtulajdonosként, sőt archiváló intézményként is. Az adatmodell erre vonatkozó részletei a háttéradatbázis-kapcsolatoknál, illetve a keresőfelület ismertetőjében olvashatók.

Háttéradatbázis (db_fewa)

A FEWA háttéradatbázisa (*db_fewa*) egy Ubuntu szerveroldali operációs rendszerre telepített, MySQL lekérdezőnyelvet használó open source MariaDB adatbázis, amely a közös nyelv révén összekötésben áll a WCT háttéradatbázisával. Az alkalmazott karakterkódolás UTF-8 (utf8mb4). Az adattípusok szöveges karakterlánc esetében CHAR(255) vagy VARCHAR(255), azonosítók esetében INTEGER(11), dátumok esetében DATETIME, míg kétállású (IGAZ = 1, HAMIS = 0) adatmezők esetén TINYINT. Utóbbiak használatával tudjuk például egyfajta kapcsolóként beállítani az archivált weboldalak jogi státuszát, nulla értéknél a felhasználó csak a VMK-s IP-tartományban fér hozzá az adott oldalakhoz. Hasonlóképp jelöljük az archivált oldalak minőségellenőrzési státuszát, vagy a weboldalon kiemelten ajánlott oldalak körét. A táblák belső azonosítói (id) elsődleges kulcsokként működnek, és az AUTO_INCREMENT funkcióval automatikusan kerülnek kiosztásra.

Maga a háttéradatbázis hármas szerepet lát el, egyrészt adattáblái és kapcsolatai révén leképeződik maga a webarchívum, rugalmas és bármikor módosítható módon tárolja az annak metaadatait, illetve ezeket kamatoztatva kiszolgálja az összes szükséges adattal a szolgáltatás weboldalát.

Entitások

Az adatmodell entitásaiból épül fel a háttéradatbázis adattábláinak többsége. A relációs adatbázisok kontextusában fontos megkülönböztetni az IFLA LRM-ben is használt entitás fogalmát, a hazai informatikai szakzsargonban használt egyedétől, mivel angol nyelvterületen mindkettő az entity-relationship elnevezést használja és mindkettő eredete az entity-relationship (egyed-kapcsolat) modellben keresendő. A relációs adatbázisban az adattáblákban tárolják az egyedekre vonatkozó adatokat (pontosabban az összetartozó egyedek alkotnak

egyed típusokat, így egy adattábla egy adott egyed típus adatait tartalmazza), a táblák adatmezőiben pedig az egyedek tulajdonságait leíró attribútumok találhatók. Az adatokkal konkretizált egyedeket nevezzük egyed-előfordulásnak. Például a *Vörösmarty Mihály Könyvtár* egyedként a *testület* egyed típus egyik előfordulása, és annak különböző névváltozatai, azonosítói, egyéb adatai alkotják az attribútumokat (egyed tulajdonságokat). Ezen tulajdonságok között lehet különféle kapcsolatokat (relációkat) képezni és ily módon az adatokat hatékonyan tárolni és lekérdezni. Ezzel összhangban, ha az IFLA-LRM-et relációs adatbázisban kívánjuk leképezni, akkor a modell entitásai egyedekké, míg azok ismérvei attribútumokká válnak.⁴ Mi is nagyrészt erre törekedtünk, felismerve a teljeskörű gyakorlati LRM-implementáció nehézségeit, nevezetesen a nomen entitások egyedként való használatát. Mivel azok jellemzően az egyes entitások neveiként, azonosítóiként fordulnak elő, a jól bevált és megszokott adatbázisgyakorlatban ezek az egyedként jelentkező entitások attribútumaivá válnak, vagyis minden egyes névváltozatot, külső és belső azonosítót, sőt magukat az egyedek neveit is önálló adattáblában kellene tárolni, megtöbbszörözve ezzel az eleve komplex adatbázis bonyolultságát, sokszorosára növelve az adatbázisépítés során bekövetkező hibák lehetőségét. A mi adatbázisunkban tehát minden könyvtáros értelemben vett entitás egyed, de nem minden egyed entitás. Lássuk az egyes entitásokat és azok a *db_fewa* adatbázisban való gyakorlati megvalósítását. A következő táblázatokban az egyes adattáblák elsődleges kulcsait félkövér sötétkék, míg a táblák közötti kapcsolatokat lehetővé tevő másodlagos kulcsokat zöld betűszínnel jelöltük.

Weboldalak (*tbl_fewa*)

Mivel a FEWA-ban csak webtartalmakat tartunk nyilván, nem használjuk a művekre és azok megjelenési és kifejezési formáira vonatkozó magentitásokat, a webtartalmakra egyetlen (integráló) műtípust alkalmazunk entitástípusként. Ezek egyedei találhatóak a *tbl_fewa* adattáblában. Gyakorlatilag ez a háttéradatbázis központi táblája, szinte minden más táblával össze van kapcsolva az adatbázisban. Attribútumai a következők:

tbl_fewa attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
standard_fewa_id	VMK azonosító		nomentípus
source_url	aktív oldal URL-je		nomentípus
title_statement	honlapon feltüntetett címszerűség		nomentípus
uniform_title_id	egységesített cím belső azonosítója		
var_form_title	egyéb címadat, címváltozat		nomentípus

⁴ Angol nyelven teljes egészében megegyezik az absztrakt könyvtudományi adatmodell és a gyakorlati adatbázisépítés nyelvezete. A magyar nyelvű eltérés hátterének oka az, hogy Magyarországon sajnos egymástól függetlenül alakult ki a két tudományterület szaknyelve.

