

NETWORKSHOP

30

Online térben - az online térért

ORSZÁGOS ONLINE KONFERENCIA

2021. április 6-9.
Eötvös Loránd Tudományegyetem



ONLINE TÉRBEN – AZ ONLINE TÉRÉRT

Networkshop 30: országos online konferencia

2021. április 6–9.

Eötvös Loránd Tudományegyetem

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2021



INNOVÁCIÓS ÉS TECHNOLÓGIAI
MINISZTERIUM



Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Networkshop

2021. április 6–9. Eötvös Loránd Tudományegyetem, (On-line)
konferencia előadásainak közleményei

ISBN 978-615-01-2956-3

DOI: [10.31915/NWS.2021](https://doi.org/10.31915/NWS.2021)

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével
Budapest
2021

Borítókép: [freepik.com](https://www.freepik.com)

TARTALOMJEGYZÉK

Előszó	7
TANULMÁNYOK A TARTALOMSZOLGÁLTATÁS KÖRÉBŐL	
Drótos László: A közösségi média archiválásának nehézségei	9
Németh Márton: A COVID járvánnyal kapcsolatos webarchiválási tevékenységek – nemzetközi és hazai körkép.....	17
Csík Tibor: A könyvtárak az új digitális világban	25
Ungváry Rudolf: A földrajzi személyi hungarikum feldolgozásának problematikája.....	35
Farkas Katalin – Nagy Dóra: Speciális dokumentumok és tartalmak – speciális metaadatszerkezetek – Muzeális tartalmak és kéziratok repozitálása a Szegedi Tudományegyetem Klebelsberg Kuno Könyvtárában	45
Kiss Margit – Palkó Gábor – Szakács Béla Benedek: Szöveg hasonlósági vizsgálatok automatizálása	53
Bajzát Tímea Borbála – Szemes Botond Bálint – Szilágyi Eszter: Az ELTE DH Regénykorpusz és lehetőségei	63
Szűcs Kata Ágnes: Automatikus kézírás-felismerés Kiss József levelezésén	73
Timári Mária – Bajzát Tímea Borbála – Palkó Gábor: 19. századi magyar regényeken végzett kísérletek a magyar nyelvű szerzőazonosítás leghatékonyabb távolságméréseinek megtalálására	81
Mihály Eszter: A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben	91
Görög Dániel – Rényi Máttyás: A Mikes Kelemen program tapasztalatai és megújult működése	111
Lendvai Miklós: A könyvtári együttműködés informatikai támogatása: adatmodell, workflow, rendszerfelépítés : Együttműködési lehetőségek a nyílt forráskódú platformokban: OKP, FOLIO, ReShare.....	121
Bolya Máttyás: A BTK Zenetudományi Intézet digitális archívum koncepciója az oktatás és a tudomány szolgálatában	133
Várnai-Vígh Adrienn E. – Kokas Károly: Tömeges könyvtári felhasználóképzés távoktatási formában az SZTE Klebelsberg Könyvtár e-learning platformján.....	143
Fülöp Tiffany – Nagy Gyula: Az online oktatás könyvtári támogatása a Szegedi Tudományegyetemen	153
Sörény Edina: A digitáliskompetencia-fejlesztés új eszköze: a DigKomp Rendszer	161
Békésiné Bognár Noémi Erika – Nagy Andor: A könyvtári szolgáltatások átalakulása és fejlesztése a pandémia idején.....	169
Zeller Rozália – Hoczopán Szabolcs – Nagy Gyula: Kutatási adatkezelést támogató szolgáltatás a Szegedi Tudományegyetemen	179
Holl András: Intézményi, támogatói szabályozások kialakítása a nyílt tudomány kutatási adatkezelési céljaihoz	189
Kalcsó Gyula: Born digital workflow tervezése a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában	195

TANULMÁNYOK A MŰSZAKI TERÜLETEKRŐL

Koczka Ferenc: OSINT technológiák és alkalmazási lehetőségeik a felsőoktatási rendszerek ellen.....	203
Mohácsi János: KIFÜ nemzetközi együttműködései és projektjei.....	211
Horváth Krisztián - Keller Zsófia - Oravecz Zsolt - Sass-Gyarmati Norbert - Geda Gábor: Komplex informatikai megoldás természettudományos ismeretek bemutatására az egri Varázstoronyban.....	221

ELŐSZÓ

Tisztelt Olvasó!

A negyedik szerkesztett Networkshop konferenciakötetet látja képernyőjén, amelyet a HUNGARNET Egyesület az MTA Könyvtár és Információs Központ közreműködésével adott közre. A Networkshop immár harmincadszor került megrendezésre, első alkalommal fővárosi egyetem – az ELTE segítségével. A hagyománytól eltérő intézményválasztás a járványhelyzet által megkövetelt on-line lebonyolítás következménye volt. Mint a három korábbi kötet esetében is, minden tanulmány DOI azonosítót kapott, és bibliográfiai adataikat az MTMT-ben is rögzítettük. A digitális kötet megtalálható lesz az MTA Könyvtár és Információs Központ Open Conference Systems rendszerében, a REAL repozitóriumban – és a CC BY felhasználási licencnek köszönhetően bizonyára máshol is.

Örömmel nyugtázzuk, hogy az idei kötet „vastagabb” lett, és még az év vége előtt megjelenhetett: ez szerzőink és szaklektoraink igyekezetének köszönhető.

Köszönjük az Információs és Technológiai Minisztérium támogatását!

Budapest, 2021. szeptember 27.

A szerkesztők

TANULMÁNYOK A TARTALOMSZOLGÁLTATÁS KÖRÉBŐL

A közösségi média archiválásának nehézségei

Drótos László
Országos Széchényi Könyvtár Webarchiválási Osztály
drotos.laszlo@oszk.hu

The difficulties of archiving social media

Web 2.0 sites designed for user-uploaded textual and audiovisual content began to grow exponentially in the 2000s, and today they have become the most important communication channels for culture and public life, beyond the sharing of private information targeted at a narrow audience. However, these platforms are even more “in present tense” than the traditional web. And this is not just about publicly published and then removed content, but about dozens of services with millions of users that have been discontinued for economic reasons, with gigabytes or terabytes of digital material disappeared overnight. Nowadays most social platforms are more optimized for mobile devices, and not archive-friendly at all, or even explicitly prevent downloading by automatic methods. This paper details the problems of their long-term preservation based on the tests carried out within the framework of the web archiving project in the National Széchényi Library, as well as international examples.

Keywords: web archiving, long-term preservation, social media, web 2.0

1. Öntörvényű világok

A World Wide Web megjelenésével az 1990-es évek elején egy új digitális univerzum született meg, mely azóta is fénysebességgel tágul. Weboldalak milliárdjai jöttek létre és tűntek el nyomtalanul, kivéve azokat, amelyeket sikerült – bár sokszor csak töredékesen – elmenteni az olyan archívumokba, mint az 1996-ban indult Internet Archive¹. A világháló első tíz évében jellemzően fájlrendszerben tárolt, statikus HTML oldalakból állt, szintén fájllokként beágyazott médiatartalmakkal, az interaktív funkciókat pedig egyszerű CGI szkriptekkel oldották meg. Ha eltekintünk az olyan, a mai böngészők által már nem támogatott technológiáktól, mint a VRML, a Java-applet, a Microsoft Silverlight, vagy az idén nyugdíjba vonult Adobe Flash, az 1.0-ás web viszonylag robot- és archívumbarát volt, jól menthető automatikus módszerekkel és évtizedek múlva is visszanezhető.

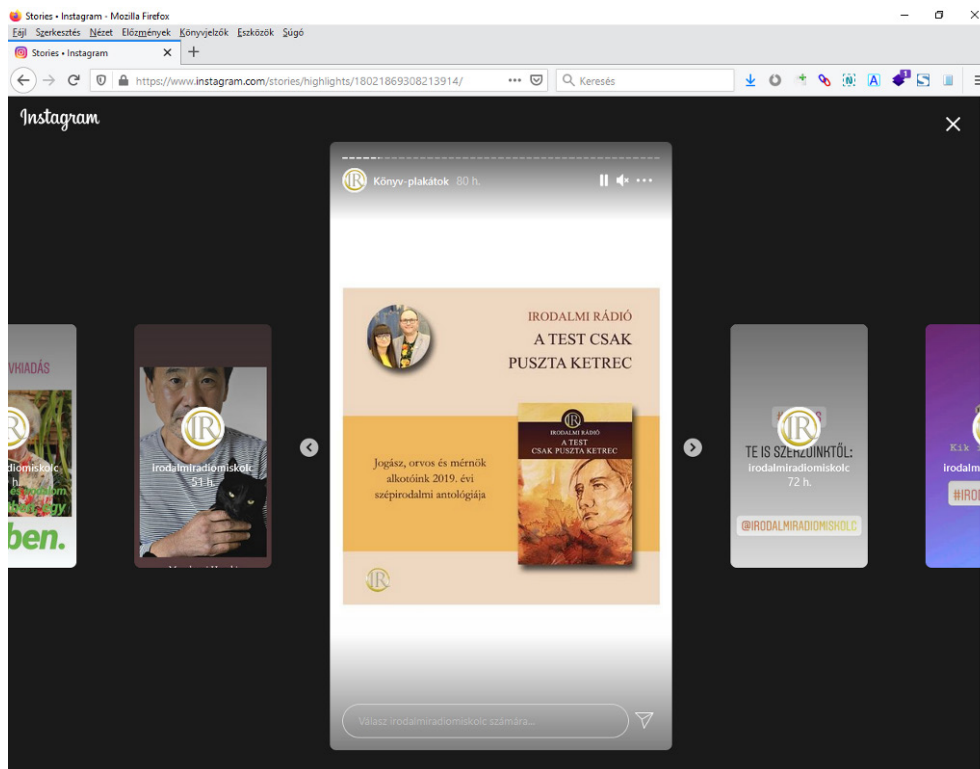
A 21. század internetjét viszont már az adatbázisokból dolgozó programok által generált „webkettes” platformok dominálják, azokon zajlik az igazi élet, ott jelenik meg a legtöbb új tartalom és onnan is tűnik el a leggyorsabban sok minden. Ezek a közösségi terek öntörvényű univerzumokként léteznek a hagyományos web mellett, hasonlóan a már nemcsak sci-fi ötletként, hanem tudományos elméletekben is megjelenő buborék univerzumokhoz. A fizikusok szerint ezekben a feltételezett párhuzamos világegyetemekben eltérő törvények uralkodnak, például más a fény sebessége vagy az elektron tömege, és emiatt teljesen más struktúrák alakulhatnak ki, mint az általunk

1 Wayback Machine, Internet Archive. Hozzáférés: 2021.05.10., <http://web.archive.org>



ismert atomok és égitestek. Az új generációs webes szolgáltatásokat üzemeltető cégek is saját világokat alakítottak ki és ezeknél sajnos egyáltalán nem szempont a hosszú távú megőrizhetőség és a kutathatóság. Az alábbi lista csak néhányat sorol fel a sokféle, archiválhatósági szempontból problémás megoldás közül:

- Egyre több helyen a nyilvános tartalom elérése is belépéshez kötött, regisztráció és azonosítás nélkül csak korlátozottan lehet böngészni a publikus felületen. Az ilyen zárt rendszerekbe sem a keresőgépek, sem a webarchívumok által futtatott szoftverrobotok nem tudnak bejutni. Az API-n keresztül való adatkinyerést is mind jobban korlátozzák a szolgáltatók a 2010-es évek közepétől a Cambridge Analytica-ügy és azt követő hasonló botrányok után.
- A nagy webes keresők robotjainak a kizárása vagy erős korlátozása miatt az archiválásra érdemes tartalom csak a platform saját belső keresőjével található meg, az viszont személyre szabott találatokat ad a felhasználó és ismerősei korábbi preferenciái alapján.
- Alternatív felületek léteznek, melyek nem teljesen ugyanazokat a tartalmakat és funkciókat szolgáltatják asztali és mobil böngészőkben, illetve applikációk formájában.
- Egyre gyakoribb, hogy már csak saját app-on keresztül érhető el a szolgáltatás, a honlapján mindössze egy letöltő link található az alkalmazásboltra. Ezek az applikációk persze egyáltalán nem jövőállóak és csak egy-két mobil operációs rendszer adott verzióin működnek.
- Egy oldalon belül tulajdonjogilag kevert a tartalom, mert az algoritmus a felhasználó számára relevánsnak gondolt más híreket és posztokat is mellé tesz, így hiába van engedélye az archiválónak egy fiók mentésére, nem tudja kizárni a más forrásokból odakerült elemeket.
- A bejegyzések és feltöltések mellett sok személyes adat is megjelenik (pl. kommentek, lájkok, nicknevek, avatar képek), melyeknek az archiválása és szolgáltatása GDPR-kérdéseket vet fel.
- A tartalomhoz való hozzáféréshez folyamatos emberi közreműködés szükséges (pl. „vég nélküli” görgetés vagy lapozás) és folyamatos a háttérben zajló adatcsere is a kliens és a szerver között. Ezek a megoldások nagyon megnehezítik a tartalom automatikus aratását, illetve a lementett anyagnak az eredetihez hasonló megjelenítését.
- Az ugyanazon az URL címen a görgetés vagy lapozás hatására megjelenő, illetve a saját URL nélküli, beágyazott vagy felugró tartalmi elemek a fájlok beazonosíthatóságát teszik lehetetlenné az archívumban. (1. ábra) A másik véglet az ugyanahhoz az elemhez véletlenszerűen generált URL-ek tömege, ami miatt az aratószoftver azt hiszi, hogy mindig új tartalmat tölt le, pedig valójában egy robotcsapdában kering.



1. ábra. Az Instagram *Stories Highlights* funkciója ugyanazon az URL címen jeleníti meg a képeket és videókat, így ezeket sem linkelni, sem robottal bejárni nem lehet.²

- Az idő is egészen sajátosan múlik ezekben az univerzumokban, a jelenidejűségük miatt tipikus az ilyen dátumkijelzés az egyes bejegyzések mellett: „40 perce” vagy „2 hónapja”, aminek természetesen semmi információtartalma nincs, ha egy archivált verziót nézünk vissza. Az pedig még az élő szolgáltatásnál is nagyon zavaró, amikor nem lehet adott dátumra ugrani a hírfolyamban. Az archiválást emiatt nem, vagy csak hosszas görgetés után tudjuk onnan folytatni, ahol korábban abbahagytuk.
- Az önmegsemmisítő üzenetek (pl. Snapchat) és posztok (pl. Instagram story) megítélése már nem ennyire egyértelmű, mert egyrészt érthetően szükség van ilyenekre a magán jellegű kommunikáció védelme érdekében, de amikor egy közszereplő a nyilvánosság számára tesz közzé időzített módon törlődő tartalmat, akkor igen kicsi az esély arra, hogy az bekerül valamilyen közgyűjteményi archívumba. Persze sokszor vannak élelmes felhasználók és újságírók, akik még időben csinálnak róla másolatokat, akár csak azokról a tartalmakról, amiket maga a közlétező töröl le, amikor azok kezdenek kellemetlenné válni számára.
- A platform üzemeltetője vagy az általa futtatott szoftver algoritmusai is folyamatosan és tömegesen töröl tartalmat a nyilvános felületről a közösségi elvek vagy a jogszabályok (pl. a szerzői jog) megsértése miatt. Gyakori az is, hogy felhasználókat tiltanak le és a teljes fiókjukat zárolják akkor is, ha annak csak kis része volt szabálysértő. Így járt az Egyesült Államok elnöke is 2021 elején, akit előbb a Twitter, majd az összes többi nagy közösségi platform kitiltott.

² *irodalmiradiomiskolc* – Az Irodalmi Rádió szerkesztő-házaspárjának könyvélményei, Instagram. Hozzáférés: 2021.02.21., <https://www.instagram.com/irodalmiradiomiskolc/>



- De nemcsak egyes posztok vagy fiókok tűnnek el a szervezett megőrzés esélye nélkül, hanem egész platformok zárnak be anélkül, hogy legalább a kutatók hozzáférést kaphatnának valamilyen ellenőrzött módon ahhoz a sok giga- vagy terabájthoz, amit a felhasználók milliói az évek során felhalmoztak (pl. iWiW, Hotdog, Myspace, Google+, Panoramio ... és a sor még hosszan folytatható).
- Arra ma már a legtöbb webkettes szolgáltató lehetőséget biztosít, hogy a használói a saját tartalmaikat kimenthessék a rendszerből, így a személyes célú archiválás legalább megoldható. Természetesen ezekből az export csomagokból az eredeti felület és kontextus nem rekonstruálható, csak a korábban feltöltött médiafájlok vannak bennük, meg a szöveges tartalmak JSON, XML, vagy valamilyen más, nem túl felhasználóbarát formában.
- Közgyűjteményi, webkurátori szempontból a legbosszantóbb jellemzője ezeknek az öntörvényű világoknak, hogy még a saját szabályaikat sem tartják be. Mire sikerül kidolgozni egy olyan technológiát és munkafolyamatot, amivel a kulturális vagy tudományos szempontból értékes/érdekes tartalmakat megfelelő mennyiségben és minőségben archiválni lehet, addigra a platform változtat valamint a felületén, szigorít a hozzáféréseken, átalakítja a képek és videók megjelenítő modulját ... és megint el lehet kezdeni az archiváló és visszanező szoftverek hozzáigazítását a megújult rendszerhez. (Ez történt tavaly is, amikor a Facebook dizájnt váltott.)