showcase	szolgáltatási felületen ajánlott oldal (INTEGER)		
rights_status	jogi státusz, van-e engedély nyilvános közlésre (INTEGER)		
rights_note	megjegyzés az engedélyezésről		
variability_id	változékonyság belső azonosítója		
qa_status	minőségellenőrzés státusza		
qa_note	minőséggel kapcs. megjegyzés		
webpage_form_id	weboldal műfaja belső azonosító		
webpage_type_id	weboldal típusa belső azonosító		
webpage_startdate	oldal létrejöttének v. az első bejegyzés dátuma		időtartamtípus
webpage_enddate	oldal megszűnésének v. az utolsó bejegyzés dátuma		időtartamtípus
cataloging_source	leírás forrása		
description_date	leírás dátuma		
description_note	megjegyzés a leírásról		
general_note	általános megjegyzés		
abstract	absztrakt		
history_note	történeti megjegyzés		
provenance	proveniencia		
archivator_id	archiváló intézmény belső azonosítója		
wct_target_id	WCT háttéradatbázis target ID		
lrm_type_id	LRM megfeleltetés belső azonosítója		

Gyűjtemények (tbl_collection)

A *tbl_collection* nevű táblában helyet kapó *gyűjtemények* a weboldalak valamilyen közös szempont szerinti halmazát alkotják. Jelen pillanatban még nem készítettünk, de terveink között szerepel, hogy a könyvtárunkat érintő online sajtómegjelenéseket archiváljuk és ezeket egy gyűjteménybe szervezzük. Ebben az esetben a *VMK sajtómegjelenés* lesz a *gyűjtemények* egyik egyede. Az archívum bővülésével tetszőleges módon alakíthatunk ki különböző gyűjteményeket az archivált weboldalakból. Attribútumai a következők:

tbl_collection attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
standard_fewa_id	VMK azonosító		nomentípus
access_url	gyűjtemény URL-je		nomentípus
collection_title	gyűjtemény címe		nomentípus
general_note	általános megjegyzés		
abstract	absztrakt		
history_note	történeti megjegyzés		

Személyek (tbl_person)

A feldolgozott weboldalakhoz valamilyen módon kapcsolódó személyeket érdemes önálló, könyvtári értelemben vett entitásként kezelni, mivel így a rájuk vonatkozó adatokat egységesen, normalizált formában tudjuk felhasználni, gyakorlatilag az adatbázison belüli vagy kívüli authorityállományként kamatoztathatjuk ezeket az adatokat. Ez azzal az előnnyel jár, hogy az adatbázisban tárolt személyek a VMK-s és külső azonosítók révén megfeleltethetők a katalógusunk, vagy bármely szabványos azonosítókat használó külső authorityállomány, névtér személynév-állományával. Az ily módon összekötött információs rendszerek az alapkövei a több forrásból dolgozó, ún. discovery jellegű keresőrendszerek kialakításának.

A FEWA háttéradatbázisában a személyek entitástípus egyedeit tárolja a *tbl_person* adattábla. Az önálló adategységként kezelt személyek többféle szerep (tárgy, létrehozó, közreműködő és jogtulajdonos) szerint kapcsolódhatnak a főtáblához. Ezen kapcsolatok részletei a következő fejezetben olvashatók. A *személyek* attribútumai megfeleltethetők a MARC21 authority formátumában leírtaknak, esetünkben az adott személyhez kapcsolódó különböző névváltozatokat, titulusokat és szöveges értelmezőket, illetve vonatkozó weboldalak és külső azonosítók linkjeit tartalmazza. Az attribútumok részletesen a következők:

tbl_person attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
std_personal_name_id	VMK azonosító		nomentípus

personal_name	személynév (előnyben részeshített)		nomentípus
pname_fuller_form	A személynév teljes alakja		nomentípus
pname_numeration	uralkodói számozás		
pname_title	titulus, foglalkozás, egyéb szöveges információ		
pname_date	kapcsolódó dátum, évszám (ált. születési és halálozási év)		időtartamtípus
website	kapcsolódó vagy saját weboldal		nomentípus
fmel	Fejér Megyei Életrajzi Lexikon linkje		nomentípus
isni	ISNI azonosító		nomentípus
wikidata	Wikidata azonosító		nomentípus
nn	Nemzeti Névtér azonosító		nomentípus
other_uri	egyéb külső forrás URI-ja		nomentípus

Testületek (tbl_corporate)

A *testületek* egyedeit tároljuk a *tbl_corporate* táblában. A személyekhez hasonlóan eltérő szerepek szerint kapcsolódnak a weboldalakat tartalmazó főtáblához és a tábla felépítése, MARC 21 kompatibilitása is hasonló, a testület nevét és alárendelt egységét, kapcsolódó dátumadatokat, weboldal és azonosítólinkeket tartalmaz. A részletes attribútumok a következők:

tbl_corporate attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
std_corporate_ name_id	VMK azonosító		nomentípus
corporate_name	testületi név (előnyben részeshített)		nomentípus

subordinate_unit	alárendelt szervezeti egység		nomentípus
cname_date	kapcsolódó dátum, évszám (ált. születési és halálozási év)		időtartamtípus
website	kapcsolódó vagy saját weboldal		nomentípus
nn	Nemzeti Névtér azonosító		nomentípus
wikidata	Wikidata azonosító		nomentípus
viaf	VIAF azonosító		nomentípus
other_uri	egyéb külső forrás URI-ja		nomentípus
wct_agency_id	VCT háttéradatbázis agency ID		

Rendezvények (tbl_meeting)

A könyvtári feldolgozó munkában bevett harmadik szokásos “bibliográfiai entitás” a személyek és testületek mellett a rendezvények. Rendezvények alatt tág értelemben vett eseményeket értünk, közös nevező bennük, hogy mind rendelkeznek valamiféle címmel vagy megnevezéssel, kapcsolódó helyszínnel és időtartammal. Részletes attribútumok a következők:

tbl_meeting attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
std_meeting_ name_id	VMK azonosító		nomentípus
meeting_name	rendezvény neve (előnyben részesített)		nomentípus
number	sorszám		
mname_date	kapcsolódó dátum, évszám (ált. születési és halálozási év)		időtartamtípus
website	kapcsolódó vagy saját weboldal		nomentípus
wikidata	Wikidata azonosító		nomentípus
viaf	VIAF azonosító		nomentípus

other_uri	egyéb külső forrás URI-ja		nomentípus
-----------	------------------------------	--	------------

Földrajzi helyek (tbl_geo)

Az ágensekhez hasonlóan az egyes weboldalakhoz kapcsolódó helyek rögzítése is fontos feladat. Amennyiben a földrajzi helyeket is önálló entitásként kezeljük, úgy alkalmassá válnak, az archívumtól függetlenül, más entitások földrajzi kapcsolatainak, például a testületek székhelyeihez, rendezvények helyszíneihez stb. leírására is. A számos külső azonosító mellett a földrajzi neveket a Köztaurusz földrajzi deszkriptorainak is megfeleltetjük. A *tbl_geo* attribútumai a következők:

Egységesített címek (tbl_uniform_titles)

A formai feltárás fejezetben már ismertettük a FEWA kontextusában az *egységesített címek* alkalmazását. Amit ehhez még szükséges hozzátenni az a standard címadáson túlmutató felhasználás lehetősége. Ez alatt annak lehetőségét értjük, hogy, bár jelenleg még nincs ez implementálva a gyakorlatban, az *egységesített címek* önálló, nomen entitásként való kezelése révén akár más weblapok, akár bármilyen dokumentum (könyv, folyóiratcikk, műalkotás, fotó, stb.) tárgyként rögzíthető egy weboldal leírásában. Ezzel megoldható például adott művekkel foglalkozó weboldalak pontos leírása. E lehetőség szem előtt tartása miatt kerültek, kvázi nomen implementációként, az *egységesített címek* egyedei a *tbl_uniform_titles* táblába. Attribútumai a következők:

tbl_uniform_titles attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
uniform_title	egységesített cím		
other_uri	egyéb külső forrás URI-ja		nomentípus
viaf	VIAF azonosító		nomentípus
oszk	OSZK Nektár azonosító		nomentípus
standard_uri	VMK URI (tervezett szabványos szótárhoz)		nomentípus

Tárgyi entitások (kategóriák, témakörök, altémakörök, tárgyszavak)

A tartalmi feltárás ismertetésénél írtuk, hogy a magasabb szintű, általánosabb tartalmi feltárás (osztályozás) és a specifikusabb (tárgyszavazás) során egységesen a Köztaurusz lexikai egységeit használtuk. A négy adattábla azonos felépítéssel készült, a különbséget a tárgyi hierarchia egyre bővülő egyedszámú típusai képezik a kategóriától a tárgyszavakig. A *kategóriák* egyedeit tartalmazó *tbl_subject_categories* attribútumai a következők:

tbl_subject_categories attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		

std_subject_categories_id	VMK azonosító		nomentípus
subject_category	kategória		nomentípus
category_koztaurusz	Köztaurusz ekvivalens		nomentípus
standard_uri	VMK URI (tervezett szabványos szótárhoz)		nomentípus

A témakörök (tbl_subject_themes) egyedtípus attribútumai:

tbl_subject_themes attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
std_theme_id	VMK azonosító		nomentípus
subject_theme	témakör		nomentípus
theme_koztaurusz	Köztaurusz ekvivalens		nomentípus
standard_uri	VMK URI (tervezett szabványos szótárhoz)		nomentípus

Az altémakörök (tbl_subject_subthemes) egyedtípus attribútumai:

tbl_subject_subthemes attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
std_standard_subtheme_id	VMK azonosító		nomentípus
subject_subtheme	témakör		nomentípus
subtheme_koztaurusz	Köztaurusz ekvivalens		nomentípus
standard_uri	VMK URI (tervezett szabványos szótárhoz)		nomentípus

A tárgyszavak (tbl_subject_headings) egyedtípus attribútumai:

tbl_subject_headings attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
--	-------------------------	----------------------	---------------------

id	belső azonosító		
std_standard_subject_heading_id	VMK azonosító		nomentípus
subject_heading	témakör		nomentípus
headings_koztaurusz	Köztaurusz ekvivalens		nomentípus
standard_uri	VMK URI (tervezett szabványos szótárhoz)		nomentípus

Weboldalak műfaja és típusa (tbl_webpage_forms és tbl_webpage_types)

Az OSZK-MIA gyakorlatát átvéve kategorizáltuk az archivált weboldalak műfaját és típusát. A műfaji meghatározáshoz alkalmazott értékek a következők: *Általános honlap, Elektronikus periodika, Blog, Mikroblog, Fórum, Közösségi oldal, Ismerettár, wiki, Hirdetési oldal, Adatbázis, Digitális gyűjtemény, Virtuális kiállítás*. Az adattábla attribútumai a következők:

tbl_webpage_forms attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
form	műfaj		

A weboldaltípus meghatározásához alkalmazott értékek: *Kormányzati oldal, Intézményi oldal, Egyéb szervezeti oldal, Üzleti oldal, Magánoldal, Közösségi oldal*. Attribútumok:

tbl_webpage_types attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
type	típus		

Weboldalak változékonysága (tbl_webpage_variability)

Az aktív weboldalak változékonyságának jelölésére saját szótárt használtunk. Ennek értékei: *Naponta többszöri, Napi, Heti, Havi, Negyedévenként vagy néhány havonta, Félévenként, Rendszertelen, Évenként, Ritkán változik, Soha, vagy rendkívüli ritkán változik*. Ez a fajta kategorizálás azonban a gyakorlatban nehezen megoldható, hiszen a feltárás során a feldolgozók egy alkalommal látják a weboldalt, és a pillanatnyi állapotból kell megállapítani annak változékonyságát. Éppen ezért jelenleg az archívumba kerülő oldalak többsége a rendszertelen besorolást kapta. A változékonyság pontosítása már webkurátori feladat, amelynek precíz ellátása az archívum mennyiségbeli növekedésével már egyre nagyobb kihívást jelent. A tbl_webpage_variability attribútumai a következők:

tbl_webpage_variability attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
variability	változékonyság		