2. OpenSocial

A 2007-ben elég nagy médiavisszhangot kapott OpenSocial³ a közösségi média szabványosítására tett első komolyabb kísérlet volt. A kezdeményezésben a Google tulajdonú Orkut mellett a Friendster, a LinkedIn, a MySpace és néhány más közösségi platform vett részt, majd 2014-től a W3C is bekapcsolódott a munkába⁴ és az Open Web Platform projektjük keretében létrehoztak egy Social Web Working Group nevű munkacsoportot⁵. Az együttműködés célja egy moduláris, szabványos API-kon keresztül kommunikáló platform, illetve az ehhez szükséges protokollok kidolgozása volt.⁶ Bár elsősorban az egyes rendszerek közötti együttműködést és az azonos funkciók (pl. ismerősök felvétele, kommentálás, képek és videók megosztása) leprogramozását könnyíti meg ez az ajánlás, de ki lehetne bővíteni a hosszú távú megőrzéshez hasznos részekkel is. Sajnos a munkacsoport 2018-ban befejezte a munkáját (2. ábra) és azóta nem igazán hallani arról, hogy a nagy közösségi média cégek a nyílt szabványok irányába fejlesztenék a rendszereiket.

3 OpenSocial, Wikipedia. Hozzáférés: 2021.05.10., <https://en.wikipedia.org/wiki/OpenSocial>

4 OpenSocial Foundation moves standards work to W3C social web activity, W3C Blog. Hozzáférés: 2021.05.10., <https://www.w3.org/blog/2014/12/opensocial-foundation-moves-standards-work-to-w3c-social-web-activity>

5 Socialwg, W3C Wiki. Hozzáférés: 2021.05.10., <https://www.w3.org/Social/WG>

6 Social Web Protocols – W3C Working Group Note 25 December 2017, World Wide Web Consortium. Hozzáférés: 2021.05.10., <https://www.w3.org/TR/2017/NOTE-social-web-protocols-20171225/>

Log in

Page Discussion Read View source View history Search

Socialwg

Welcome to the **Social Web Working Group** (SocialWG) home page (<https://www.w3.org/SocialWG/>)!

The Social Web Working Group was active 2014-07-21 to 2018-02-13.

Please see the [Social Web Incubator Community Group](#) for current and future activity.

Social Web Working Group

- [Announcement 2014-07-21](#)
- [Charter](#) (closed on 2018-02-13)

Chairs for the duration

[Tanlek Çelik](#) (Mozilla)
[Evan Prodromou](#) (e14N)

Staff contacts during and at close

[Amy Guy](#) (W3C/University of Edinburgh)
[Sandro Hawke](#) (W3C/MIT)

Specifications

Specifications that the Social Web WG has produced

Some use cases have been addressed by multiple specifications that provide alternative approaches, so trying to figure out what's what can be a bit confusing. [Social Web Protocols](#), an overview of these protocols, should help with this.

- [Document Status](#)

Recommendations

- [Webmention](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Linked Data Notifications](#) ([Sarven Capadmal](#), [Amy Guy](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Micropub](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Activity Streams Core](#) ([James Snell](#) and [Evan Prodromou](#)) (Editor's Draft, Issues, Repo, Implementations)
- [Activity Streams Vocab](#) ([James Snell](#) and [Evan Prodromou](#)) (Editor's Draft, Issues, Repo)
- [WebSub](#) ([Julien Genestoux](#), [Aaron Parecki](#)) (Issues, Repo) - formerly PubSubHubbub
- [ActivityPub](#) ([Chris Webber](#), [Jessica Tallon](#)) (Editor's Draft, Issues, Repo)

NOTES

- [JF2](#) ([Ben Roberts](#)) (Editor's Draft, Issues, Repo)
- [Post Type Discovery](#) ([Tanlek Çelik](#)) (Editor's Draft, Issues, Repo)
- [Social Web Protocols](#) ([Amy Guy](#)) (Editor's Draft, Issues, **Contains: Guide to all these drafts**)
- [IndieAuth](#) ([Aaron Parecki](#)) (Editor's Draft, Issues, Repo)

Specification Related

- List of github aliases of Social Web WG contributors: [Socialwg/Github](#)
- Resolutions - e.g. to accept & publish drafts, for controversial issues, etc.

Discussion

The Working Group welcomes participation from everyone. We currently use IRC, email, telcons, and face-to-face meetings for discussion. You have to be a WG member to attend telcons and F2F meetings, but anyone is free to observe these on IRC!

Extensions to Working Group specs are done in the [Social Web Incubator Community Group](#) (SocialCG), which is open for anyone to join as a member.

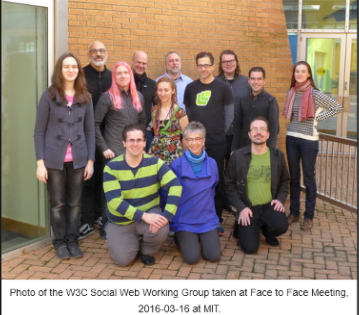


Photo of the W3C Social Web Working Group taken at Face to Face Meeting, 2016-03-16 at MIT.

Contents [hide]

- 1 Social Web Working Group
- 2 Specifications
 - 2.1 Recommendations
 - 2.2 NOTES
 - 2.3 Specification Related
- 3 Discussion
 - 3.1 IRC
 - 3.2 Email
 - 3.2.1 Past Meetings
- 4 Documentation
 - 4.1 Naming
 - 4.2 Historical
 - 4.3 Use Cases
 - 4.4 Social Data Syntax
 - 4.5 Social Vocabulary
 - 4.6 Social API
 - 4.6.1 API research
 - 4.6.2 API candidates
 - 4.7 Federation
 - 4.7.1 Federation candidates
 - 4.8 Inputs into SocialWG
 - 4.8.1 Activity Streams
 - 4.8.2 IndieWeb
 - 4.9 Dependent Work

2. ábra. A 2014 és 2018 között működő Social Web Working Group szócikke a W3C wikijében.

3. BESOCIAL

2020-ban a Belga Királyi Könyvtár három ottani egyetemmel együttműködve BESOCIAL néven indított egy két éves projektet⁷ a közösségi média archiválhatóságával kapcsolatban. A céljaikat az International Internet Preservation Consortium (IIPC) által szervezett Research Speaker Series című webinariumon ismertették.⁸ Megpróbálnak egy fenntartható stratégiát kidolgozni ennek a komplex és gyorsan változó digitális tartalomnak a válogatására, begyűjtésére, megőrzésére és hozzáférhetővé tételére. Első lépésben a Belgiumban és más országokban már létező közösségi média archívumokat, az általuk használt technikai megoldásokat és a jogi környezetet mérték fel. Ennek a fázisnak a kutatási jelentése jelen cikk írásakor még nem volt elérhető, de várhatóan rövidesen megjelenik a könyvtár honlapján. Ezt azután hat további szakasz követi majd, köztük egy pilot projekt, amely során a tartalom arathatóságát, a lementett anyag minőségét és hosszú távú megőrizhetőségét fogják megvizsgálni.

⁷ BESOCIAL, Koninklijke Bibliotheek van België. Hozzáférés: 2021.05.10., <https://www.kbr.be/en/projects/besocial/>

⁸ BESOCIAL research project @KBR: towards a sustainable social media archiving strategy for Belgium, YouTube. Hozzáférés: 2021.05.10., <https://www.youtube.com/watch?v=bX7A5pDoMmQ>



4. Külföldi példák

Európában több nemzeti könyvtár és levéltár is foglalkozik webkettes tartalmak archiválásával, vagy a közeli jövőben tervezi ezt. Többnyire a Facebook, a Twitter, a YouTube, a Flickr és az Instagram oldalairól válogatnak fiókok és/vagy témák szerint a közszereplők (pl. politikusok, újságírók, sportolók, más hírességek) és a szervezetek (pl. hírügynökségek, médiacégek, pártok, kormányzati szervek, kulturális intézmények) által közzétett tartalmakból. Ebből a szempontból kiemelkedő a dán és a brit webarchívum, utóbbinak kiváló nyilvános felülete is van.⁹

Érdekes még megemlíteni az amerikai Library of Congress esetét a Twitterrel, amely a közösségi média megőrzésének egy további nehézségére világít rá. A Kongresszusi Könyvtár 2010-ben jelentette be, hogy 2006-ig visszamenőleg megkapja a Twitter teljes anyagát és előremenetben pedig minden publikus tweetet. 2017 végén viszont már egy olyan közleményt adtak ki¹⁰, mely szerint a jövőben már csak válogatni fognak a Twitter csatornák közül, mert a teljes üzenetfolyam mérete olyan drámaian nő, hogy annak kezelése meghaladja a könyvtár lehetőségeit.

5. Módszerek és eszközök

Mint a fentiekből kiderült, a webkettes tartalmak archiválása komoly technikai kihívás. Szerencsére egyre több olyan szolgáltatás, illetve szoftver jelenik meg, amelyek erre a célra (is) alkalmasak. Itt most csak felsorolásszerűen teszünk ezekről említést, de többségükről külön szócikk van az OSZK Webarchívum honlapjáról¹¹ elérhető wikiben.¹² A csillaggal jelletteket mi is használjuk a nemzeti könyvtárban.

Fizetős platformok, szolgáltatások: MirrorWeb, ArchiveSocial, PageFreezer, Smarsh, Intradyn, Jatheon, Archive-It

API-n keresztül archiváló szoftverek: Social Feed Manager, Twarc, F(b)arc, TAGS, Munin
Fájlletöltők: youtube-dl*, DownThemAll!, FastVid, Story saver, Social Downloader

Böngészőn keresztül archiváló eszközök: Brozler*, Browsertrix, WAIL*, Webrecorder Desktop*, ArchiveWeb.page*, Conifer*, Instamancer, Crocoite

Megjelenítők: ReplayWeb.Page*, WebrecorderPlayer*, Conifer*, PyWb*, OpenWayback*

6. Saját tapasztalatok

Az Országos Széchényi Könyvtárban 2017-ben indult webarchiválási projekt első éveiben a hagyományos web aratására koncentráltunk és bár már 2019 elején megpróbálkoztunk az OSZK saját online szolgáltatásai, köztük a webkettes oldalai

⁹ UK Government Social Media Archive, The National Archives. Hozzáférés: 2021.05.10., <https://webarchive.nationalarchives.gov.uk/social/search/>

¹⁰ Update on the Twitter Archive at the Library of Congress, Library of Congress Blog. Hozzáférés: 2021.05.10., <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/>

¹¹ OSZK Webarchívum honlapja. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu>

¹² MIA Wiki, OSZK Webarchívum. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu/mediawiki/>

mentésével (3. ábra), komolyabban csak tavaly kezdtünk el kísérletezni a közösségi média archiválásával. Február végétől május közepéig kb. 700 publikus magyar Instagram fiókból minden képet és videót letöltöttünk (összesen több mint százezer bejegyzést), majd ősszel frissítettük is ezt a részgyűjteményt az időközben közzétett anyagokkal. Ezután 147 könyvtári Facebook oldalról mentettünk közel 30 ezer bejegyzést, többnyire képpel, illetve néhol videóval együtt. 35 esetben sikerült a teljes oldalon végigmenni (a legrégebbi letöltött poszt dátuma 2010.04.28. volt), a többinél pedig az utolsó egy-két év bejegyzéseiből mentettünk le átlagosan körülbelül kétszázat. Mindkét platform esetében a Webrecorder Desktop programot és annak online változatát, a Conifert használtuk. 2021 tavaszán pedig az ArchiveWeb.page Chrome böngészőbe integrálható modulját teszteltük a Twitteren. Az első alkalommal 89 könyvtár, múzeum és galéria csatornáját töltöttük le 600 MB összméretben. 75 esetben sikerült egészen az első bejegyzésig visszagörgetni, a legrégebbi elmentett tweet 2007 decemberi volt.

Webkettes oldalak

A webhely neve	URL címe	Archiváló szoftver	Megjelenítő szoftver					Oldalkép	IA mentés	Eredeti
			OpenWayback	PjWb	SoiWayback	Conifer	Webszerver			
DKA – Facebook oldal (mobil felület)	m.facebook.com/digitaliseparchivum	Webrecorder	🔴	🔴	🔴	🔴	—	🔵	🟪	🟠
MEK – Facebook oldal	www.facebook.com/Magyar-Elektronikus-Könyvtár-338675729150	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – Facebook oldal	www.facebook.com/nemzetikonyvtar	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – Google+ oldal	plus.google.com/u/0/114309059631430797547/	Webrecorder	🔴	🟡	🔴	🔴	—	🔵	—	—
OSZK – Indafotó oldal	indafoto.hu/nemzetikonyvtar/	Webrecorder, HTTrack	🟡	🟡	🟡	🟡	🟡	🔵	🟪	🟠
OSZK – Instagram oldal	www.instagram.com/nemzetikonyvtar/	Webrecorder	🔴	🟢	🔴	🟢	—	🔵	🟪	🟠
OSZK – Pinterest oldal	hu.pinterest.com/nemzetikonyvtar/	Webrecorder	🔴	🟡	🔴	🟡	—	🔵	🟪	🟠
OSZK – Tumblr mikroblog	nemzetikonyvtar.tumblr.com	Webrecorder, HTTrack	🟢	🟢	🟢	🟢	🟡	🔵	🟪	🟠
OSZK – Twitter mikroblog	twitter.com/nemzetikonyvtar/	Webrecorder	🔴	🔴	🔴	🟡	—	🔵	🟪	🟠
OSZK – YouTube csatorna	www.youtube.com/user/nemzetikonyvtar/	Webrecorder, Conifer	🔴	🟡	🔴	🟡	—	🔵	🟪	🟠

3. ábra. Az OSZK saját webkettes oldalainak mentései különféle megjelenítőkkal. A piros nyilak a teljesen hibásak, a sárgák a kissé hibásak vagy hiányosak, a zöldek pedig a jók.¹³

Az eddigi tapasztalataink nem túl biztatóak. A webkettes tartalmak üzemserű és tömeges archiválásának egyik nagy akadálya a magas élőmunka igény. Mivel a munkafolyamatok zöme nem automatizálható, ezért valakinek ott kell ülnie a számítógép előtt, bejelentkezni az adott platformra, elindítani az archiváló szoftvert, megnyitni a letöltendő URL címet, végiggörgetni és végigkattintgatni az oldalt, lementeni az archivált fájlt, majd feltölteni a szerverre, végül pedig rögzíteni a fontosabb metaadatokat. Ezen a hosszadalmas folyamaton az archiváló szoftverekbe beépített „robotpilóta” funkciók, illetve az általunk készített billentyű makrók sem gyorsítanak érdemben, ráadásul ezeknél gyakoribb a hibázás vagy a lefagyás esélye, ami miatt újra kell kezdeni az egészet. A problémák másik része az archiváló, az indexelő és a megjelenítő eszközök hiányosságai és inkompatibilitásával magyarázható. Rengeteg ilyenbe futottunk már bele, a magyar ékezeteket tartalmazó URL címeiktől kezdve, a Facebook-ról készített WARC fájlok indexelhetetlenségén át, a lementett, de lejátszhatatlan videókig. Sajnos azt kell mondanunk, hogy a jelenlegi létszámmal és technológiával a közösségi média

¹³ OSZK-s webhelyek archívuma. Hozzáférés: 2021.03.20., <https://webarchivum.oszk.hu/oszk-s-archivum-kezdolap>



archiválásának mostani metodikája nem hatékony, nem robosztus, nem skálázható és nem fenntartható.

7. Útkeresés

A témában az utóbbi években megsaporodott szakcikk, az említett BESOCIAL felmérés előzetes eredményei, az IIPC konferenciáin és webináriumain elhangzó előadások azt mutatják, hogy más nemzeti könyvtárak és egyéb archiváló szervezetek is küzdenek ezzel a feladattal. Van, ahol még csak a tervezés vagy a kísérletezgetés szintjén tartanak, de akad köztük néhány olyan is, ahol már évtizedes tapasztalat gyűlt össze és van egy bejáratott rendszerük. Utóbbiak esetében az a jellemző, hogy nagyon megsűrik az archivált tartalmat, egy-két platformra koncentrálnak (leggyakrabban a Twitterre), és azokon belül is csak kiemelten fontos fiókokat vagy hashtag-eket mentenek. Az eredeti kezelőfelület megőrizhetőségének problémáját pedig néhol úgy kerülik meg, hogy csak a tényleges tartalmat töltik le API-n keresztül vagy valamilyen célszoftverrel, majd egy saját fejlesztésű megjelenítővel teszik böngészhetővé.

A honlapunk *Szakembereknek*¹⁴ menüpontja alatt elérhető forrásokban további információkat találhatnak az érdeklődők erről a speciális szakterületről is.

Bibliográfia

- Drótos László. „Az OSZK webarchívumának 2020-as újdonságai.” *Könyvtári Figyelő* 67, 1. sz. (2021), 31–38. Hozzáférés: 2021.05.10. <http://ojs.elte.hu/kf/article/view/2296>
- Drótos László, Németh Márton. „Egyedi mentésekre szolgáló webarchiváló szoftverek.” *Könyv, Könyvtár, Könyvtáros* 29, 12. sz. (2020), 3–11. Hozzáférés: 2021.05.10. <http://ki2.oszk.hu/3k/2020/12/egyedi-mentesekre-szolgalo-webarchivalo-szoftverek/>

¹⁴ OSZK Webarchívum honlapja – Szakembereknek menüpont. Hozzáférés: 2021.05.10., <https://webarchivum.oszk.hu/szakembereknek/>

A COVID járvánnyal kapcsolatos webarchiválási tevékenységek – nemzetközi és hazai körkép

Németh Márton
Országos Széchényi Könyvtár (Budapest)
nemeth.marton@oszk.hu

Abstract

The article describes the evolution of COVID-19 special web archive collection in Hungary. It offers an overview about the main conception, about searchability and usability issues that are appearing related to the collection and about the international environment of our activities. There is a major emphasis on those technical and human resource limits that take a major challenge to the full success of the project.