Nyelvek (tbl_language)

Mivel még Fejér megyei kontextusban is elég gyakori egy weboldal többnyelvűsége, ezért a leírásban az oldalak nyelvét is rögzíteni kell. A nyelvek szabványos azonosítására háromkarakteres ISO nyelvi kódokat használunk, míg a felhasználói felületen a magyar nyelvű szöveges megnevezést használunk. A *tbl_language* attribútumai a következők:

tbl_language attribútum mezőnév	megnevezés / funkció	MARC21 ekvivalens	IFLA-LRM ekvivalens
id	belső azonosító		
language	ISO nyelvi kód		
language_caption	nyelv szöveges megnevezése		

Kapcsolatok és kapcsolótáblák

A relációs adatbázisok előnye a táblákban tárolt adatok közötti kapcsolatok tetszőleges kifejezése, ami lehetővé teszi nagyszámú adat hatékony tárolását. Egy bizonyos mennyiségig lehet Excelhez hasonló táblázatkezelőben nyilvántartani egy webarchívum háttéradatait, ám egy bizonyos mennyiség fölött már korlátozó tényező lesz a táblázatkezelő kapacitása és az egyes adattípusok közötti bonyolultabb kapcsolatrendszer kialakítása. Bibliográfiai-könyvtári információs környezetben pedig alapkövetelmény az adattípusok közötti kapcsolatok hatékony kezelése és kifejezése.

Egyszerű kapcsolatok

A FEWA háttéradatbázis központi eleme az archívumba bekerülő weboldalakat nyilvántartó tábla (*tbl_fewa*). A weboldalaknál előfordulnak olyan kapcsolatok, amikor egyetlen webdalegységhez egy másik táblából csupán egy másik egyed kapcsolódik. Ebben az esetben a *tbl_fewa*-ban egy másodlagos kulcsokat (a gyakorlatban a kapcsolódó tábla belső azonosítóit) tartalmazó mező kapcsolódik a másik tábla elsődleges kulcsához, vagyis belső azonosítójához. A *tbl_fewa* esetében ilyen a weboldalak elsődleges címeinek (*uniform_title_id* – *tbl_uniform_titles.id*), a változékonyságuknak (*variability_id* – *tbl_webpage_variability.id*), a műfajaiknak (*webpage_form_id* – *tbl_webpage_forms.id*), a típusainak (*webpage_type_id* – *tbl_webpage_types.id*), az archiváló intézményeinek (*archivator_id* – *tbl_corporate.id*) a kapcsolata. Hasonló egyszerű kapcsolattal van összekötve a FEWA és a WCT adatbázisa is. A *tbl_fewa.wct_target_id* másodlagos kulcsként kapcsolódik a *DB_WCT TARGET* táblájának *T_AT_OID* mezőjéhez, ahol az archiválási célként (target) kijelölt weboldalak adatai vannak.

Kapcsolótáblák

Az adatbázison belüli kapcsolatok többsége azonban nem olyan egyszerű, mint a fenti esetekben, ugyanis például egy weboldal adattípusánál számos esetben többféle érték is kapcsolódhat. Nézzük például a tartalmi feltárás gyakorlatát, ahol egy weboldal akár többféle kategóriába, és még több témakörbe, altémakörbe besorolható, nem is említve a hatékony feltáráshoz szükséges kiosztható tárgyszavak számát. Ilyen esetekben kapcsolótáblákat kell létrehozni az egy a többhöz kapcsolatok rögzítésére. Ezek a megnevezésükben *reltbl_* előjellel rendelkező táblák egységes felépítésűek és kizárólag kulcsokat tartalmaznak. Minden kapcsolótábla rendelkezik egy saját belső azonosítóval (id, elsődleges kulcs), illetve a két kapcsolódó tábla megfelelő mezőjéhez tartozó másodlagos kulcsokat. Az archívum

weboldalainak kapcsolórendszerében, az egyik másodlagos kulcs mindig a *tbl_fewa* tábla belső azonosítója, míg a másik a kapcsolódó tábla belső azonosítója. A kapcsolótáblák kapcsolati szerepeket is definiálnak. Például az ágens (testületek és személyek) táblák és a weboldalak között eleve adott négyféle kapcsolattípus: az ágensek lehetnek a weboldalak tárgyai, létrehozói, közreműködők a tartalom előállításában, illetve jogtulajdonosok. Ráadásul minden ilyen szerepben egyetlen weboldalhoz több ágens is kapcsolódhat. Ilyenkor előnyös a kapcsolótáblák használata, hiszen ez lehetővé teszi az értékek ismétlődését oly módon, hogy egyrészt az adatbázisban ne kelljen felesleges nullértékeket tárolni, másrészt ez módszer elősegíti a tetszőleges, komplex kapcsolatok létrehozását. A következő táblázat a főtábla (*tbl_fewa*) és a többi tábla (mint céltábla) közötti kapcsolótáblás relációkat szemlélteti.

Kapcsolótábla neve	Főtáblára irányuló mezőnév (főtábla)	Céltáblára irányuló mezőnév (céltábla)	Reláció funkciója
reltbl_fewa_collection	fewa_id (tbl_fewa)	collection_id (tbl_collection)	weboldalakat gyűjteményekbe szervezi
reltbl_fewa_contributor_corporate	fewa_id (tbl_fewa)	corporate_id (tbl_corporate)	weboldalak tartalmi bővítésében részt vevő testületek
reltbl_fewa_contributor_person	fewa_id (tbl_fewa)	person_id (tbl_person)	weboldalak tartalmi bővítésében részt vevő személyek
reltbl_fewa_creator_corporate	fewa_id (tbl_fewa)	corporate_id (tbl_corporate)	weboldalak létrehozásában részt vevő testületek
reltbl_fewa_creator_person	fewa_id (tbl_fewa)	person_id (tbl_person)	weboldalak létrehozásában részt vevő személyek
reltbl_fewa_language	fewa_id (tbl_fewa)	language_id (tbl_language)	weboldalakon használt nyelvek
reltbl_fewa_owner_corporate	fewa_id (tbl_fewa)	corporate_id (tbl_corporate)	weboldalak testületi jogtulajdonosai
reltbl_fewa_owner_person	fewa_id (tbl_fewa)	person_id (tbl_person)	weboldalak személy jogtulajdonosa
reltbl_fewa_subject_categories	fewa_id (tbl_fewa)	category_id (tbl_subject_categories)	weboldalakhoz kapcsolódó tárgyi kategóriák
reltbl_fewa_subject_corporate	fewa_id (tbl_fewa)	corporate_id (tbl_corporate)	weboldalakhoz tárgyként kapcsolódó testületek
reltbl_fewa_subject_geo	fewa_id (tbl_fewa)	geo_id (tbl_geo)	weboldalakhoz tárgyként kapcsolódó földrajzi helyek
reltbl_fewa_subject_heading	fewa_id (tbl_fewa)	subject_heading_id (tbl_subject_headings)	weboldalakhoz kapcsolódó tárgyszavak

reltbl_fewa_subject_meeting	fewa_id (tbl_fewa)	meeting_id (tbl_meeting)	weboldalakhoz tárgyként kapcsolódó rendezvények
reltbl_fewa_subject_person	fewa_id (tbl_fewa)	person_id (tbl_person)	weboldalakhoz tárgyként kapcsolódó személyek
reltbl_fewa_subject_subtheme	fewa_id (tbl_fewa)	subtheme_id (tbl_subject_subthemes)	weboldalakhoz kapcsolódó tárgyi altémakörök
reltbl_fewa_subject_theme	fewa_id (tbl_fewa)	theme_id (tbl_subject_themes)	weboldalakhoz kapcsolódó tárgyi témakörök
reltbl_fewa_subject_uniformtitle	fewa_id (tbl_fewa)	uniformtitle_id (tbl_uniform_titles)	weboldalakhoz tárgyként kapcsolódó művek (más weboldalak, könyvek, filmek, stb.)

Komplex kapcsolatokat azonban nem csupán az archivált weboldalak fő táblájának kontextusában lehet kialakítani. A weboldal-relációk jelentős része megismétlődik a tervezett gyűjtemények esetében is. Az entitások adatbázisban való kezelése során is szükséges a kapcsolótáblák használata, hiszen az egyes entitások más entitásokkal is relációban állnak. Például a földrajzi helyek valamilyen kontextusban kapcsolatban állnak testületekkel (székhely) és rendezvényekkel (helyszín) egyaránt. A következő táblázat az egyes entitások táblái (mint fő tábla) és a többi entitás táblája (mint céltábla) közötti kapcsolótáblás relációkat szemlélteti.

Kapcsolótábla neve	Fő táblára irányuló mezőnév (fő tábla)	Céltáblára irányuló mezőnév (céltábla)	Reláció funkciója
reltbl_collection_language	collection_id (tbl_collection)	language_id (tbl_language)	gyűjteményekben használt nyelvek
reltbl_collection_subject_category	collection_id (tbl_collection)	subject_category_id (tbl_subject_categories)	gyűjteményekhez kapcsolódó tárgyi kategóriák
reltbl_collection_subject_corporate	collection_id (tbl_collection)	corporate_id (tbl_corporate)	gyűjteményekhez tárgyként kapcsolódó testületek
reltbl_collection_subject_geo	collection_id (tbl_collection)	geo_id (tbl_geo)	gyűjteményekhez tárgyként kapcsolódó földrajzi helyek
reltbl_collection_subject_heading	collection_id (tbl_collection)	subject_heading_id (tbl_subject_headings)	gyűjteményekhez kapcsolódó tárgyszavak
reltbl_collection_subject_meeting	collection_id (tbl_collection)	meeting_id (tbl_meeting)	gyűjteményekhez tárgyként

			kapcsolódó rendezvények
reltbl_collection_subject_person	collection_id (tbl_collection)	person_id (tbl_person)	gyűjteményekhez tárgyként kapcsolódó személyek
reltbl_collection_subject_subtheme	collection_id (tbl_collection)	subtheme_id (tbl_subject_subthemes)	gyűjteményekhez kapcsolódó tárgyi altémakörök
reltbl_collection_subject_theme	collection_id (tbl_collection)	theme_id (tbl_subject_hemes)	gyűjteményekhez kapcsolódó tárgyi témakörök
reltbl_collection_subject_uniformtitle	collection_id (tbl_collection)	uniformtitle_id (tbl_uniform_titles)	gyűjteményekhez tárgyként kapcsolódó művek (más weboldalak, könyvek, filmek, stb.)
reltbl_corporate_geo	corporate_id (tbl_corporate)	geo_id (tbl_geo)	testületekhez kapcsolódó földrajzi helyek (székhely)
reltbl_meeting_geo	meeting_id (tbl_meeting)	geo_id (tbl_geo)	rendezvényekhez kapcsolódó földrajzi helyek (helyszín)
reltbl_uniformtitle_contributor_corporate	uniformtitle_id (tbl_uniform_titles)	corporate_id (tbl_corporate)	művek tartalmi bővítésében részt vevő testületek
reltbl_uniformtitle_contributor_person	uniformtitle_id (tbl_uniform_titles)	person_id (tbl_person)	művek tartalmi bővítésében részt vevő személyek
reltbl_uniformtitle_creator_corporate	uniformtitle_id (tbl_uniform_titles)	corporate_id (tbl_corporate)	művek létrehozásában részt vevő testületek
reltbl_uniformtitle_creator_person	uniformtitle_id (tbl_uniform_titles)	person_id (tbl_person)	művek létrehozásában részt vevő személyek