Keywords: Covid-19, web archiving, digital preservation, epidemic

Bevezetés

Ez a rövid tanulmány az Országos Széchényi Könyvtár Webarchiválási Osztálya által készített tematikus webarchívum gyűjtemény létrehozásának és fejlesztésének legfontosabb lépéseit és kihívásait foglalja össze. Ezt követően a nemzetközi háttér is ismertetésre kerül.

A szervezett webarchiválási tevékenység az OSZK E-Könyvtári Szolgáltatási Osztályának keretei között az Országos Könyvtári Platformhoz kötődő szakmai és informatikai fejlesztések keretében 2017-től indult el kísérleti projekt keretében. Jogilag (törvényben, illetve rendeletben szabályozott) szakmai alaptevékenységként 2021. január elsejétől tartozik a nemzeti könyvtár feladatkörébe.¹

A webarchívum gyűjteménye főként a magyarországi tudományos, oktatási, tudományos és közéleti webes forrásokat próbálja meg feltárni és archiválni. Alapvetően háromfajta webaratósi tevékenységet végzünk²:

- Általános pillanatfelvételek készítése a .hu domain alatti webtartalomról, illetve egyéb magyar nyelvű, illetve magyar vonatkozású webhelyekről.³
- Eseményalapú webarchiválás, a webes információforrások (blogok, hírportálok) címkék, rovatok szerint elkülöníthető tematikus tartalmak szerint (ebbe a kategóriába tartozik a Covid-19 gyűjtemény)⁴

1 Alapinformációk és adatok, Magyar Internet Archívum, hozzáférés: 2021.05.17., <https://webarchivum.oszk.hu/ujsagiroknak/alapinformaciok-es-adatok/>

2 Tájékoztató a honlapról. Magyar Internet Archívum, hozzáférés:2021.05.17., <https://webarchivum.oszk.hu/>

3 Webtér szintű aratások. Magyar Internet Archívum, hozzáférés:2021.05.17., <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/webter-szintu-aratasok/>

4 Esemény alapú aratások. Magyar Internet Archívum, hozzáférés:2021.05.17., <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/esemeny-alapu-aratasok/>



- Tematikus, műfaji, intézményi alapon gyűjteményekbe válogatott weboldalak, webhelyek archiválása⁵

A webarchívum jelenleg mintegy 45 TB adatot tartalmaz, több mint 35000 kiinduló címet (seed URL) gyűjtünk a tematikus és eseményalapú archiválás keretében, a magyar webtér általános aratása során pedig mintegy 250000 URL cím anyagából indulunk ki. Mintegy 700 Instagram profil, 500 Twitter csatorna illetve 1500 Facebook oldal archivált anyagai tartoznak jelenleg a közösségi média tartalmak terén a webarchívum gyűjteményébe.

A webarchívum gyűjteményeinek anyaga jórészt nyilvánosan még nem érhető el, mivel a könyvtár jelenleg az infrastrukturális és szolgáltatási rendszert érintő átalakuláson megy keresztül. A jövőben a szerzői jogi korlátozások miatt korlátozottan hozzáférhető tartalmak elérhetővé válnak a könyvtár olvasótermében. Három gyűjtemény azonban (legalább részben) online elérhető a nyilvánosság számára: a II. Rákóczi Ferenc-émlékév gyűjteménye⁶, a felhasználási engedély alapján szolgáltatható webhelyekből összeállított demó archívum⁷ és az Országos Széchényi Könyvtár online szolgáltatásainak archívuma⁸.

1. A COVID-19 gyűjtemény létrehozásának háttere

2020 elején a webarchívum munkatársai elkezdtek megvitatni, hogy milyen eseményalapú archiválási tevékenységeket folytassunk tovább és milyen új gyűjteményeket alakítsunk ki, amikor az első hír megjelent a médiában, hogy talán világméretű járvány kezdődött el. Mielőtt még Magyarországon is megjelent volna, úgy gondoltuk, hogy az éves tervünk részeként esetleg létre kell hoznunk egy eseményalapú gyűjteményt, a globálisnak ígérkező járvány magyar vonatkozásainak rögzítésére⁹. Elkezdtük tehát gyűjteni a főbb magyarországi hírportálokon lassanként megjelenő tematikus címkékhez kötődő tartalmakat. A pandémia kibontakozásával a téma már sok helyen önálló rovat, kategória szintjén is megjelent. Amennyiben a címkézési, kategorizálási gyakorlat konzekvensen lett a hírportálokon, s egyéb tartalomforrásokban kialakítva, az jelentősen megkönnyítette a dolgunkat. Természetesen mindez fordítva is igaz, ha nem volt lehetséges a járvánnyal kapcsolatos információk elkülönítése az adott webes forrás kapcsán, akkor néhány kiemelkedően fontos egyedi hír kivételével annak gyűjtéséről le kellett mondanunk. A látóköreinkbe került címeket heti rendszerességgel mentjük. A kezdetektől törekedtünk arra, hogy a magyar nyelvű határon túli információforrásokat is igyekezzünk számba venni. Így került látóköreinkbe a hivatalos szlovákiai magyar nyelvű információs oldal,

5 Tematikus és műfaji aratások. Magyar Internet Archívum, hozzáférés:2021.05.17., <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/tematikus-aratasok/>

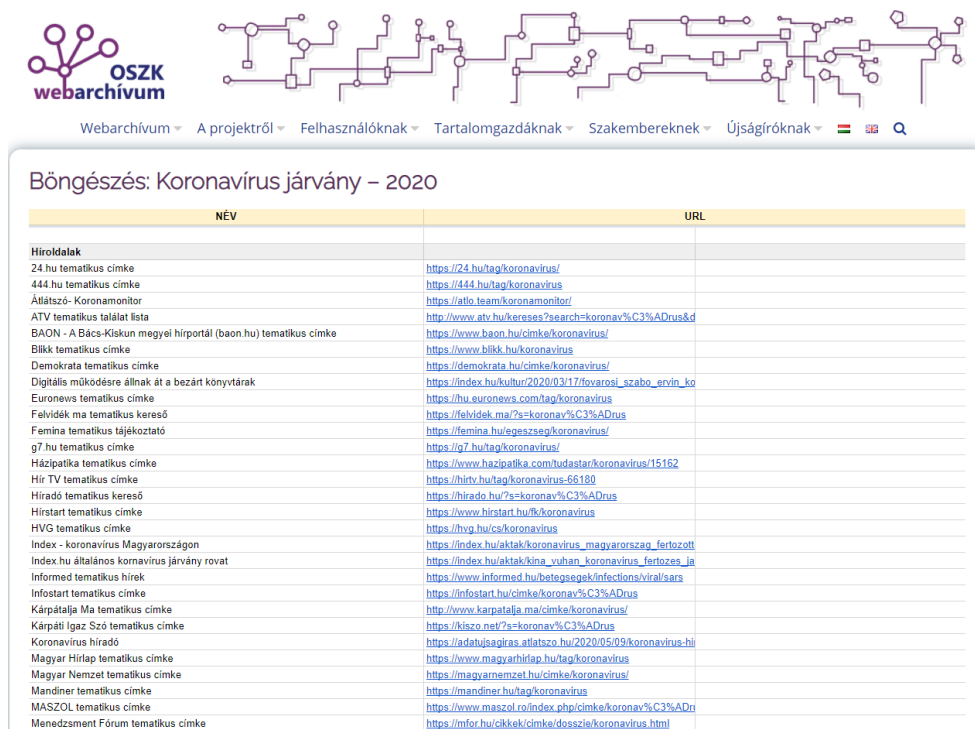
6 Rákóczi- gyűjtemény. Magyar Internet Archívum, hozzáférés:2021.05.17., <http://rakoczi2019.webarchivum.oszk.hu/>

7 Demó Archívum, Magyar Internet Archívum, hozzáférés:2021.05.17., <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/demo-kezdolap/>

8 OSZK-s webhelyek, Magyar Internet Archívum, hozzáférés.2021.05.17., <https://webarchivum.oszk.hu/webarchivum/reszgyujtemenyek-szerint/oszk-s-archivum-kezdolap/>

9 Bőngészés: Koronavírus járvány – 2020, Magyar Internet Archívum, hozzáférés.2021.05.17., <https://webarchivum.oszk.hu/webarchivum/bongesz/bongesz-az-esemeny-alapu-gyujtemenyekben/bongesz-koronavirus-jarvany-2020/>

illetve az RMDSZ által üzemeltetett nem hivatalos magyar nyelvű webhely is. Olyan kuriózumokra is akadtunk, mint a Török Rádió magyar nyelvű tematikus webes hírrovata. A teljes gyűjtemény seed listája (az archiválás kiindulópontjaként szolgáló URL-címek listája) elérhető a webarchívum weboldalán. A munkánk során tehát így 129 seed címről 56 aratást végeztünk 435 GB összméretben, az archivált tartalmakat pedig tömörített formában 435 WARC formátumú konténerfájlban tároljuk a 2021. márciusi állapot szerint. (1. ábra)



NÉV	URL
Híroldalak	
24.hu tematikus címke	https://24.hu/tag/koronavirus/
444.hu tematikus címke	https://444.hu/tag/koronavirus
Átlátszó- Koronamonitor	https://atlo.team/koronamonitor/
ATV tematikus találat lista	http://www.atv.hu/kereses?search=koronav%C3%ADrus&d
BAON - A Bács-Kiskun megyei hírportál (baon.hu) tematikus címke	https://www.baon.hu/cimke/koronavirus/
Blikk tematikus címke	https://www.blikk.hu/koronavirus
Demokrata tematikus címke	https://demokrata.hu/cimke/koronavirus/
Digitális működésre állnak át a bezárt könyvtárak	https://index.hu/kultur/2020/03/17/fovarosi_szabo_ervin_ko
Euronews tematikus címke	https://hu.euronews.com/tag/koronavirus
Felvidék ma tematikus kereső	https://felvidek.ma/?s=koronav%C3%ADrus
Femina tematikus tájékoztató	https://femina.hu/egeszseg/koronavirus/
g7.hu tematikus címke	https://g7.hu/tag/koronavirus/
Házipatika tematikus címke	https://www.hazipatika.com/tudastar/koronavirus/15162
Hír TV tematikus címke	https://hirtv.hu/tag/koronavirus-66180
Híradó tematikus kereső	https://hirado.hu/?s=koronav%C3%ADrus
Hírstart tematikus címke	https://www.hirstart.hu/%/koronavirus
HVG tematikus címke	https://hvg.hu/cs/koronavirus
Index - koronavírus Magyarországon	https://index.hu/aktak/koronavirus_magyarorszag_fertozott
Index.hu általános koronavírus járvány rovat	https://index.hu/aktak/kina_vuhan_koronavirus_fertozes_la
Informed tematikus hírek	https://www.informed.hu/betegsegek/infections/viral/sars
Infostart tematikus címke	https://infostart.hu/cimke/koronav%C3%ADrus
Kárpátalja Ma tematikus címke	http://www.karpatalja.ma/cimke/koronavirus/
Kárpáti Igaz Szó tematikus címke	https://kiso.net/?s=koronav%C3%ADrus
Koronavírus híradó	https://adatujsgiras.atlatso.hu/2020/05/09/koronavirus-hi
Magyar Hírlap tematikus címke	https://www.magyarhirlap.hu/tag/koronavirus
Magyar Nemzet tematikus címke	https://magyarnemzet.hu/cimke/koronavirus/
Mandiner tematikus címke	https://mandiner.hu/tag/koronavirus
MASZOL tematikus címke	https://www.maszol.ro/index.php/cimke/koronav%C3%ADru
Menedzsment Fórum tematikus címke	https://mfor.hu/cikkek/cimka/dosszie/koronavirus.html

1. ábra: A Covid-19 seedlista részlete

Ezt a speciális gyűjteményt nem csak URL-címek alapján lehet böngészni. Az anyaga teljesszövegű indexelésre került a dán fejlesztésű SolrWayback szoftver segítségével. (2. ábra) Ez a szoftver különféle statisztikai módszereket is biztosít a begyűjtött fájlok elemzésére (3. ábra).

2. A COVID-19 gyűjtemény építésének és gondozásának főbb kihívásai

A közösségi média archiválásának problémáival e tanulmánykötetben Drótos László foglalkozik egy másik tanulmányban részletesebben. Itt elég most csak arra utalni, hogy ezen információforrások gyűjtése nem automatizálható. Fokozottan ki vagyunk szolgáltatva a tartalomfejlesztőknek, akik rengeteg archiválási nehézséget tapasztalnak az alapvető tartalomszerkezet nem archivumbarát kialakítása, s állandó módosítása révén. Az archivált anyag megjelenítésekor is rengeteg hibába futunk bele.

Sajnos a rendelkezésre álló munkaerő hiánya miatt a gyűjtőmunka minőségbiztosításának gondozására sem jut elég energia. A hibásan archivált webhelyeket más beállításokkal, illetve más szoftverekkel újra kellene menteni, fel kellene tárni egyedi webhely szinten a legjobb megoldásokat.



SOLRWAYBACK

Search results for "astrazeneca vérrög" in SolrWayback. The interface shows search options, facets, and three result cards. A visualization of search results by domain is also provided.

Facets:

- domain
- hazipatika.com (1,411)
- mandiner.hu (969)
- 444.hu (254)
- hvg.hu (183)
- origo.hu (145)
- pestisracok.hu (135)
- nepszava.hu (116)
- egeszsegkalauz.hu (109)
- webbeteg.hu (78)
- 24.hu (63)
- promotions.hu (49)
- infostart.hu (47)
- koronavirus-kisokos.eu (47)
- blikk.hu (46)
- hirstart.hu (31)
- ripost.hu (31)
- vg.hu (29)
- portfolio.hu (28)
- nlc.hu (25)
- euronews.com (22)
- telex.hu (22)
- azonnali.hu (18)
- mfor.hu (17)
- informed.hu (15)
- wikipedia.org (14)
- magyarhirlap.hu (8)
- qubit.hu (8)
- napl.hu (7)
- vasarhely24.com (7)
- magyarnemzet.hu (6)
- parameter.sk (6)
- uj szo.com (6)
- felvidek.ma (5)
- noizz.hu (5)
- intima.hu (4)
- life.hu (4)
- merce.hu (4)

RESULTS
Showing 0 - 20 of 2,278 unique entries matching query (total hits: 3,997).

Result #1: ASTRAZENECA VÉRRÖG | 24.HU
Type: html, web page @ 24.hu
Date: 13/04-2021
Url: https://24.hu/tag/astrazeneca-verrog/
Highlighted content: "vérrög Ausztrál AstraZeneca-méreg: 700 ezer oltott, két vérrögösödés Ausztráliában sem tűnik kevésbé"

Result #2: AZ ASTRAZENECA-VAKCINA ÉS A VÉRRÖGKÉPZŐDÉS: VILÁGOS A KAPCSOLAT :: EGÉSZSÉGPOLITIKA - INFORMED ORVOSI ÉS ÉLE TMÖD PORTÁL :: KORONAVÍRUS JÁRVÁNY, ASTRAZENECA, VÉRRÖG, THROMBÓZIS
Type: html, web page @ informed.hu
Date: 13/04-2021
Url: https://www.informed.hu/egpo_new/az-astrazeneca-vakcina-es-a-verrogekpezodes-vilagos-a-kapcsolat-242464.html
Highlighted content: "Koronavírus Járvány, AstraZeneca, vérrög, thrombózis 2021-04-07 13:54:17 Kapcsolódó cikkek Az AstraZeneca"

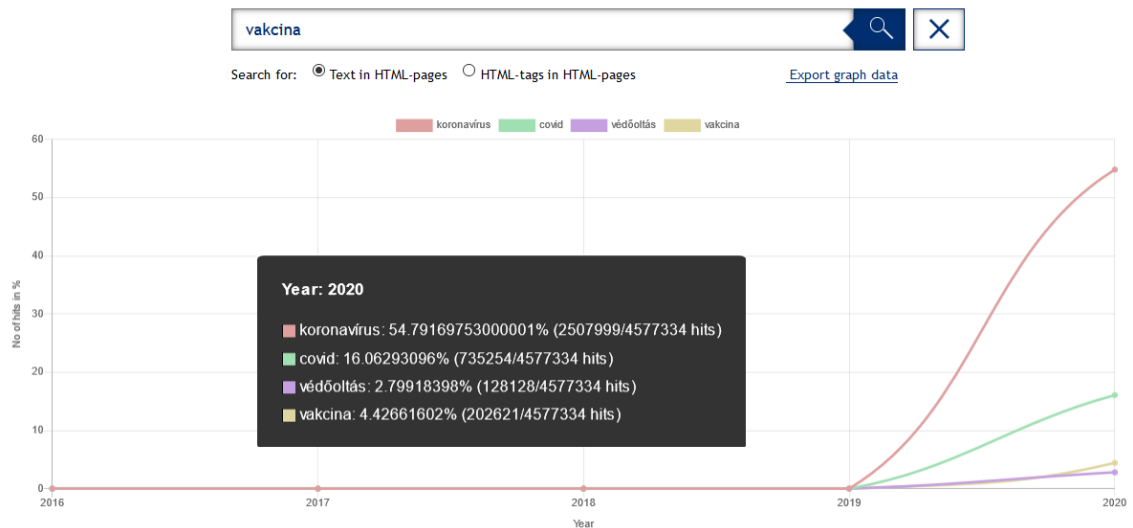
Result #3: ASTRAZENECA ÚJABB KÖZLEMÉNYE AZ OLTÓANYAGRÓL - NLC.HU
Type: html, web page @ nlc.hu
Date: 16/03-2021
Url: https://nlc.hu/ezvan/20210315/gyarto-ujabb-kozlemeny-astrazeneca-oltas-verrogekpezodes/

VISUALIZATION OF SEARCH RESULT BY DOMAIN

2. ábra: Keresés a gyűjtemény teljes szövegében

AVAILABLE TOOLS: Wordcloud Link graph Domain stats Link graph Gephi export Ngram Netarchive

NGRAM NETARCHIVE



3. ábra: A SolrWayback rendszerrel készített szógyakorisági görbék

A másik fő kihívást a hozzáférés szerzői jogi korlátai jelentik. Miközben a gyűjtemény seed listájának anyaga nyilvánosan hozzáférhető, a begyűjtött anyagnak csupán azok az elemeit szolgáltathatjuk nyilvánosan, melyekre engedéllyel rendelkezünk az adott tartalomtulajdonosától. Ennek ügyintézésének is nagyon jelentős munkaerőszükséglete, az adminisztratív munkaterhe. Némi javulást e téren újabb webkurátor munkatárs(ak) bevonása hozhat majd a későbbiekben.

A fentebb vázolt a közösségi média tartalomszerkezetből, illetve munkaerőhiányból levezethető nehézségek azt is eredményezik, hogy nem tudunk olyan átfogó, megfelelő minőségű gyűjteménnyel beilleszkedni a nemzetközi együttműködés keretei közé, melyet amúgy szakmai képességeink lehetővé tennének. Szerencsére a gyűjtemény különféle nemzetközi összehasonlító jellegű hasznosítási lehetőségeinek feltárására különféle partnereink segítségével így is mód nyílik, a továbbiakban erről adunk rövid áttekintést.

3. A COVID-19 gyűjtemény nemzetközi hasznosítási lehetőségei

A webarchiválást végző nemzeti közgyűjtemények, illetve egyéb intézményi szereplők ernyőszervezetekenként szolgáló International Internet Preservation Consortium (IIPC)¹⁰ keretei között a Content Development Working Group¹¹ (tartalomfejlesztési munkacsoport) keretében zajlanak a kiemelt nemzetközi eseményalapú gyűjtemények megalapozásához és kezeléséhez szükséges munkálatok. Az OSZK tagja a konzorciumnak, s korábbi nemzetközi eseményalapú gyűjtési projektekbe is bekapcsolódtunk már a nemzetközi munkacsoport felhívásai szerint. Így a mintegy egy-két hónappal a saját gyűjtési tevékenységünket követően meg tudtuk osztani folyamatosan a COVID-19 járványra vonatkozó seed listánk tartalmát is egy átfogó nemzetközi táblázat keretei között, számos európai nemzeti könyvtárral közös platformra kerülve így. Az Internet Archive infrastruktúrájára is támaszkodó munkacsoport saját mentéseket készít a begyűjtött címekről és átfogó gyűjteményt alakít ki. Az archiválási munka 2020. január végén indult az addig összegyűlt anyaggal, a nemzetközi seedlista ugyanazon év március végétől vált nyilvánossá, az egyes tételekhez kötődő legfontosabb metaadatokkal együtt. 2021 márciusáig több mint tizenkétezer seed URL cím került be a gyűjteménybe, több mint harmincmillió webes dokumentum került begyűjtésre, 4,1 TB terjedelemben (4. ábra). A begyűjtött anyag az Internet Archive kereskedelmi üzletágának az Archive-IT-nak a szerverein kerül tárolásra¹² (5. ábra).

10 International Internet Preservation Consortium, hozzáférés: 2021.05.17. <https://netpreserve.org/>

11 Training Working Group. International Internet Preservation Consortium, hozzáférés: 2021.05.17. <https://netpreserve.org/about-us/working-groups/training-working-group/>

12 Novel Coronavirus (Covid19) international collection of the IIPC, Hozzáférés, 2021.05.17. <https://archive-it.org/collections/13529>



About the IIPC Covid-19 Collection

- A project of the IIPC Content Development Group
- Capture tools: Heritrix via Archive-It (and Brozzler, selectively)
- Start date: **January 29, 2020**
- **Collection made public March 9, 2020 with 2,011 archived seeds**
- # of seeds March 9, 2021: **12,283**
another 500 seeds that are in process of capture or were not capturable
- # of documents at present: **30,394,204**
- Current data size: **4.1 TB**

<https://archive-it.org/collections/13529>

Updated 9 March 2021

4. ábra: IIPC Covid-19 gyűjteményének adatai

5. ábra: IIPC Covid-19 gyűjtemény az Archive-IT szerverén

Az archivált anyag tudományos, kutatási célú hasznosítására két szálon is formálódna tervek. Az egyik elképzelés szerint az Alexandriai Könyvtár (Egyiptom) koordinálásával egy grálapadú, vizuális megjelenítő felületet is felvonultató szemantikus adatbázis kialakítása kezdődött el, a szintén az IIPC által finanszírozott LinkGate projektben fejlesztett szoftverkörnyezetre alapozva.

A másik kutatási célú hasznosítási elképzelés a gyűjtemény tartalmának szöveg és adatbányászati vizsgálatára irányul a kutatók és webarchiválásért felelős közgyűjteményi szakemberek által közösen gondozott WarcNet¹³ projekt keretei között. Ezt az Aarhusi Egyetemről, Dániából hangolják össze. Az első konkrét eredmények 2021. nyarára várhatóak. Terveink szerint a könyvtári szaksajtóban a későbbiekben erről is hírt adunk majd. A projekt hangsúlyos célja továbbá az egyes európai országok COVID 19 gyűjteményépítési gyakorlatának összehasonlító jellegű feltárása is¹⁴. Ennek keretében számos interjút készítettek a gyűjteményeket gondozó szakemberekkel, többek között e sorok írójával is az OSZK gyűjteményének kapcsán¹⁵. Az interjúk a projekt honlapján kerültek publikálásra. Egyedülálló lehetőséget kapunk így arra is, hogy össze tudjuk hasonlítani számos európai ország tematikus archiválási gyakorlatát és tág képet kapunk az ennek kapcsán jelentkező kihívásokról az egyes nemzeti intézmények szintjén, illetve összehasonlító jelleggel egyaránt.

Epilógus

A gyűjtőmunka jelenleg is zajlik még, mert a járvány még nem zárult le. Annyit azonban megállapíthatunk, hogy a fent vázolt munkaerő kapacitás korlátok ellenére a webarchívumunk keretei között egy értékes tematikus gyűjteményt tudtunk kialakítani. A gyűjtőmunka során pedig értékes nemzetközi együttműködéseknek is részesévé tudtunk válni.

Ábrák jegyzéke

1. A Covid-19 seedlista részlete
2. Keresés a gyűjtemény teljes szövegében
3. A SolrWayback rendszerrel készített szógyakorisági görbék
4. IIPC Covid-19 gyűjteményének adatai
5. IIPC Covid-19 gyűjtemény az Archive-IT szerverén

13 WARCnet, Hozzáférés, 2021.05.17. <https://cc.au.dk/en/warcnet/>

14 WARCnet Papers. Warcnet, Hozzáférés, 2021.05.17.
<https://cc.au.dk/en/warcnet/warcnet-papers/>

15 Geeraert, Friedel & Németh, Márton, Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary, Warcnet, Hozzáférés 2021.05.17.
https://cc.au.dk/fileadmin/user_upload/WARCnet/Geeraert_et_al_COVID-19_Hungary.pdf



A könyvtárak az új digitális világban

Csík Tibor

Eötvös Loránd Tudományegyetem Könyvtár- és Információtudományi Intézet

csik.tibor@btk.elte.hu

ORCID [0000-0003-2061-1506](https://orcid.org/0000-0003-2061-1506)

Abstract

Tibor Csík: Libraries in the new digital world

According to Umberto Eco, books share their fates with their readers. Carrying the idea forward, not only books but also libraries share their fates with their readers. The state of a society is well illustrated by the state of its libraries and how to access information resources. The aim of the presentation is to examine the impact of libraries and their services on the neoliberal economic environment and the expanding process of commodification. Nicholas Negroponte (MIT) stated in 2010 that the physical book will be dead in five years. Although his provocative forecast has not been confirmed, the ratio of online resources in the acquisition of libraries is increasing. Technology companies have approached libraries to digitize their printed collection. Then the digital contents were sold, or used in accordance with their commercial interests to expand the advertising network. Information companies offered e-journals in large bundles for sale to libraries. Despite their promise, the price of digital publications has not become cheaper, but prices have risen steadily. The economic crisis of the early 2000s led to the closure of many public libraries. However libraries cannot become an information soup kitchen, where equal access to data and information hides the fact that there is inequality in access to meaningful information or important knowledge. The tasks of teaching library is not only to teach library and research skills, navigation on web, information retrieval from databases and how acting a law-abiding information consumer. Libraries can teach about information production and current information economics in part of information literacy. Libraries play a role in electronic publishing through their digital collections and repositories.

Keyword: library services, content industry, digitisation, e-lending, commodification



Umberto Eco A rózsza neve című művében található az a gondolat, mely szerint a könyvek osztoznak olvasók sorsában. Továbbíve a gondolatot az olvasók sorsában nemcsak a könyvek, hanem a könyvtárak is osztoznak, tehát a társadalom viszonyait jól jellemezi könyvtárainak helyzete: hogyan lehet az információforrásokhoz hozzáférni, milyen tudástárak állnak a köz és a különböző közösségek rendelkezésére. Az ezredfordulóra végbement alapvető változások következtében, a kulturális és a szellemi élet intézményrendszerében a redisztribúció és a reciprocitás egyre jelentéktelenebbé válik, miközben a kulturális és szellemi javak elérésében, megosztásában szélesedik a kommodifikáció folyamata.

A könyvtárak folyamatosan megújulva, a digitális technológiák fölhasználásával akarták szolgálni a közösségek, illetve a társadalom egészének információellátását. A tartalom-előállítás és -forgalmazás gazdasági modelljének átalakulásával a kiadók és az információtechnológiai cégek a könyvtárakra egyre inkább úgy tekintettek, mint kiaknázandó forrásokra vagy a profitot csökkentő tényezőkre. Tanulmányunkban, két fontos terület, a könyvtárak gyűjteményében és szolgáltatásaiban végbement változásokat kívánjuk bemutatni. A téma összetettsége és a terjedelmi korlátok miatt a legmeghatározóbb szempontok tárgyalására törekedhetünk.

A tudásmegosztás szolgálatába állított technológia

A XIX. század végének, illetve a XX. század első harmadának, a társadalmi és a tudományos haladásban hívő gondolkodói megvalósíthatónak vélték a tudás egyetemes megosztását. Közülük is kiemelkednek a dokumentációs mozgalom megindítói, Paul Otlet (1868–1944) és Henri La Fontaine (1854–1943), akik 1895-ben kezdeményezték az emberi tudást magukban foglaló kiadványok egyetemes számbavételét (Repertoire bibliographique universel – RBU) és gyűjteménnyé szervezését.¹ 1930-as évekre a tudásmegosztás valamennyi formáját kidolgozták az „analóg” technológiára alapozva.

A szellemi javak egyetemes gyűjteményének eszméje ma is hat. Az ezredfordulót követő első évtized közepén kezdeményezték például a Digitális Világkönyvtár (World Digital Library – WDL) fölállítását, amely ma, több mint 190 ország közel 200 intézménye közreműködésével épül, s az UNESCO és a Library of Congress (Egyesült Államok) üzemelteti.² A nyílt tudomány (Open Science) koncepciójának kialakítása az Európai Unióban a stratégiai célok közé tartozik,³ az UNESCO-ajánlás a nyílt tudományról pedig éppen most van születőben.⁴ Valamennyi felfogás épít a könyvtárakra, számít a közreműködésükre.

1 Rayward, Boyd W. „Visions of Xanadu: Paul Otlet (1868-1944) and hypertext.” *JASIS*. 45. 4. sz. (1994) 235-250. Hozzáférés: 2021. 04 06 [https://doi.org/10.1002/\(SICI\)1097-4571\(199405\)45:4<235::AID-ASI2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199405)45:4<235::AID-ASI2>3.0.CO;2-Y); <https://web.archive.org/web/20051227184732/http://people.lis.uiuc.edu/~wrayward/otlet/xanadu.htm>

2 World Digital Library honlapja: <https://www.wdl.org/en/> Hozzáférés: 2021. 04. 06.

3 „Open innovation, open science, open to the world – a vision for Europe.” <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe> Hozzáférés: 2021. 04 06.

4 UNESCO. „Recommendation on Open Science.” <https://en.unesco.org/science-sustainable-future/open-science/recommendation> Hozzáférés: 2021. 04 06.

A digitális forma megtöltése tartalommal

A XX. századra a könyv metaforává vált, jelképezte a már meghaladott tudásrögzítési és megosztási technológiákat és a meghaladni kívánt ismereteket is. Thomas A. Edison, a feltaláló 1913-ban úgy ítélte meg, hogy a könyvek hamarosan elavulnak az állami iskolákban. Marshall McLuhan, a forrómédiák apostola, nemcsak a Gutenberg-galaxis végét hirdette meg, hanem a könyvet ugyanolyan mulandónak tekintette, mint a szabásmintát vagy a csíkot a női harisnyán. Nicholas Negroponte, a digitális létezésünk kutatója 2010-ben arról beszélt, hogy a könyv mint tárgy öt év múlva halott lesz.

A számítógépes technológia a könyvtárak életében az 1950-es években jelent meg. A szakirodalom visszakereshetővé tétele érdekében a bibliográfiákból, a katalógusokból, a referáló-indexelő kiadványokból számítógépes adatbázisokat építettek. Az orvostudomány szakirodalmát feldolgozó – National Library of Medicine (NLM US) készítette – Index Medicus kiadványból az 1960-as évek közepétől a bibliográfiai adatbázis vált, mely 1971-től online kereshető (MEDLINE), 1990-es évek második felétől a világhálón elérhető, s már teljes szöveget is tartalmazó adatbázissá alakították (PubMed). A fejlesztés mögött kormányzati stratégia állt, s bőséges szövetségi (pl. National Institutes of Health – NIH) és vállalati anyagi forrásokkal. Ennek eredményeként ma az Egyesült Államok szakirodalmi adatbázisai jelentik a vonatkoztatási pontot a tudományos élet számára.

Hazánkban is megindult a bibliográfiák, dokumentációs és referáló kiadványok stb. adatbázissá alakításának folyamata. Például az Országos Pedagógiai Könyvtár nemzeti szakkönyvtárjából (Magyar Pedagógiai Irodalom) és referáló lapjából (Külföldi Pedagógiai Információ) 1989-től számítógépes adatbázis épül, amely 2007-től folyóiratok hivatkozásait is feldolgozza. A magyar szakirodalmi adatbázisok – a tudományos szakkönyvtárak térvesztésével párhuzamosan – az 1990-es évektől folyamatosan hanyatlanak, szűnnek meg. Sikeres viszont az 1990-es évek második felétől a hazai könyvtárak katalógusainak adatbázissá szervezése és online elérhetővé tétele.

Digitalizálás

A könyvtárak remek lehetőséget láttak a digitális technológiában arra, hogy kiteljesítve küldetésüket a gyűjteményüket is mind szélesebb kör számára tegyék elérhetővé. A digitalizálás tömegessé válása és az abból profitot termelő vállalkozások megjelenése azonban teljesen új helyzetet teremtett. Például a 2000-es évek elején az amerikai nagykönyvtárak anyagát (pl. Harvard Library, New York Public Library) a Google – a Google Print Program keretében – eddig nem látott mértékben digitalizálta, s tette a tartalmakat a szolgáltatása részévé. A Bibliothèque nationale de France élesen elutasította a „kedvező” digitalizálási lehetőséget, mert a francia kulturális örökség megőrzését és közzétételét nemzeti feladatnak tekintette. A Google digitális könyvtári



terveit az amerikai szerzői jogi (Authors Guild) és a kiadói szervezet (Association of American Publishers) bíróságon támadta meg.⁵

A könyvtárak az 1990-es évek közepétől aktívan bekapcsolódnak a digitális megőrzési programokba, melyek az ezredfordulót követően lendültek föl, s amelyeket támogat az UNESCO és a nemzetközi könyvtárosszervezet, az IFLA.⁶ Az Egyesült Államokban a nemzeti könyvtár (Library of Congress) tevével az 1990-es évek végén indult a folyamat, majd 2010-től a kongresszus támogatta nemzeti programmal folytatódott (National Digital Information Infrastructure and Preservation Program – NDIIPP), 2010-től az intézményi kört szélesítik (National Digital Stewardship Alliance – NDSA), 2016-tól már digitális könyvtárak szövetségében (Digital Library Federation – DLF) folyik a munka.⁷ A kis nemzetek könyvtári rendszere is nagyon lelkesítő példával tud szolgálni. Ilyen lehet a norvég nemzeti könyvtár (Nasjonalbiblioteket) stratégiája, mely online elérhető digitális könyvtár létrehozását célozza.⁸

A stratégiai tervezettséghez képest a másik végletet az állományvédelmi szükségintézkedésként megvalósuló digitalizáció képviseli, melyet csak úgy tud a könyvtár elvégeztetni, ha lemond a tartalom online terjesztéséről a digitalizáló cég javára. Az eredményes nemzeti program nem nélkülözheti a megbízható és hosszú távon alkalmazható technológiát, az intézmények széles körét átfogó, jól szervezett együttműködést, a digitalizálandó művek tervszerű kiválasztását és a folyamatos munkát lehetővé tevő, biztos pénzügyi háttérrel.