Alternatív névváltozatok kapcsolótáblái

Az egyes entitások névváltozatainak kezelésében is nélkülözhetetlenek a kapcsolótáblák, mivel egy entitásnak több alternatív névváltozata is lehet. Ezek a táblák annyiban különböznek az eddig felsorolt kapcsolótábláktól, hogy nem két tábla között teremtenek kapcsolatot, hanem egy entitástípus táblájához köt hozzá névváltozatokat, tehát a kapcsolótáblában az egyik másodlagos kulcsot felváltotta az alternatív név karakterláncát tároló CHAR(255) adattípusú mező. A következő táblázat az adott entitások egyéb névváltozatait és az azokhoz kapcsolódó entitások közötti kapcsolótáblás relációkat szemlélteti.

Kapcsolótábla neve	Egyéb névadatot tartalmazó mező	Kapcsolódó táblára irányuló mezőnév	Reláció funkciója
--------------------	---------------------------------	-------------------------------------	-------------------

reltbl_corporate_ alt_name	corporate_ alt_name	corporate_id (tbl_corporate)	testületek névváltozatai
reltbl_geo_ alt_name	geo_alt_name	geo_id (tbl_geo)	földrajzi helyek névváltozatai
reltbl_meeting_ alt_name	meeting_ alt_name	meeting_id (tbl_meeting)	rendezvények névváltozatai
reltbl_person_ alt_name	person_ alt_name	person_id (tbl_person)	személyek névváltozatai
reltbl_subject_ category_altname	category_ alt_name	category_id (tbl_subject_ categories)	tárgyi kategóriák névváltozatai
reltbl_subject_ heading_altname	headings_ alt_name	subject_ heading_id (tbl_subject_ headings)	tárgyszavak névváltozatai
reltbl_subject_ subtheme_altname	subtheme_ alt_name	subtheme_id (tbl_subject_ subthemes)	tárgyi altémakörök névváltozatai
reltbl_subject_ theme_altname	theme_ alt_name	theme_id (tbl_subject_ themes)	tárgyi témakörök névváltozatai

Honlap és keresőfelület

Az elkészült webarchívum tartalmának megjelenítésére önálló honlap létrehozása szükséges. Ennek koncepcióját az Aranybulla Webarchívum (AWA) 2023-ban elkészült szolgáltatási honlapjának tulajdonságai alapozták meg. Ezek a következők: könyvtári igényességű részletes keresés, autocomplete (vagyis az adatbázisban elérhető tartalmak megjelennek a keresőmezőkben), könnyen használható és érthető, felhasználóbarát felület. Magának az Aranybulla Webarchívumnak a műszaki tartalma, tehát a háttéradatbázis használata, az archívum tartalmának egyedi tárgyszavazása és metaadatolása, is előfutára volt a megyei helyismereti webarchívumnak. A fő különbség a FEWA méreteiben és általános jellegű tematikájában figyelhető meg, hiszen az AWA egyetlen történelmi esemény weboldalai és online elérhető dokumentumait hivatott megőrizni. Ez azonban nem állja útját az itt kigondolt és megvalósított technológiai innovációk ismételt alkalmazásának, illetve továbbfejlesztésének.

A <https://fewa.vmk.hu/> webcímen elérhető weboldal minden ízében az AWA szolgáltatási honlapjának újragondolása. A kezdőoldal kellemes küllemű, korszerű weboldal. A VMK munkatársainak dizájn- és funkcióötleteit külső webfejlesztő szakemberek keltették életre. Maga a weboldal egy kereskedelmi Wordpress-sablonra épül, a dinamikus tartalommegjelenítéséért is a FEWA háttéradatbázisa szolgál, ugyanis az oldal PHP-scriptek révén kérdez le onnan adatokat. Ezek költségeit a már ismertetett összegkeretekkel a pályázat finanszírozta. A weboldal arculatának központi elemei az állandó fej- illetve lábléc, amely az azonosíthatóságért felelős logók és adatok, köztük az impresszumoldal linkje, mellett a legördülő “hamburgermenü”, illetve a VMK online platformjaira irányuló linkeket tartalmazza. A kezdőoldalon a felhasználó számára jól kiemelkedik az alapértelmezett keresőmező. Ennek alapértelmezett funkciójában minden szempont szerint lehet keresni.

Magát a keresést az autocomplete funkció segíti, egy a weboldal PHP-kódjába beágyazott SQL-lekérdezés, amely a beírt keresőkifejezést behasonlítja a háttéradatbázis megfelelő adatmezőivel és visszaadja a keresőmezőbe az elérhető adatbázisértékeket, jelentősen megkönnyítve a webarchívum tartalmának áttekintését és a tartalmak visszakeresését. Tehát, ha valaki beírja a keresőmezőbe a “könyv” kifejezést, a legördülő autocomplete-fülön megjelenik minden olyan lehetséges tárgyi kifejezés, címszerűség, kapcsolódó ágens és hely, amelynek nevében a “könyv” előfordul. A keresőmező feletti szűkítő szempontok a keresés során lekérdezett adattáblák számát csökkentik, vagyis értelemszerűen a *Honlap neve* alatti keresőmező kizárólag az archivált honlapok címszerűségei és URL-jei szerint keres, a *Téma* a tárgyi kifejezések (kategória, témakör, altémakör, tárgyszó), a *Földrajzi hely* a kapcsolódó földrajzi nevek között, míg a *Személyek és intézmények* a kapcsolódó személyek és testületek értékeit adja vissza az adatbázisból. Mivel az adatbázisban rögzítve vannak az egyes entitások alternatív névalakjai is, ezek is kereshetők a weboldal keresőmezőiben, azonban mivel ezek összezavarhatják a felhasználókat, az autocomplete nem dobja fel találatként. A keresési lehetőség azonban adott, így ha valaki gyorsan gépelve rutinból keres, meg sem várva az autocomplete megjelenítését, akkor, amennyiben a pontos alternatív névalakot írja be, szintén eljut a releváns tartalomhoz.