A digitális kiadványok és a könyvtárak, avagy az Amazonnak megvan a hatalma ahhoz, hogy a könyvtárakat a XX. században tartsa⁹

A könyvtárak beszerzésében nagy változást hoztak a hordozó nélküli, elektronikus/digitális dokumentumok. A költségvetésben az időszaki kiadványok beszerzésére fordított összegek emelkedtek a leginkább. A felsőoktatási könyvtáraknál – az amerikai

5 The Authors Guild. „Authors Guild v. Google: Questions and answers.” In: AG. Hozzáférés: 2021. 04 06. <https://www.authorsguild.org/authors-guild-v-google-questions-answers/>

6 Baucom, Erin. „A brief history of digital preservation.” In: Myntti, Jeremy – Zoom, Jessalyn (ed.): *Digital preservation in libraries. Preparing for a sustainable future*. Chicago: American Library Association (An ALCTS monograph) 2019. 3 –19. Hozzáférés: 2021. 04 06. [Digital Preservation in Libraries: Preparing for a Sustainable Future \(An ALCTS Monograph\) | ALA Store](https://www.ala.org/advocacy/digital-preservation/digital-preservation-in-libraries-preparing-for-a-sustainable-future)

7 „Digital Preservation” (Library of Congress) Hozzáférés: 2021. 04 06. <https://www.digitalpreservation.gov/>

8 „A national library for everyone. Digitisation of the collection is creating new opportunities for public enlightenment and widespread access to knowledge and information. Through our online digital library, Nasjonalbiblioteket aims to offer a modern, user-focused service providing democratic access to our collection. Nasjonalbiblioteket is putting in place agreements with rights holders to allow us to make copyright-protected material available as well.” Hozzáférés: 2021. 04 06. <https://www.nb.no/en/strategy>

9 „...Amazon to have the power to unilaterally force libraries to stay in the 20th century. It’s a price we pay for letting Big Tech get so big.”
Fowler, Geoffrey A. „Want to borrow that e-book from the library? Sorry, Amazon won’t let you.” *The Washington Post*. March 10, 2021. Hozzáférés: 2021. 04 06. <https://www.washingtonpost.com/technology/2021/03/10/amazon-library-ebook-monopoly/>

Association of Research Libraries (ARL) statisztikai adatai szerint – a beszerzési kiadásokban az időszaki kiadványokra fordított összeg 1998 és 2018 között 166%-kal nőtt, miközben a monográfiák esetében csak 23%-kal.¹⁰

Az elektronikus könyvek könyvtári kölcsönzését a kiadók határozzák meg, megszabva a kölcsönözhető művek körét, hogy a könyvet hány olvasó kaphatja meg egyidejűleg, hányszor és mennyi ideig lehet kölcsönadni, és hogy a mű bekerülhet-e a könyvtár állományába vagy törlődik a kínálatból stb. Az amerikai könyvtárosszervezet (American Library Association – ALA) hosszú évek óta küzd az öt legnagyobb kiadóval (Big Five)¹¹, hogy a könyvtárak korlátozásoktól mentesen kölcsönözhessenek elektronikus könyveket, hasonlóan a fizikai hordozókhoz. A kiadók láthatólag az új technológia valamennyi előnyét a maguk javára kívánják fordítani, sőt, még a törvényességen is túllépnek (pl. árkartell).¹² A könyvtárosszervezet ezért az #eBooksForAll keretében támogatókat gyűjt, és kész jogi eljárást kezdeményezni a digitális piac monopolizálása ellen (antitrust).¹³

Az Amazonnál például 2011-ben indult Kindle Lending Library, mely 2021 januárjában megszűnt, de helyette ajánlja a cég a 2016-tól működő Prime Reading szolgáltatását több mint 1000 könyvvel és folyóirattal, vagy a 2014-es alapítású Kindle Unlimited¹⁴ előfizetését (havi 9,99\$), mely kétfélmilliónál több művet kínál (könyvek, folyóiratok, hangskönyvek stb.). Avállalatnak természetesen van hasonló zenei és videós szolgáltatása, s használt könyveket is forgalmaz. Az Amazon Publishing, melyhez több könyvkiadó tartozik, nem adja el a könyvtáraknak az e-könyveit, megakadályozva azok könyvtári kölcsönzését. A vállalat kinyilvánított célja olyan – úgymond – megközelítés kidolgozása, amely egyrészt az olvasóközönség bővítését és a könyvtárhasználók szolgálatát, másrészt a szerzők érdekeinek érvényesülését és a jogdíjakat, jövedelmeket egyaránt biztosítja.¹⁵

Online szolgáltatások

Akönyvtári rendszernekkülönbözhetetlen szolgáltatásai közé tartozik a közös-, illetve központi katalógus, amely tartalmazza a lelőhelyadatokat is. A könyvtárak együttműködésével készülő leghíresebb, számítógépes közös katalógus az Ohio College Library Center

10 „ARL Statistics Survey Statistical Trends.” In: *Association of Research Libraries*. Hozzáférés: 2021. 04 06. <https://www.arl.org/arl-statistics-survey-statistical-trends/>

11 Az öt nagy kiadó: Hachette, HarperCollins, Macmillan, Simon & Schuster és a Penguin Random House

12 Albanese, Andrew „Amazon hit with e-book price fixing suit.” *Publisher Weekly*. Jan 14, 2021 Hozzáférés: 2021. 04 06. <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/85318-amazon-hit-with-e-book-price-fixing-suit.html>

13 American Library Association. „The State of America’s Libraries 2020. A report from the American Library Association.” Steve Zalusky, ed. 2020. 24. Hozzáférés: 2021. 04 06. <http://www.ala.org/news/state-americas-libraries-report-2020>

14 „Kindle Unlimited” (Amazon) Hozzáférés: 2021. 04 06. https://www.amazon.com/gp/feature.html?docId=1002872331#faq/ref=insider_ar_reading_primereading

15 Fowler (2021)



(OCLC) 1967-ben indult, s ma 107 országból való 15.637¹⁶ tagintézményének munkáját összefogva készíti és szolgáltatja a WorldCat nevű katalógust, mely közel 500 millió bibliográfiai rekordot tartalmaz.

Ohio államban a könyvtárak együttműködéseként 1992-ben jött létre az Ohio Library and Information Network (OhioLINK)¹⁷, mely 1997-től a világhálón működik. Az OhioLINK küldetésében kertilés nélkül, világosan megfogalmazták a célt: az állam felsőoktatása számára teremtsen versenyelőnyt azzal, hogy együttműködve és költséghatékonyan megszerzi és rendelkezésre bocsájta a folyamatosan gyarapodó, nyomtatott és digitális, tudományos forrásokat, a tanítás, a tanulás és a kutatás előrehaladása, továbbá Ohio tudásalapú gazdaságának növekedése érdekében.¹⁸

A hazánkban 2001-ben 15 könyvtár közreműködésével létrejött a Magyar Országos Közös Katalógust (MOKKA). A MOKKA az elmúlt húsz évben számos területen ért el igen komoly eredményeket, de az OCLC-hez hasonló, az állam gyűjteményeit oly teljességgel átfogó közös katalógust nem sikerült kiépíteni. A tudás- és információmegosztás érdekében létrehozott könyvtári hálózat, amelynek keretében együttműködnek az ország tudományos intézményei, ugyancsak várat magára.¹⁹

A tudományos szakkönyvtárak és a felsőoktatási könyvtárak kiemelt szolgáltatásai közé tartozott az 1970-es évektől az online adatbázisokban való szakértői keresés. Az 1990-es évek végére a nagy szakirodalmi adatbázisok kialakították a webes felületüket, aminek következtében a könyvtárosok elsődleges feladata azt lett, hogy a kutatókat és a hallgatókat fölkészítsék az adatbázisban való keresésre. Ezek a kurzusok a felsőoktatásban fokozatosan kötelezővé váltak, a kutatástámogatás viszont egyre szűkebb, de szakmailag és tudományosan sokkal igényesebb szolgáltatássá vált.

Az internetes majd webes források közötti eligazodást már a kezdetektől segíteni kívánták a könyvtárosok. A meglehetősen különböző forrásokat, weboldalakat megpróbálták könyvtárrá szervezni: minőségi kritériumok szerint szigorúan átválogatva, s alapos formai és tartalmi feldolgozással lehetővé tenni a sokszempontú keresést. Az Internet Public Library (IPL) például a michigani egyetem könyvtároshallgatói munkájaként indult 1995-ben.²⁰ A webböngészők versenye és az anyagi támogatás megszűnése miatt az IPL

16 OCLC. „About.” Hozzáférés: 2021. 04 06. https://www.oclc.org/en/about.html?cmpid=md_ab

17 Mares, Halle. „OhioLINK’s 25th anniversary 1992–2017. A concise history.” [2017] Hozzáférés: 2021. 04 06. <https://www.ohiolink.edu/sites/ohiolink.edu/files/uploads/25th%20Anniversary%20History%20Booklet.pdf>

18 „OhioLINK Mission.” In: https://www.ohiolink.edu/content/ohiolink_mission

19 Vö. Az Elektronikus Információszolgáltatást Nemzeti Program (EISZ) nemzeti licenc vásárlásával nyújt hozzáférést elektronikus tudományos tartalmakhoz az intézmények meghatározott köre számára. <http://eisz.mtak.hu/index.php/hu/informaciok.html>

20 Abels, Eileen G. „Transforming the internet public library into the ipl2 virtual learning laboratory.” *Reference Librarian*, 52. 4.sz. (2011) 284-290. doi:<http://dx.doi.org/10.1080/02763877.2011.586922>

2015-ben befejezte működését.²¹ A vállalkozásnak azonban olyan jól csengő neve volt, hogy a Barnes & Noble Education megvette, de a szolgáltatás tartalmát megváltoztatva, 2020-tól a diákok számára több mint fél millió esszét kínáló portálként működteti.

Hazánkban az Országos Pedagógiai Könyvtár tervezte, hogy magyar iskolai könyvtárak munkáját egy közösen épített Web Iskolai Könyvtárral (WebIK) segíti. A végül pályázati pénzből megvalósult WebIK 2010 és 2015 között működött, s hasonlóképpen a pénzhiány miatt maradt abba. Ma már ilyen átfogó munkára a könyvtárak nem vállalkoznak, de minden rangos felsőoktatási intézmény könyvtára állít össze listát az egyetemi polgárok számára az online elérhető adatbázisokról, tematikus portálokról, jól használható forrásokról.

A könyvtári szolgáltatások iránti igény megváltozása

A tájékoztató szolgálatot – az ARL adatai szerint – 2018-ban 1998-hoz képest 84%-kal kevesebben keresték meg, és a kölcsönzések száma is több mint kétharmad résszel esett vissza. A könyvtárközi kölcsönzés 2009-ig folyamatosan nőtt (59%), majd ekkortól csökkent, és 2018-ra már csak 15%-kal haladta meg a húsz évvel korábbi szintet. A hallgatónak tartott csoportfoglalkozások száma viszont látványosan nőtt, 2016-ra 88%-kal, 2018-ban pedig 67%-kal haladta meg a kiinduló szintet, s a résztvevők száma pedig folyamatosan gyarapodott (41%).²²

A tájékoztatás iránti igény csökkenéséhez hozzájárult, hogy az adatbázis-forgalmazók kifejlesztették az ilyen irányú szolgáltatásaikat: használóképzés (személyes, online, Youtube-anyagok, webinárium stb.), tájékoztató szolgáltatás (telefon, ímél, cset, közösségi háló stb.), sőt tudományos konzulens rendelkezésre bocsátása. Az óriásvállalatok mindegyike olyan portál fejlesztésén dolgozik, amelyről a legkülönbözőbb referenz és teljes szövegű adatbázisok éppúgy elérhetők, mint tudományos hírek, rendezvények, virtuális közösségi terek, ingyenes források stb., de igénybe lehet venni a használatot messzemenően támogató szolgáltatásokat is. Mindez jelentősen szűkítette a könyvtárak adatbázis-keresésben és a kutatástámogatásban betöltött szerepét.

A könyvtári foglalkozások, a könyvtárak oktatási tevékenységei iránt viszont nőtt az igény. Ez összhangban áll azzal a folyamattal, melynek során az információs műveltség (information literacy) megszerzése a felsőoktatás valamennyi szintjén elvárás lett. A képzés egészét átfogó szabályozás 2000-től indult,²³ majd elkészült tudományáganként

21 Janes, Joseph. „A fond farewell to the Internet Public Library.” In: *American Libraries*. RIP IPL. Posted November 13, 2014 Hozzáférés: 2021. 04 06.

<http://www.americanlibrariesmagazine.org/article/rip-ipl>

22 „ARL Statistics Survey Statistical Trends.” In: *Association of Research Libraries*. Hozzáférés: 2021. 04 06. <https://www.arl.org/arl-statistics-survey-statistical-trends/>

23 „Information literacy competency standards for higher education.” 2000. American Library Association. Association of College and Research Libraries. Hozzáférés: 2021. 04 06.

<https://alair.ala.org/bitstream/handle/11213/7668/ACRL%20Information%20Literacy%20Competency%20Standards%20for%20Higher%20Education.pdf?sequence=1&isAllowed=y>



is,²⁴ 2006-ban pedig rendezték a könyvtárosok oktatói státuszát²⁵ a felsőoktatásban. Az információs műveltség átadásán túl igény mutatkozik a tudományterületek tájékoztató eszközei, illetve adatbázisai bemutatására, valamint a nyílt hozzáférésű platformok használatának és a publikációk megjelentetésének segítése iránt.

Tudományos közlemények nyílt adatbázisai a világhálón

A felsőoktatási és a tudományos szakkönyvtárak gyűjteményének és szolgáltatásainak igénybevételét jelentősen csökkenti, hogy egyre növekvő számú forrás érhető el szabadon az online térben. Például Paul Ginsparg fizikus 1991-ben létrehozta a tudományos közlemények nyílt hozzáférésű tárá, az arXiv-ot. Mára már a diszciplínák és a kutatási területek preprint-adatbázisai, -szerverei ugyanúgy a szakterület tudományosságának nélkülözhetetlen attribútumai közé tartozik, mint a XIX. században az akadémiai folyóiratok. A tudományos ismereteket előállító intézményekkel szemben általános elvárásá vált, hogy saját anyagaikból, nyílt hozzáférésű repozitóriumo(ka)t alakítsanak ki.

A világhálón föllelhető tudományos közlemények visszakeresésére a Google külön szolgáltatást indított 2004-ben Google Scholar néven. Az elérhető tartalmak szélsőséges bírálói szerint, a használó nem tudhatja, hogy a szolgáltatás jelmondatának megfelelően éppen egy óriás vállán áll vagy „halkan belelépett” (vö. ragadozó-folyóiratok). A Google Scholar közölte bibliometriai adatok azonban, összevetve a tudományosan elismert hivatkozási adatbázisok – Web of Science, Scopus – értékelésével relevánsnak bizonyultak.²⁶ Ez más cégeket is arra ösztönöz, hogy kiépítsék a saját vállalkozásukat. A Microsoft 2016-tól szállt be az üzletbe, s elindította a Google Scholarhoz hasonló Microsoft Academic nevű szolgáltatását.

Összegezve megállapítható, hogy a könyvtárak ellenőrzött, hiteles forrásaival szemben ott áll a világháló hihetetlen információbősége, a könyvtári tájékoztatással, kutatástámogatással szemben pedig a keresés szabadságának érzése és saját ítéletünkbe, tudásunkba vetett bizalom.

24 Például a természet- és mérnöki tudományok oktatásához: „Information literacy Standards for Science and Engineering/Technology”, American Library Association, July 24, 2006. <http://www.ala.org/acrl/standards/infolitscitech>. Hozzáférés: 2021. 04 06. Document ID: b65e0fb0-5ded-a6d4-2d06-d3cb068960af

25 „ACRL Standards for Faculty Status for Academic Librarians”, American Library Association, September 6, 2006. Hozzáférés: 2021. 04 06. <http://www.ala.org/acrl/standards/standardsfaculty> [2021. 04. 24] Document ID: 5cb594d4-66de-4ee4-159b-38e1e71288f8

26 Li, Jie, Burnham, Judy F., Lemley, Trey & Britton, Robert M. „Citation analysis: Comparison of Web of Science, Scopus, SciFinder, and Google Scholar.” *Journal of Electronic Resources in Medical Libraries*, 7. 3.sz. (2010). 196-217. doi: <http://dx.doi.org/10.1080/15424065.2010.505518>
Bar-Ilan, J. „Citations to the „introduction to informetrics” indexed by WOS, Scopus and Google Scholar.” *Scientometrics*, 82. 3.sz. (2010) 495-506.

Merre tartson a könyvtári tájékoztatás és oktatás?

Általánosan elfogadott, hogy az ismeretek megszerzéséhez és megosztásához kapcsolódó kompetenciák elsajátítása kulcsfontosságú, akár az információ (information literacy) akár a digitális technológia (digital literacy) felől közelítünk. Birger Hjørland fölhívja a figyelmet, hogy az információs műveltség és a digitális műveltség közös, központi eleme az ismeretek létrehozásának kritikai értelmezése és az a tudás, mely képessé tesz az információk megítélésére.²⁷

Az UNESCO a média- és információs műveltséget az alapvető emberi jogok részének tekinti, amely az információkhoz való megfelelő hozzáféréssel együtt a társadalmi egyenlőség érvényre juttatásának előfeltétele.²⁸ A szervezet nevesíti is a könyvtárakat, amelyek közreműködése nélkülözhetetlen ezen ismeretek átadásában, valamint a média- és tartalomipar működésének megismertetésében. Azt jelölve meg célként, hogy az információhoz, a tudáshoz való hozzáférésben az igazságosság elve érvényesüljön, mind a társadalmon belül, mind a globális világ társadalmi között.

A másik vonatkoztatási pont az ún. digitális intelligenciához tartozó kompetenciák, amelyeket az Organisation for Economic Co-operation and Development (OECD) közreműködésével az Institute of Electrical and Electronics Engineers (IEEE), a világ egyik legnagyobb mérnöki szervezete dolgozta ki.²⁹ A digitális intelligencia azokat a technikai, kognitív, metakognitív és szocio-emocionális kompetenciákat fogja át, melyek lehetővé teszik az egyénnek, hogy a digitális élet kihívásaival szembenézzen és alkalmazkodjon hozzájuk.³⁰ A meghatározásban szerepel, hogy ezek a kompetenciák egyetemes értékeken alapulnak, de részletesen nem fejtik ki, s csak az „oktatásipari” (educational industry) vonatkozásokra történik utalás, a könyvtárakhoz való viszonyra azonban nem.

Sokan tartanak attól, hogy a tartalomipar szolgálatába állított könyvtári tájékoztatás már nem a szellemi szabadság elvét fogja szem előtt tartani, mert a fogyasztó, adott termékkínálatból való választásának segítése lesz a feladata. A digitális kompetenciáknak pedig nem arra kell irányulniuk, hogy az egyén hogyan állítsa céljai szolgálatába a technológiát, hanem hogy alkalmazkodni tudjon digitális világhoz, elsajátítsa a jogszerű és etikus fogyasztói magatartást (vö. Aaron Swartz vs JSTOR).

27 Hjørland, Birger. „Information literacy and digital literacy.” *PRISMA.COM*. 7. (2008) Hozzáférés: 2021. 04 06. http://www.competenzedocenti.it/Documenti/competenze_digitali/information_literacy_and_digital_literacy_en_es.pdf

28 „Global media and information literacy assessment. Framework: country readiness and competencies.” Paris: UNESCO, 2013. Hozzáférés: 2021. 04 06. <https://unesdoc.unesco.org/ark:/48223/pf0000224655>

29 Park, Yuhyun „DQ. Digital intelligence. Global standards report 2019. Common framework for digital literacy, skills and readiness.” (s.l.) 2019. DQ Institute. Hozzáférés: 2021. 04 06. <https://www.dqinstitute.org/wp-content/uploads/2019/11/DQGlobalStandardsReport2019.pdf>

30 „IEEE 3527.1-2020 - IEEE Standard for Digital Intelligence (DQ)--Framework for Digital Literacy, Skills, and Readiness.” In: IEEE Standards Association. Hozzáférés: 2021. 04 06. https://standards.ieee.org/standard/3527_1-2020.html



Vajon a könyvtári oktatás feladata-e annak tudatosítása, hogy a böngészők megosztotta ingyenes tartalmaknak és kommunikációs szolgáltatásoknak, a közösségi kapcsolattartásnak mi az ára? Elvárható-e a könyvtártól annak bemutatása, hogyan válnak az egyén személyes adatai és a viselkedését leíró adatok áruvá, s miként használhatók azok a viselkedés előrejelzésére és módosítására, vagy akár az egyén gondolkodásának megváltoztatására? Ugyancsak kérdés, hogyan határozzák meg a XXI. század uralkodó eszméi a könyvtárak társadalmi szerepét, például a platformkapitalizmusról vagy a felügyeleti kapitalizmusról³¹ szóló, új elméletekben.

Záró gondolatok

A számítógépes adatbázisokat és a digitalizációt a könyvtárak abban a hitben támogatták, hogy ezzel – küldetésükhöz híven – a szellemi javakhoz való hozzáférést szolgálják. A 2000-es évek második évtizedére azonban egyértelművé vált, hogy az átadott javak nem a közt, hanem a tartalomipari és információtechnológiai cégeket szolgálják.

Láttuk, hogy a cégek az általuk forgalmazott digitális művek, illetve az online szolgáltatások könyvtári beszerzését és terjesztését egyre szigorúbban szabályozzák, sőt, korlátozzák. A fizikai hordozón forgalmazott művektől eltérő eljárásrendet, különös módon a szerzői jogok és ahhoz kapcsolódó jövedelmek védelmével szokták igazolni.

Mindezzel egyidejűleg ezek az óriásvállalatok a legismertebb könyvtári szolgáltatásokhoz – mint a kölcsönzés, tájékoztatás – hasonló szolgáltatásokat üzemeltetnek. A technológia révén tehát, új életre kelt a fizetős kölcsönzés, s egyre nagyobb teret nyernek az előfizetéses tartalomszolgáltatások és az ún. streaming media.

A kiadók és a forgalmazók – úgy tűnik – készek a legvégsőkig elmenni az új technológia adta lehetőségek kiaknázásában. És ami nem elég nyereséges vagy a rászorulókat segíti, azt kell közpénzből fenntartani. A könyvtárak azonban nem válhatnak információs népkonyhává, ahol az adatokhoz, információkhoz való egyenlő hozzáférés elrejti azt, hogy nincs egyenlőség, mert a jelentős információkhoz vagy a fontos ismeretekhez nem lehet hozzáférni.

Ha a szellemi örökségünk védelmét, a tudományos eredményeink számba vételét, a felsőoktatás és a kutatás információellátását, valamint a nemzet tudásvagyonával való gazdálkodást tekintjük, akkor föltehető a kérdés, hogy a globális óriásvállalatok vagy egy országos könyvtári rendszer lássa-e el a feladatokat. Ugyanennek a problémának egy másik vetülete a könyvtári oktatás tematikája: az információhoz, az ismeretekhez való hozzáférés álljon-e középpontban vagy pedig az, hogy az egyén a „digitális élet kihívásaival szembenézzen és alkalmazkodjon hozzájuk”.

Vajon az eszmék és gondolatok sokféleségének megismerésén, valamint a tudás lehetőleg a társadalom egésze számára hozzáférhetővé tételén alapuló könyvtárfelfogás érvényes lesz a 21. században?

³¹ Zuboff, Shoshana „The age of surveillance capitalism. The fight for a human future. At the new frontier of power.” New York: 2019. PublicAffairs

A földrajzi személyi hungarikum feldolgozásának problematikája

Ungváry Rudolf
Országos Széchényi Könyvtár; PIM Digitális Bölcsészeti Központ
ungvaryr@gmail.com

Bevezető

A magyar vagy magyar vonatkozású szerzőknek, csakúgy, mint külföldi társaiknak, van két alapvető földrajzi meghatározottsága: a születési és a halálozási helyük. Ezeknek azért van jelentőségük, mert arról tájékoztatnak, hogy honnan származik a szerző, illetve arról, hogy hol halt meg. A magyar és magyar vonatkozású szerzők esetében tárgyalt földrajzi nevek mind magyar vonatkozásúak, ezért hungarikumnak tekinthetők. Így bevezethetjük a *földrajzi személyi hungarikum* fogalmát. Ennek minősül minden olyan földrajzi hely, amely magyar vonatkozású személyek születési és halálozási helye, külföldi állomáshelye, hungarikum kiadási helye.

A feldolgozás nehézségei

A PIM Digitális Bölcsészeti Központban jelenleg az adatbázisban szereplő személyek helyadatainak a névtéri¹ feldolgozása folyik. Ennek lényeges része, hogy a forrásokban szereplő szerzők születési és halálozási helyneveit azonosítani kell:

- meg kell állapítani, hogy a forrásban szereplő földrajzi név a valóságban milyen formában létezik, és
- milyen magasabb igazgatási egységhez tartozik, végül pedig
- be kell hasonlítani a jelenleg épülő névtérbe, ill. annak földrajzinév-terébe, meg kell állapítani, hogy van-e megfelelője, ha van, akkor az milyen hivatalos alakban szerepel, ha pedig nincs, akkor föl kell venni²

A könyvtárak egységesített besorolási adataiban szereplő helynevek mind ellenőrzött, megbízható névalakok, ezért feldolgozásuk – elsősorban informatikai eszközökkel – jól támogatható. A PIM adatbázisában szereplő források túlnyomó része ezzel szemben hagyaték, kis része pedig nyomtatott könyv. A bennük szereplő földrajzi nevek az eddigi tapasztalatok alapján megbízhatatlanok.

A helynevek egy részének ma van hivatalos, államilag nyilvántartott nevük.³ A történelmi Magyarországnak (beleértve Horvátországot is) volt hivatalos helységnévtára.⁴ A feldolgozandó születési és halálozási nevek jelentős többsége település (pl. *Csíkménaság*, *London* [Finnország], *Pozsony*).

1 A földrajzi nevek és névterek irodalmát illetően lásd Ungváry [11], [12] és [13].

2 A jelenleg operatíván fejleszhető földrajzi névtér [8] a Relex [9] kezeli. Az eredményt majd egyeztetve véglegesíteni kell a Nemzeti Névtér béta változatának [7] végleges, „élesben” is használható állományával. Az utóbbinak ugyanis nincs még szerkesztősége, és maga a szerkesztőségi rendszer sincs még véglegesen befejezve <http://abcd.hu/2019/05/13/nemzeti-nevter-beta/>.

3 A szakirodalomból itt csak Fábrián Pál és Földi Ervin [3], továbbá Győry Erzsébet [4] munkájára utalok.

4 Ennek a térképészeti gyakorlathoz hasonlóan a legutolsó érvényes kiadását vettem figyelembe [5].



A másik része településrész (külterületi lakott hely, mint pl. *Ajtós*, *Csórompuszta*, *Kisújszállási csorba*). Ezek központi, állami, hivatalos névadása jelenleg sem teljes, ezeket a földrajzi neveket többnyire csak a helyi igazgatás hatáskörében tartják nyilván.⁵

A helynevek másik nagy csoportját a névváltozatok (szinonimák) alkotják: a ma is általánosan használt szinonimák (pl. *Peleske* (=Pölöske)), a helytélő⁶ és tájnevek, amelyeket csak a szóban forgó településen és környékén használnak (pl. *Hosszúhavas* (=Rakottyástelep [Csík vármegye]), *Mándibokor* (=Mandabokor [Nyíregyházi határ])), idegen nyelvű névváltozatok (pl. *Găiceana* (=Gajcsána [Moldva])), végül pedig a korábbi (pl. *Eszék* (=Osijek)) és a történelmi nevek (pl. *Boroszló* (=Wrocław), *Nyulak-szigete* (=Margit-sziget)).

Aminek nincs hivatalos neve, az a magyar térképészeti gyakorlatnak megfelelően szerepel a földrajzi névtérben. Számos történelmi név esetében a magyar történelemtudomány és régészet gyakorlatának megfelelő forma a mérvadó.

Nagyon sok külföldi földrajzi név is szerepel az állományban. Jelentős részük nem hivatalos, hanem elavult vagy hibás névalak, vagy csak a településen vagy annak környékén használt, ún. helytélő név. Akisnyomtatvány-forrásokban (pl. gyászjelentések), iskolai értesítőkből, katonai veszteség- és hadifogolyjegyzékekben rendkívül gyakoriak a helytélő nevek és az akár az érthetlenségig eltorzult elírások. A keleti és a balkáni fronthoz kapcsolódó nevek esetében, hivatalos katonai topográfiai térképek hiányában különösen nehezen rekonstruálható, hogy a megadott helynév milyen mai, hivatalosnak számító névnek felel meg.

Mindezek miatt nem lehet a forrásokban rendelkezésre álló névállományt egészében intellektuális előfeldolgozás nélkül, automatikusan összehozni sem a meglévő hivatalos földrajzinév-állományokkal (pl. a KSH vagy a FÖMI állományával), sem nemzetközi rendszerek (pl. a Geonames vagy a Wikipédia) neveivel, de magával a Relex [9] által kezelt állománnyal [8] sem.⁷ Lényegében minden egyes nevet ellenőrizni kell, amely nem felel meg maradéktalanul a Relex földrajzi névtérben szereplővel. De még azokat is ellenőrizni kell, amelyek első lépésre névalak dolgában teljesen azonosak a földrajzi névtér valamelyik névalakjával, mivel a homonim nevek következtében egy *Alexandriából*, *Alexandrovóból*, *Almásból* vagy egy *Annamajorból* akár több tucat is létezhet.

- 5 A normatív (hivatalosnak tekinthető) külterületi településrészeket a történelmi Magyarország vonatkozásában a teljesség igényétől meglehetősen távolálló, 1913-as helységnévtár [5] tartalmazza. A mai magyarországi külterületeket pedig a Földi Ervin által szerkesztett 1971-1982-es földrajzinév-tár [1].
- 6 A helytélő név a tájnévtől (*Rábatöttös* helyett *Rábatóttös*) abban különbözik, hogy nem nyelvjárási-tájnyelvi név (*Mezőkeresztes* helyett *Borsodkeresztes*). A két típus többnyire nem választható szét, a helytélő nevek többsége tájnévnek is tekinthető. A kifejezést a KSZ/5 vezette be [8].
- 7 A földrajzi névtér Relex által kezelt változata maradéktalanul tartalmazza a KSH és a FÖMI névállományát. Ezekben szinonim nevek nincsenek. A kitéüntetett névalak (deszkriptor) mindig az utóbbiakban nyilvántartott név. A FÖMI állománya, a „Magyarország Földrajzinév-tára” csak belső használatban tekinthető meg (<http://www.ftf.bfkh.gov.hu/portal/index.php/termekeink/magyarorszag-foeldrajzinev-tara>) [15]

A forrásokban rögzített születési és halálozási helyek feldolgozásának eredményei

A feldolgozandó nevek mennyiségéről az 1. táblázat ad felvilágosítást.

település és egyéb nem közterület	26728 db
budapesti közterület	1620 db
összesen	28348 db

1. táblázat. A feldolgozandó nevek

A feldolgozáshoz betöltöttük a PIM teljes névállományát és behasonlítottuk (összevetettük) a Relex által kezelt földrajzi névtér neveivel. Voltak nevek, amelyek megfeleltek a névtérben szereplő neveknek, és voltak, amelyek nem. Az utóbbiakhoz vagy meg lehetett találni a megfelelő helyes névtéri nevet, vagy rá lehetett jönni a névtéren kívüli, hosszas, intellektuális kereséssel a megfelelő névre. Ennek során a földrajzi névtér 130 000 névből (lexikai egységből) álló állománya kb. 1%-kal bővült.

Mindez arra utal, hogy a magyar vonatkozású belföldi, történelmi magyarországi és ezen kívüli neveket tartalmazó jelenlegi földrajzi névtér magyar vonatkozásban már közel teljesnek tekinthető. A mostani feldolgozás során keletkezett településnevek kisvárosok (pl. Bellaire, Ohio állam egyik négyezer lakosú városkája), olykor városrészek (pl. Gersthof, Bécs egyik városrésze), vagy rendkívül kis falvak (pl. Boldirjevka a Don közelében) voltak.

Az állomány 33%-ának feldolgozása után voltak

- a meglévő névalakok megegyezése következtében könnyen azonosítható nevek, a feldolgozott állomány 55%-a;
- behasonlítandó (nem azonosított) nevek, amelyeket szellemi munkával azonosítani kellett, a feldolgozott nevek 45%-a;
- ezen belül a sikeresen behasonlított (azonosított) nevek aránya (szellemi ráfordítással, egyeztetéssel és behasonlítással, más forrásokban való kutatással): 81% (teljes feldolgozott állomány 36%-a)
- a sikeresen azonosított nevek közül a szinonimák, a helytélők, az elavultak és a hibás nevek aránya, amelyekhez meg lehetett találni a megfelelő hivatalos, ill. a térképészeti gyakorlatban használt vagy a névtérben meglévő, vagy pedig a névtérben addig nem szereplő nevet: 57% (a teljes behasonlítandó állomány 46%-a)
- a sikeresen azonosított nevek közül azok, amelyeket közvetlenül meg lehetett feleltetni a névtérben szereplő vagy addig még ott nem szereplő, megfelelő, hivatalos, ill. a térképészeti gyakorlatban használt nevekkel: 43% (a teljes behasonlítandó állomány 35%-a)
- ezen belül a végképp azonosítatlannak bizonyult nevek aránya: 19% (a teljes feldolgozott állomány 9%-a)

Figyelemre méltó tehát, hogy a forrásokban szereplő nevek közel fele nem olyan volt, hogy a névalakjukból egyértelműen kiderült volna, hogy pontosan hol található a születési vagy halálozási hely. Ezek egy részénél ránézésre csak sejthető volt vagy az ország vagy a nyelvterület, ahol fekszenek.



A feldolgozási jellemzőket a 2. alábbi táblázat foglalja össze:

	teljes állományon belül	behasonlítható nevekben belül	sikeresen azonosítottakon belül
megegyező nevek	55%		
behasonlítható nevek	45%		
	100%		
végképp azonosítatlan név	9%	19%	
sikeresen azonosított (azonosított) név	36%	81%	
	(45%)	100%	
sikeresen azonosított név azonos a hivatalos, ill. térképészeti gyakorlat nevével (deszkriptor, kitüntetett név)	15%	35%	43%
sikeresen azonosított név azonos valamelyik névváltozattal (nem deszkriptor)	20%	46%	57%
	(35%)		100%

2. táblázat. Feldolgozási jellemzők

A hibák egy része egy vagy több karakter elírásából adódott. Ezek egy részénél csak asszociációk, hasonlóságok révén lehetett rátalálni a valószínűsíthetően helyes névalakra. Például:

Alabor völgy

Talabor-völgy

Más esetekben a vonatkozó személlyel kapcsolatos lehetséges adatbázisokban kellett keresni. Például

Alsómartic

Alsónastic (1899 után Alsóneszte).