Amennyiben a felhasználó nem keres konkrét tartalmat, a kezdőoldalon az archívumból véletlenszerűen megjelenő ajánlott oldalak, illetve a könyvtárosok által manuálisan összeállított válogatott oldalak kínálnak ízelítőt a FEWA tartalmából. Ha a keresőmezőben keresnek tartalmat, a kezdőoldal dinamikusan legenerálja az adatbázis alapján a találati listát. Mind az ajánló/válogatott oldalak, mind a keresés találati halmaza virtuális kártyaként jeleníti meg az archivált oldalak hivatkozásait az oldalon. Egy kártya az adott weboldal képernyőképéből és egységesített címéből áll, arra kattintva jut el a felhasználó az oldal adatlapjához, ahol a részletes metaadatok kerülnek feltüntetésre. A metaadatok három vizuálisan elkülöníthető blokkban jelennek meg: a felsőben az egységesített cím, FEWA-azonosító és eredeti URL hármasa, a középsőben a tárgyra vonatkozók, míg alul az egyéb metaadatok olvashatók. Az első és második adatblokk között érhetők el a pywb megjelenítési felületére mutató hivatkozások, vagy jogtulajdonosi jóváhagyás hiányában a könyvtári IP-tartományból való kizárólagos megtekintésről szóló tájékoztató szöveg. A pywb-re két link mutat: az egyik a legfrissebb aratásra mutat, míg a másik a pywb naptárfelületére, ahonnan a felhasználó választhatja ki, hogy melyik dátum szerinti aratást kívánja megtekinteni.

Az Aranybulla Webarchívumhoz hasonlóan a FEWA-nál is lehetőség kínálkozik részletes keresésre a legördülő menüből kiválasztható Keresés fül alatt. Itt az imént ismertetett funkciókkal felvértezett keresés szabható testre felhasználói igény szerint. A következő szempontok szerint lehetséges a keresés: téma, tárgyszó, cím-url, szerző-közreműködő, földrajzi név, személynév, testületi név, nyelv. Ezek közül egyszerre három szempont kombinálható BOOLE-operátorokkal (ÉS, VAGY, NEM). Az autocomplete és az oldalon megjelent találati halmaz tulajdonságai megegyeznek az egyszerű keresésnél leírtakkal. A különbség a PHP-kódba ágyazott lekérdezéseknél jelentkezik.

Tapasztalt hibajelenségek, kihívások, megoldások

Műszaki jellegű problémák, kihívások

Az aratások során a legnagyobb hibalehetőséget a nem szabványos és rossz minőségben elkészített weboldalak képezik, amelyeket nem tud rendeltetésszerűen bejárni az aratórobot. Ilyenkor a be nem járt oldalrészek URL-jét megadhatjuk seedként, sok esetben javul is a mentett weboldal, de bizonyos esetekben ez sem megoldás, ilyenkor az adott oldalt nem tudjuk a megfelelő minőségben menteni.

Általános probléma még a weboldalakba beágyazott médiaelemek (videók, online térképek, közösségimédia-alkalmazások stb.) aratása, amelyeket kevés kivétellel nem sikerül sikeresen reprodukálni a mentett oldalaknál. Az egyik esetben a kezdőoldalon a háttérben menő videónál hiába kerestük meg annak YouTube-linkjét, az aratás nem sikerült. Sokszor a videók mellett a képek is hiányoznak, gyaníthatóan olyan esetekben, amikor a robot nem tudja feltérképezni az adott mellékletek elérési útjait a weboldal könyvtárszerkezetében. Videók esetében is volt olyan eset, hogy annak elérési útját nem találta meg az aratórobot. Néhány esetben a forráskódokból sikerült megállapítani ezeket az elérési utakat, és ezáltal azokat seedként megadni beolvasatni a hiányzó képeket, de ez a módszer sem vezetett mindenkor eredményre. Az iframe-ben beágyazott Google Térkép és a hasonló open source webtérképek, illetve Facebook-komponensek mentése egyszer sem vezetett eredményre. Mivel ezek webarchiválása általános kihívás és nehézség, ezek hiányát egészséges kompromisszumnak tekintjük, mivel az elsődleges cél a weboldaltartalmak, és nem a kapcsolódó egyéb médiumok megőrzése.

A székesfehérvári oldalak archiválását célul kitűző pilotszakasz végén az archivált 329 weboldal közül 147 oldal mentése hiba nélkül, tökéletesen sikerült. További 62 még elfogadhatónak minősül, apróbb hibákat, illetve a nem archiválható beágyazott tartalmakat foglalja magában. A fennmaradó 115 esetben súlyosabb akadályokba ütközött az archiválás, szétesőszerű oldalak, komoly tartalomhiány jellemzi ezeket a mentéseket. Összességében tehát elmondható, hogy jelen pillanatban az archívum harmadánál jelentkeznek hiányosságok. Reményeink szerint ez a mennyiség az archiváló technológia fejlődésével, illetve a saját tapasztalataink gyarapodásával fokozatosan kevesebb lesz a jövőben. Az idő múlásával még a tárhelybővítéssel is számolni kell, különösen ami a nagyon gyakran frissülő, méretes weboldalakat illeti.