Palló János (1882-) tábornok hibás születési helyét a Hadtörténeti Intézet és Múzeum egyik eldugott adattárában lehetett azonosítani a születési hely helyes nevével, amelyet az életrajzában adtak meg.

Apolda

Apold

Mivel Krausz András Kis- és Nagy-Küküllő vármegyékben volt lelkész, átnézve a két vármegye összes települését, a megtalált Apold lehetett a legvalószínűbb helyes név.

Balazkoje

Hrístoforivka

Szúnyoghy Szúnyog Péter itt eshetett el 1941 augusztusában. A Hadtörténelmi Levéltárban talált adatok szerint a magyar csapatok akkor Dél-Ukrajnában állomásoztak.

Egyedül az „Ukrainiendeutsche” honlapjain böngészve volt felfedezhető, hogy ez egy 1842-ben alapított német falu (Neu-Danzig) Dél-Ukrajnában, amelynek egyik neve a 19. század végétől Hrístoforivka, és egy helyt élő neve Balackoje, amelyet a korabeli magyar hadvezetés használt hallomásból és pontatlanul.

Braile Mocrin mocsarak

Mačini mocsarak

Az előtagról fel kellett tételezni, hogy az Braila kikötőváros a Duna-delta közelében, mert ezen a területen vannak a mocsarak. A „Braila” és a „Bräila” és a „mocsár” keresőszavakkal előkerült egy forrás (*Magyar Botanikai Lapok*) és benne a szövegrész: „Macintól Isaccea-ig a Duna dobrogeai oldala elmocsarasodik”. A Braila közelében fekvő Macsin szócikkben szerepel a „Brațul Măcin” („Macsniholtág”). Feltehetően ennek mocsaras környékéről van szó.

A nehézségek más részét az eldugott, félreeső helyeken fekvő, vagy nagyon távoli települések nevei okozták. Például:

Agoncy

Ambohitalanana

Benyovszky Móric halálának helye, Agoncy tengerparti település Madagaszkár északi részén, Sava tartomány Antalaha kerületében; Benyovszky Móric angol nyelvű Wikipédia-szócikkében szerepelt a mai, földrajzilag azonosítható név.

Akhalciay

Tiroyan P. Jákob (1863–) apostoli vagy keleti örmény katolikus pap születési helye, feltehetően Örményország vagy Törökország vagy Bulgária egyik aprófalva, amely egyetlen weben elérhető térképen sem szerepel ezen a néven, és az is lehet, hogy névelírás. Hosszas kutatás után is kinyomozhatatlan maradt.

A hibás adatok száma viszonylag magas, és viszonylag nagy a megállapíthatatlan nevek aránya is. Algoritmikus módszerek csak kis határfokkal alkalmazhatók, mivel csak ritka esetekben van csupán egy-két karakterben eltérő névalak, de még a valamennyire hasonló név sem gyakori a hibás nevek esetében. Ebből következően a földrajzi személyi hungarika feltérképezése meglehetősen hosszú időt és nagy behasonlítási-keresési ráfordítást igényel.



Az azonosításhoz azonban *nem elég, hogy azonosítani kellett a pontos, hivatalos vagy a magyar térképészeti gyakorlatban elfogadott helynevet*, mert a puszta név alapján nem lehet egyértelműen megállapítani, hogy hol fekszik. Legfeljebb úgy, hogy a kézikönyvekben vagy az interneten utána nézünk, de ez többnyire nem vezet könnyű eredményre, mert a nevek jelentős része homonim. Azaz ugyanaz a névalak sok különböző települést jelöl.

Például a *London* helynevet szerte a világon 17 település viseli, csak az Egyesült Államokban 10 település. *Antalmajor*ból Magyarországon több mint egy tucat található, és ezek bármelyike lehet születési vagy halálzási hely a forrásokban.

A nevekhez meg kellett állapítani azt az igazgatási egységet is, ahová tartozik, és ha ez nem szerepelt még a névtérben, akkor azt is újként föl kellett venni. Ha pedig nem szerepelt még a névtérben, hogy az adott igazgatási egység milyen átfogóbb igazgatási egységhez tartozik, akkor az egész láncolatot föl kellett venni újként.

Egy 19. századi németsséggel írt Lemberg környéki település nevének azonosítása a mai, hivatalos nevével a következő, egymás után feltárt láncolaton keresztül vezetett eredményre: *Blatocerkiew*, onnan *Belaja Zerkow*, onnan *Bjelaja Cerkovi*, végül *Bila Cerkva*. A település azóta többször is államot, nyelvet és írásrendszert cserélt, egyes névváltozatok el is tűntek, ki is veszték a köztudatból. Az ilyen jellegű munka időigényes kereséssel jár.

A földrajzi személyi hungarikum területi szórása

A személyek gyakran nem a mindenkori Magyarországon, nem is magyar nyelvterületen születtek vagy haltak meg, és sok esetben soha nem is éltek a mindenkori Magyarországon. Ez arról tanúskodik, hogy milyen a magyar kultúra területi integrációs képessége. Hogy honnan, mely országokból, régiókból történt meg a tudáshordozók beáramlása és megsemmisülése. Más szóval mennyire ágyazódott be Magyarország személyileg a világ – és legfőképpen Európa és vele együtt az euroatlanti kultúra – vérkeringésébe?

Hipotézisként megkockáztatom, hogy más nemzeti kultúrákhoz képest a magyar kultúra esetében valószínűleg nagyobb a külföldi releváns születési helyek aránya. Ezt más államok adatbázisaival való összehasonlító vizsgálatokkal lehetne tisztázni.

A halálzási helyek külföldi szórása a beáramlással és befogadással ellentétesen, áttételesen a magyarországi magyar kultúra veszteségeiről ad területi képet. Az utóbbiak száma ugyancsak rendkívül magas. Feltűnően sok a nyugat-európai, az amerikai egyesült államokbeli és a dél-amerikai. Mindez szorosan összefügg a különböző időkben (1849, 1919, 1938–39, 1945 és 1956 utáni) lezajlott kivándorlási, menekülési hullámokkal. Ugyanakkor szinte nincs olyan kontinense, országa/gyarmata a világnak, amelyhez ne kapcsolódna magyar vagy magyar vonatkozású szerző halálzási helye. A másik átlagon felüli sűrűsödési hely a keleti hadszíntér, Oroszország és a Szovjetunió, beleértve a közép-ázsiai és távoli-keleti területeket is. Ez utóbbi dolgában bizonyára a német nyelvterületen is hasonló a helyzet. Nemzetközi összehasonlításhoz pontos adatok egyelőre nem állnak rendelkezésre.

Hasonló munkák

Király Péter és Kiséry András jelenleg Demeter Tibor részben kézíratos írói bibliográfiája [2] alapján dolgozza fel azoknak a nyomdáknak a földrajzi helyeit, amelyekben magyar könyveket nyomtattak a világon. A munkáról a Facebook közösségi hálózaton számolt be Király Péter [6].

Ebben a munkában kb. 400 települést jelölő, 2500 körüli névalakot (névváltozatot) emeltek ki, ezek mindegyike nyomdahely, következésképp város, amelyekről sok dokumentáció és interneten letölthető forrás áll rendelkezésre. Ebből adódóan azonosításuk egyszerűbb, mint egy XIX. századi németességgel írt pusztanév Lemberg környékén. A nehézség itt elsősorban a kisebb-nagyobb elírásokból adódott, és a hivatalos névalakok megkeresése igényelt több munkát. Sok mindent meg lehetett oldani számítástechnikai eszközökkel (pl. a Geonames nevű földrajzinév-adatbázis kereső alkalmazásprogramozási felületével), többek között a fontossági előfordulási statisztikák, hasonlósági keresések lefuttatása, közeli írásképek esetében. Legfeljebb az egykori szovjet városok cirill és transliterált neveivel akadtak gondok.

Munkájuk konklúziójaként – a kiadási-nyomtatási évek alapján – kronológiai rétegződést is megvalósítottak, azaz időbeli bontásban, kattintások sorozatával interaktív térképek jeleníthetők meg.



1. ábra. Részletek a magyar könyveket nyomó nyomdák interaktív térképéből



Nagyjából ezzel egy időben Szakadát István a Nemzeti Névtér béta változatának személynévtérében szereplő emberek születési helyeinek térképét mutatta be a közösségi hálózaton Horvátország nélkül a történelmi Magyarország területén [10]

<https://www.facebook.com/photo?fbid=10158153057952981&set=a.10150592363867981> .



2. ábra. A Nemzeti Névtérben szereplő, és a történelmi Magyarországon született személyek születési helyei.

A pontok nagysága jelzi, hogy az adott településen hány névtérben szereplő ember született. Ezt még kontrasztosabban jeleníti meg a következő térkép:

<https://www.facebook.com/photo.php?fbid=10158153125377981&set=p.10158153125377981&type=3> [14]

Az ország kelet-alföldi, például nyírségi része feltűnően szegényes nevekben. Az is felismerhető, milyen mértékben növekszik a születési helyek sűrűsége nyugat és a Felvidék felé.

Összegzés, kitekintés

Mind a PIM-ben zajló, itt ismertetett munka, mind Király és Kiséry, mind pedig Szakadát kutatásai lehetőséget adnak arra, hogy az általuk alkalmazott informatikai eszközök segítségével számszerű eredmények szülessenek a kulturális folyamatok történelmi háttéréről, többek között az integrálóképesség és a veszteségek területi eloszlásáról. Az interaktív megjelenítés egyben a közérdeklődést is jobban kielégíti.

Hasonlóképpen feldolgozhatók és elemezhetők a szomszéd államok, de nem utolsó sorban Németország, Lengyelország és Oroszország nemzeti könyvtári adatbázisaiban az egységesített besorolási személynevek születési és halálozási helyei és évei.

Bibliográfia

- [1] *Magyarország földrajzinév-tára I.* Fontosabb domborzati, táj- és víznevek. szerk.: Földi Ervin. Budapest, Mezőgazdasági és Élelmezésügyi Minisztérium Országos Földügyi és Térképészeti Hivatal, 1971.
Magyarország földrajzinév-tára I. Fontosabb domborzati, táj- és víznevek. szerk.: Földi Ervin. 2. kiadás. Budapest, Kartográfiai Vállalat, 1982.
Magyarország földrajzinév-tára II. [Baranya megye; Győr-Sopron megye; Somogy megye; Vas megye; Veszprém megye; Zala megye; Csongrád megye; Fejér megye; Komárom megye; Tolna megye; Bács-Kiskun megye; Békés megye; Borsod-Abaúj-Zemplén megye; Heves megye; Pest megye; Budapest; Nógrád megye; Szolnok megye; Hajdú-Bihar megye; Szabolcs-Szatmár megye] szerk.: Földi Ervin. Budapest, Kartográfiai Vállalat, 1978–1981. [19 kötetben közreadva]
- [2] *Bibliographia Hungarica. Magyar szépirodalom idegen nyelven.* 1–15. kötet. Gyűjt. és leírta Demeter Tibor. Budapest, M. Külügymin., 1957–1958.
- [3] Fábrián Pál, Földi Ervin, Hőnyi Ede: *A földrajzi nevek helyesírása.* Budapest, Akadémiai Kiadó, 1998.
- [4] Gyórfy Erzsébet: A hivatalos név terminus. In: *Magyar Nyelvjárások.* 50. (2012.) p. 27–35. <http://mnytud.arts.klte.hu/mnyj/50/03gyorffy.pdf>
- [5] *A magyar szent korona országainak helységnevtára : Orts-Lexicon der Länder der Ungarischen Krone.* szerk. és kiad. a Magyar Királyi Központi Statisztikai Hivatal. Budapest, Magyar Királyi Központi Statisztikai Hivatal, 1913.
- [6] [Király Péter: Európai magyar nyomdák]. Facebook. 2021.02.18. <https://www.facebook.com/pkiraly/posts/10160071953561808>
- [7] Nemzeti Névtér béta változata. <http://abcd.hu>
- [8] Földrajzi névtér. A Nemzeti Névtér béta változatának Relexszel kezelt operatív formája. Szerk. Ungváry Rudolf. [Budapest, Országos Széchényi Könyvtár]. 2018–
<http://abcd.hu/foldrajzi-nevter/>
- [9] Relex (RELációk és LEXikai egységek) több felhasználós kliens-szerver alkalmazás névterek... osztott felhasználására egy és többnyelvű környezetben az interneten keresztül. Budapest: Országos Széchényi Könyvtár; MOKKA, 2012–.
<http://mokka.hu/relex/index.html>
- [10] [Szakadát István: a Nemzeti Névtér béta változatának személynévtére. Születési helyek]. Facebook. 2021.01.06.
<https://www.facebook.com/photo?fbid=10158153057952981&set=a.10150592363867981>
- [11] Ungváry Rudolf: A földrajzi nevek és szerepük az információkeresésben. Nemzetközi fejlődés, problémák és tapasztalatok a Köztauruszban és Geotauruszban. In: *Könyvtári figyelő.* 54. (2008.) 3. sz. <http://ki.oszk.hu/kf/2010/10/a-foldrajzi-nevek-es-sz-erepuk-az-informaciokeresesben/>
https://epa.oszk.hu/00100/00143/00068/pdf/EPA00143_konyvtari_figyelo_2008_3_395-428.pdf



- [12] Ungváry Rudolf, Pászti László: Földrajz nevek és szerepük az információkeresésben. Nemzetközi fejlődés, problémák és tapasztalatok a Köztauruszban és Geotauruszban. In: *Könyvtári figyelő*. 54. (2008.) 3. sz., p. 395–428.
<http://ki.oszk.hu/kf/2010/10/a-foldrajzi-nevek-es-szerepuk-az-informaciokeresesben/>
https://epa.oszk.hu/00100/00143/00068/pdf/EPA00143_konyvtari_figyelo_2008_3_395-428.pdf
- [13] Ungváry Rudolf: Névterek és földrajzinév-tárak. Észrevételek a Magyarország Földrajzinév-tára jövőbeni továbbfejlesztéséről. In: *Tudományos és Műszaki Tájékoztató*. 63. (2016.) 4. sz., p. 135–157.
https://epa.oszk.hu/03000/03071/00095/pdf/EPA03071_tmt_2016_04_135-157.pdf
- [14] [Szakadát István: a Nemzeti Névtér béta változatának személynévtére. Születési helyek. Újabb térkép]. Facebook. 2021.01.06.
<https://www.facebook.com/photo.php?fbid=10158153125377981&set=p.10158153125377981&type=3>
- [15] Lechner Tudásközpont. Magyarország Földrajzinév-tára.
<http://www.ftf.bfkh.gov.hu/portal/index.php/termekeink/magyarorszag-foeldrajzinev-tara>

Speciális dokumentumok és tartalmak – speciális metaadatszerkezetek
Muzeális tartalmak és kéziratok repozitálása a Szegedi Tudományegyetem
Klebelsberg Kuno Könyvtárban

Farkas Katalin

Szegedi Tudományegyetem Klebelsberg Kuno Könyvtár (Szeged)

Katalin.Farkas@ek.szte.hu

Nagy Dóra

Szegedi Tudományegyetem Klebelsberg Kuno Könyvtár (Szeged)

dora.nagy@ek.szte.hu

At the beginning of the 2020s, the SZTE Klebelsberg Library will need to store the digitised museum documents and manuscript legacies that have been and are still being generated in a repository. The content was placed in the library's thematic EPrints software-based repository Miscellanea. Different types of documents required different solutions, and their metadata required the development of a different metadata form. A solution was needed to create a structured structure in the repository to house the digital versions of the paper documents in the collection and the content transferred for digital preservation. The task was to define the different levels of access, the ongoing cleaning of the data and the establishment of the catalogue-repository link. The report presents possible solutions to these problems.

Keywords: digitised museum documents, manuscript legacies, metadata, EPrints software-based repository

Bevezetés

A 2020-as évek elején az SZTE Klebelsberg Könyvtárban szükségszerűvé vált a korábban és azóta folyamatosan keletkező digitalizált muzeális dokumentumok és kéziratok repositóriumban történő elhelyezése. A tartalmak a könyvtár tematikus EPrints szoftver alapú repositóriumaiban közül a Miscellanea-ban kerültek elhelyezésre. Többféle dokumentumtípus különböző megoldási lehetőségeket igényelt, metaadatokra különböző metaadat-szerkezetű űrlap kialakítására volt szükség. Megoldásra várt a repositóriumban a strukturált szerkezet kialakítása, melyben a gyűjteményben lévő papíralapú dokumentumok digitális változatai és a digitális megőrzésre átadott tartalmak elhelyezésre kerültek. Feladat volt a különböző elérési szintek meghatározása, a folyamatos adattisztítás és a katalógus-repositórium kapcsolat kialakítása. A beszámoló ezeknek a problémáknak lehetséges megoldását mutatja be.



1. Örökség

Az SZTE Klebelsberg Könyvtár szervezeti felépítése következtében a muzeális dokumentumokat¹ és a kézirattári anyagokat² a Különgyűjteményi Osztályon belül a Régi Könyvek Tára és Kézirattár könyvtárosai kezelik. A könyvtár MARC 21 szabványon alapuló katalógusában közel 16 000 rekord tartalmazza az összes muzeális könyv leírását³. A kézirattári állomány feldolgozottsága nem teljes körű. A tudóshagyatékok formai és tartalmi feltárása⁴ folyamatos, az elkészült leírásokat szintén a könyvtár integrált katalógusa tartalmazza.