Munkaerőt érintő kihívások

A FEWA-projekt több csoport együttműködésének keretében került kialakításra, de intézményi kihívást jelent, hogy a mindennapi munkába integráljuk ezt az új archiválási módot. További kihívást jelent, hogy a webarchiválás mint olyan megjelenése újraszabja az intézmény teljes archiválási gyakorlatát, amivel a projekt elején nem számoltunk, hiszen nem csupán arról van szó, hogy a digitális tartalmak mentésre kerülnek, hanem arról is, hogy a szolgáltatási felületet és a gyakorlatot hogyan hozzuk úgy létre, hogy az discovery jellegű legyen. Egyéb megoldás csak tovább szabdalja a felhasználók figyelmét, hajlandóságát. Ennek megfelelően újragondoltuk a Helyismereti és digitalizálási csoport helyét a szervezetben, új, elfogadásra váró Szervezeti és működési szabályzatunkban már a webarchiválást tartalmazó névvel és új szervezeti struktúrában kerül megjelenítésre.

Állományalakítást érintő kihívások

A pilotszakasz során nem történt specifikus tartalomkeresés a webarchívum tartalmát illetően, mivel az előzetes munka miatt rendelkezésre állt egy relatíve nagy URL-lista. Ezek archiválását követően azonban új tartalmakkal kell bővíteni a FEWA-t. Helyismereti webarchívumként a már ismertetett módon három szempont (földrajzi, személyi/testületi, történeti) szerint lehetséges az állomány alakítása. Eddig alapvetően a legegyszerűbb, a földrajzi szempont érvényesült, ugyanis csak székesfehérvári illetőségű oldalakat archiváltunk. A szempontok között persze van átfedés, mivel például a földrajzi megközelítés már eleve magában foglalja az adott településen székelő testületek weboldalainak felkutatását. A tartalmak felkutatása e szempont szerint viszonylag magától értetődő, azonban a többi szempont érvényesítése számos kihívást rejthet magában. A feladat a helyismereti különgyűjtemény gyakorlatára alapul, tehát az ott nyilvántartott személyekre, testületekre és eseményekre vonatkozóan kell egyéni kereséseket lefuttatni annak érdekében, hogy további a FEWA-ban archiválandó weboldalakat találjunk. Ezek a helyismereti authorityadatok jó kiindulási alapot biztosítanak, a potenciális problémát leginkább úgy lehetne jellemezni, mint „tűt a szénakazalban”. Minél népszerűbb, feltártabb, ismertebb a keresett téma, annál inkább megnövekszik a potenciális weboldalak száma, amelyeket egyesével a helyismereti szempontok szerint mérlegelni kell, hogy bekerüljön-e az archívumba. Rengeteg kitűnő helyismereti tartalom bújik meg jól vagy kevésbé ismert blogok, elrejtett weboldalak és portálok mélyén, amelyeket még a korszerű keresőmotorokkal sem feltétlenül egyszerű megtalálni, egyszerűen azért, mert a könyvtáros nem ismeri ezeket a lelőhelyeket, és mivel „eldugottnak” számítanak, tehát, hirdetési felületként sem vehető számításba, ezért a találati listákban is oly mértékben hátrra kerülhet, hogy elkerüli a kollégák figyelmét. A legcélravezetőbb tartalomkeresési stratégia talán az lehetne, hogy a megszokott helyismereti módszerek mellett, lelkes kollégákat és érdeklődőket felkérni kedvenc témáik keresésére. Ők valószínűleg ismernek ilyen, a net mélyére „jól eldugott” értékes forrásokat. Tehát összességében elmondhatjuk, hogy az állományalakítás legnagyobb kihívása ezesetben az, hogy a legértékesebb online tartalmak felkutatására a megszokottnál jóval alaposabban kell ismerni a keresett témát.

A webarchívum mint szolgáltatás, kommunikáció

A FEWA (<https://fewa.vmk.hu/>) mint minden webarchiválási kezdeményezés a kezdeti időkben szolgáltatásként valószínűleg nem számíthat általános érdeklődésre. Míg az archivált oldalak „élnek”, tehát a mindennapokban használt keresőmotorok lévén könnyedén megtalálhatók, az átlagos felhasználók látókörén kívülre esik a webarchívum, és ez így is marad, míg az időközben elhalt, eltűnt weboldalak száma el nem éri a kritikus tömeget. Ekkor fognak az érdeklődők a már megszűnt oldalak felkutatása iránt érdeklődni.

Ennek érdekében a 2024. évi első megyei szakmai napon előadás szolt a webarchiválás fontosságáról a megyei könyvtárosok részére, megjelölve többek között a kistépülési kollégák szerepét a webarchiválás folyamatában. Készítettünk egy URL-beküldő űrlapot⁵, amin bárki meg tudja osztani velünk a helyi értékeket tartalmazó weboldalak elérhetőségét.

⁵Webarchiválási javaslat: Fejér vármegyei webtartalmak: <https://forms.office.com/e/9vsw5Ya2LP>

A FEWA-ról a könyvtárszakmának Bódog András tartott bemutatót, ismertetőt *A metaadatolás és az adatbázis-építés kihívásai a Fejér Vármegyei Webarchívumban* címmel.⁶

A közönség számára a témáról Horváth Adrienn beszélt a Fehérvár Televízió Fehérvári Beszélgetések című adásában.⁷

Az oldal további bemutatása, népszerűsítésére adatbáziskeresésre épülő foglalkozásokat dolgozunk ki, valamint szakmai napokon népszerűsíteni kívánjuk oldalunkat – utóbbi időpontja 2025. március 3.

⁶ 404 Not Found – Ki őrzi meg az internetet? Konferencia és workshop 2024, Országos Széchényi Könyvtár https://www.oszk.hu/rendezvenyek/404-not-found_ki-orzi-meg-az-internetet-konferencia-es-workshop-241127

⁷ Fehérvári Beszélgetések: <https://www.youtube.com/watch?v=JVf2qGcrSzE>