2. Kezdetek

Az SZTE Klebelsberg Könyvtárban digitalizált muzeális tartalmak 2006-tól projektek következtében keletkeztek, mint például SZTE Bőrclinika Poór Imre Könyvtár (2006-2011) 248 db tétel, Kassán kiadott könyvek (2011-2013) 285 db tétel, 1956-os kézirattári dokumentumok (2017) 197 db tétel. A projekteken kívül külön kérésre digitalizált, úgynevezett „szórvány” példányok is létrejöttek.

2021-ben kb. 500 darab muzeális és 2000 hagyatéki tétel vált elérhetővé, valamint 1000 rekord betöltésre vár. A meglévő tartalmak előzetes, autopszia alapján történő szűrésen estek át, mivel a korai digitalizálás folyamán létrejöttek rossz, gyenge minőségű fájlok is.

3. Digitalizálás

Jelenleg a digitalizálás egy Zeutschel OS 12002 könyvszkenneren történik, mely állományvédelmi szempontból alkalmas a kényes, sérülékeny dokumentumok digitalizálására.

A digitalizálás során JPG vagy TIF formátumú fájlokat készítünk. A szöveges dokumentumok esetében ezeket a képfájlokat PDF formátumba konvertáljuk. A gépirásos szövegeken Abby FineReader programmal karakterfelismertetést (OCR) végzünk, így kétrétegű PDF állomány keletkezik, mely alkalmas a teljes szövegű keresésre. A kéziratok esetében nem lehetséges az OCR, így ezeket sima kép PDF-ként mentjük el.

- 1 Muzeális dokumentumok: 22/2005. (VII. 18.) NKÖM rendelet a muzeális könyvtári dokumentumok kezelésével és nyilvántartásával kapcsolatos szabályokról, hozzáférés: 2021. 05. 17. <https://net.jogtar.hu/jogszabaly?docid=a0500022.nkm>
- 2 Kézirattári dokumentumok: 27/2015. (V.27.) EMMI rendelet a közlevéltárak és a nyilvános magánlevéltárak tevékenységével összefüggő szakmai követelményekről, hozzáférés: 2021. 05. 17. <https://net.jogtar.hu/jogszabaly?docid=A1500027.EMM>
- 3 Muzeális dokumentumok bibliográfiai leírása: Régi nyomtatványok Könyvtári és szakirodalmi tájékoztatói szabályzat 2009. KSZ/6. Tankönyv: Varga András: Régi könyves ABC. Budapest MTA KIK, 2018, hozzáférés: <http://misc.bibl.u-szeged.hu/45521/>
- 4 Kézirattári dokumentumok bibliográfiai leírása: Berlász Jenő: Kéziratok katalogizálása. Budapest, OSZK, 1958, hozzáférés: 2021. 05. 18. <https://mek.oszk.hu/03100/03127/03127.htm>

Nemcsak a saját állományunk kerül digitalizálásra, hanem előfordulnak kölcsön kapott dokumentumok is (pl. Kristó Gyula hagyaték). Ezeket csak a szkennelés idejére kapjuk meg, majd visszaadjuk az eredeti tulajdonosnak. Minden esetben kérünk engedélyt a későbbi szolgáltatásra. Érdekes kérdés lesz a jövőre nézve a már eleve digitálisan született dokumentumok archiválása és esetleges szolgáltatásának biztosítása.

4. Tartalmak és metaadatok elhelyezése repositóriumban

A korábban digitalizált és jelenleg keletkező tartalmaknak visszakereshetőség szempontjából is megfelelő metaadatszerkezeteket kellett létrehozni, melyet a könyvtár repositóriumainak egyikében kellett elhelyezni. A muzeális tartalmak és kéziratok Klebelsberg Könyvtár tizenegy, tartalmában tematikusan elkülönülő digitális gyűjteménye közül a Miscellanea repositórumba⁵ (SZTE Miscellanea Digitális mindenest gyűjtemény) kerültek.

5. Metadatstruktúra

Az EPrints egyik előnye a rugalmas metaadat- és űrlapkezelés. Szabadon létrehozhatók a számunkra szükséges mezők, melyeket a megfelelő űrlaphoz lehet társítani. A metaadatstruktúra kialakítása dokumentumtípusonként történt, ami kilenc űrlap létrehozását igényelte: könyv, könyv része, cikk-tanulmány-mű, folyóirat, hírlap, kézirat, kisnyomtatvány, különlenyomat, konferencia vagy workshop anyag. Négy csoportba soroltuk a metaadatokat: cím adatok; név adatok; megjelenési adatok és egyéb adatok. Az egyes űrlapok az alap metaadatokon kívül (pl. szerző, dátum) a megadott dokumentumtípushoz tartozó különböző mezőket tartalmazzák. Esetünkben ilyen volt például a kézirat típusnál releváns 'Kézirat jellege' mező, így oldható meg, hogy az entitások leírhatók legyenek és összhangban álljanak a többi, már feldolgozott adattal. A megfelelő űrlapba rögzítés a digitális gyűjtemény saját és közös keresőfelületének alapja⁶.

5 SZTE Miscellanea. Gyűjtőköre a legszélesebb a repositóriumaink közül: az SZTE Klebelsberg Könyvtár muzeális dokumentumai és tudóshagyatékai valamint a Szegedi Tudományegyetemhez szorosan nem köthető, vegyes témájú és műfajú periodikák, könyvek, könyvfejezetek, cikkek kincsestára, hozzáférés: 2021. 05. 18. <http://misc.bibl.u-szeged.hu/>

6 SZTE Klebelsberg Könyvtár Contenta Repositóriumok közös keresője, hozzáférés: 2021. 05. 18. <https://contentas.bibl.u-szeged.hu/> és Farkas Richárd, Sándor Ákos: Digitalizált tartalmaink közös keresője VuFind alapokon az SZTE Klebelsberg Könyvtárban in.: Országos Networkshop 2020 ONLINE, Budapest HUNGARNET, MTA KIK, 2020 DOI: [10.31915/NWS.2020.3](https://doi.org/10.31915/NWS.2020.3), hozzáférés: 2021. 05. 18. <http://real.mtak.hu/119185/1/farkas.pdf>



Tétel nézet: **Balogh Jenő levele gróf Tisza Istvánhoz. "Másolat (fogalmazvány külön)"**

Ez a tétel URL címmel van az Repozitóriumban:
<http://misc.bibl.u-szeged.hu/57344/>

Tétel visszavonása Továbbítás az adminisztrátornak Tétel megsemmisítése

Részletek Akciók Történet Kiadások

Típus Adatok szerkesztése

Dokumentum típusa: Kézirat

Feltöltés Adatok szerkesztése

Dokumentum: [Teljes mű](#) **Típus:** Teljes mű. **Nyelv:** Hungarian. **Hozzáférés szintje:** Nyilvános.
• [polner_009_002.pdf](#)

Metaadatok Adatok szerkesztése

Szerző: Szerzők
Balogh, Jenő

Cím: Balogh Jenő levele gróf Tisza Istvánhoz. "Másolat (fogalmazvány külön)"

Nyelv: magyar

Dátum: 1914. január 14.

Kiadás, keletkezés helye: Budapest

Oldalszám megjegyzés: 6 fol.

Kapcsolódó URL: <http://misc.bibl.u-szeged.hu/57343/>

Gyűjtemény: [Hagyatékok > Polner Ödön hagyaték > Levelezés](#)

Kulcsszavak: Levelezés - Balogh Jenő - Tisza István, Jogtudomány - 1914 - forrás

Megjegyzések: Leltári szám: MS 2127/IX/2. Címoldal és a levél aláírása Polner Ödön autográf

Kézirat jellege: gépirás

Méret: 25,5 x 19 cm

Meghatározatlan mezők: [Közreműködők](#), [Egyéb cím](#), [Testületi név](#), [Absztrakt \(kivonat\)](#), [Oldalszám](#), [Projekt neve](#), [Proveniencia](#), [Hivatkozás](#)

Szakterület Adatok szerkesztése

1. ábra. Kézirat űrlap metaadat-struktúrája

6. Strukturált gyűjteményfelépítés, kutatóbarát környezet

EPrints-ben létre tudunk hozni többszintű témarendszert, melyet sokrétűen felhasználhatunk. Beépíthető például egy több lépcsős szakterületi besorolás⁷, illetve használható a gyűjtemény kialakításakor. A struktúra létrehozására, módosítására és kiegészítésére az adminisztrátori felületen van lehetőség. A külső, nyilvános felületen csak azok a kategóriák láthatóak, melyekhez van már rekord rendelve.

⁷ Bőngészés szakterület szerint az SZTE Elektronikus Tananyag Archívum oldalán, hozzáférés: 2021. 05. 21. <http://eta.bibl.u-szeged.hu/view/subjects/>

Téma szerkesztése: Szent-Györgyi Albert hagyaték

- (Felső szint)
 - Gyűjtemény
 - Hagyatékok
 - Szent-Györgyi Albert hagyaték

Gyerekek	Eprints	
Gratuláló levelek	295	Nem linkelt
Újságkivágat-album	341	Nem linkelt
Téma azonosító:		Létrehoz
Link Existing Subject:	1_0_0_0_0	Link

col_hagyatek_szgya

Név	Sorrend	Nyelv
Szent-Györgyi Albert hagyaték		NEM RÉSZLETEZETT ▼
		NEM RÉSZLETEZETT ▼
		NEM RÉSZLETEZETT ▼

További sorok

Letét: Igen, a felhasználók hozzáadhatnak. Nem.

Változások mentése

2. ábra. Gyűjtemény felépítés az adminisztrátori felületen

A Miscellanea repozitóriumban jelenleg három fő gyűjtemény található: hagyatékok, muzeális dokumentumok és szegei helytörténeti dokumentumok. A strukturált felépítés következtében a hagyatékok „Böngészés, Gyűjtemény szerint”⁸ felhasználói nézetben elkülönülnek és jól áttekinthetővé válnak.

A dokumentumok hozzáférési szintje lehet teljesen nyilvános, egyetemi IP tartományon belüli és csak az archívum karbantartójának elérhető.

8 Böngészés gyűjtemény szerint az SZTE Miscellanea Repozitórium oldalán: <http://misc.bibl.u-szeged.hu/view/collection/> hozzáférés: 2021. 05. 21.



Böngészés, Gyűjtemény szerint

Kérjük, válasszon az alábbi listából!

- [Gyűjtemény](#) (4549)
 - [Hagyatékok](#) (2026)
 - [Bálint Gábor hagyaték](#) (16)
 - [Halasy-Nagy József hagyaték](#) (8)
 - [Levelezés](#) (8)
 - [Herrmann Antal hagyaték](#) (397)
 - [Kalmár László hagyaték](#) (127)
 - [Kína](#) (114)
 - [Levelezés](#) (79)
 - [Klebsberg Kunoné hagyaték](#) (56)
 - [Levelezés](#) (36)
 - [Kristó Gyula hagyaték](#) (313)
 - [Újságkivágatok](#) (313)
 - [Márki Sándor hagyaték](#) (4)
 - [Németh Gyula hagyaték](#) (176)
 - [Polner Ödön hagyaték](#) (43)
 - [Levelezés](#) (14)
 - [Poór Imre hagyaték](#) (246)
 - [Szent-Györgyi Albert hagyaték](#) (640)
 - [Gratuláló levelek](#) (295)
 - [Újságkivágat-album](#) (340)
 - [Muzeális dokumentumok](#) (322)
 - [Szegedi helytörténeti dokumentumok](#) (2293)
 - [1956-os plakátok és szórólapok](#) (197)

3. ábra. Gyűjtemény felépítés a nyilvános felületen

7. Katalógusok és adatbázisok közötti kapcsolat

A repozitórium adatszerkezetének kialakítása közben alapvető szempont volt, hogy olyan környezet jöjjön létre, amelyben a felhasználó sok adat között könnyen kereshet, valamint felhívjuk a figyelmét a könyvtár egyéb elektronikus gyűjteményeiben található kapcsolódó dokumentumokra. Ennek megoldása a rekordkapcsolatok létrehozása.

7.1. Katalógusok és adatbázisok közötti kapcsolat: katalógus-repozitórium

A MARC 21 szabvány 856-os mezője alkalmas arra, hogy elektronikus dokumentumot is kapcsoljuk a leíráshoz. A 856-os mező első indikátora „4”, mely a http protokollra utal a második helyen „1” áll, ami a kapcsolódó forrás változatát definiálja, ebben az esetben a leírt dokumentum elektronikus változatát⁹. A rekordkapcsolatot 856 41 \$3Full text leltári szám\$hURL/URI\$z szabadon hozzáférhető szerkezetben kell megadni. A \$h almezőben a leírásokban a Miscellanea repozitórium URI elektronikus hozzáférése, a \$z almezőben pedig a hozzáférés szintje határozandó meg. A kapcsolatra kiváló példa Klebsberg Kunoné Botka Sarolta hagyatéka. A hagyatékról összefoglaló leírás készült a könyvtár integrált katalógusában, digitalizált kéziratári része a Miscellanea repozitóriumában, a

9 MARC 21 856 Electronic Location and Address, hozzáférés: 2021. 05. 19. <https://www.loc.gov/marc/bibliographic/bd856.html> és Gyuricza Andrea, Kasza Zsófia: MARC21 magyar fordítás, Budapest MTA KIK, 2017 DOI: [10.14755/MTAKIK.MARC21.2017](https://doi.org/10.14755/MTAKIK.MARC21.2017), hozzáférés: 2021. 05. 19.

szintén digitalizált fényképgyűjtemény pedig a Képtár és Médiatéka gyűjteményben található.¹⁰

7.2. Katalógusok és adatbázisok közötti kapcsolat: repositórium-repositórium

Egy repositóriumon belül és a könyvtár egyéb repositóriumainak összekapcsolására az EPrints szoftver is megoldást kínál. Két dokumentum között mellérendelő viszony hozható létre a „Megjelenési adatok” adatsorozatban található „Kapcsolódó URL” mező kitöltésével. Például Polner Ödön jogászprofesszor hagyatékában az összekötés egyes fogalmazványokon olvasható utalások kivetítésére szolgál. Ugyanez a megoldás alkalmazható két különböző repositórium vonatkozásában. A korábban példaként említett Klebelsberg Kunoné Botka Sarolta hagyaték kézirat és fényképgyűjteményt tartalmazó hagyatékreszeinek összekötése a Miscellanea repositóriumban szintén a „Kapcsolódó URL” alkalmazásával történt.

7.3. Katalógusok és adatbázisok közötti kapcsolat: repositórium-külső adatbázis

Muzeális és kéziratári hagyatékok estén is fontos szempont a minél szélesebb körű nyilvánosság elérése. Vannak nyelvi- illetve tartami szempontból különleges dokumentumok, melyek katalogizálásakor a tudományterület kutatójának segítségére van szükség. Az értő dokumentumhasználat a speciális gyűjtemények könyvtári feldolgozásában nagy hozzáadott értéket képvisel, mely a repositórium metaadatainak minőségét emeli. Eredményes kutató-könyvtáros/kéziratáros együttműködés eredménye az SZTE Klebelsberg Könyvtár Miscellanea repositóriuma és a Neumann János Számítógép-tudományi Társaság adatbázisa közötti kapcsolat. Utóbbiban több, Kalmár László matematikusprofesszor hagyatékában meglévő kéziratról található leírás¹¹. A Számítógép-tudományi Társaság az egyetem a szerverét „külső, tároló szerverként” használja, a felhasználó linken keresztül jut el a digitalizált tartalomhoz, ugyanakkor az elektronikus dokumentumok adatolva megtalálhatók a Miscellanea repositóriumban is¹².

10 A fényképeket Klebelsberg Kunoné Botka Sarolta jegyzetei alapján dr. Gausz Ildikó metaadatolta, hozzáférés: 2021. 05. 20. https://mediateka.ek.szte.hu/find?advanced%5B0%5D%5Bjoiner%5D=and&advanced%5B0%5D%5Belement_id%5D=70&advanced%5B0%5D%5Btype%5D=contains&advanced%5B0%5D%5Bterms%5D=Klebelsberg+album&sort=43&order=a&layout=1

11 Ezúton is köszönetet mondunk dr. Szabó Máté PhD matematikusnak (University of Oxford, Mathematical Institute), aki felhívta a Neumann János Számítógép-tudományi Társaság figyelmét a Kalmár hagyaték dokumentumaira.

12 Fidrich Ilona: Beindítóprogramok az M-3 géphez. Kézirat. Szeged, 1959 hozzáférés: 2021. 05. 19. <https://itf.njszt.hu/objektum/beinditoprogramok-az-m-3-gephez> és Az M-3 elektronikus számológép programozása. Az MTA Kibernetikai Kutató Csoportja 1958-59 évben tartott előadássorozatának anyaga. Budapest, MTA KKC, 1959, hozzáférés: 2021. 05. 19. <https://itf.njszt.hu/objektum/az-m-3-elektronikus-szamologep-programozasa>



8. Visszacsatolás: statisztika

A könyvtári repozitórium-külső adatbázis vonatkozásában esett szó arról, hogy a leírások, az adatszolgáltatás felhasználó központú. A gyűjtemények menedzselése, reklámozása gyűjteménykezelés fontos része. Ennek a munkának visszacsatolása a statisztikai adatok lekeresése, melyre az EPrints szoftver megoldást kínál az IRStats2 bővítmény segítségével. A statisztika menüpont minden Contenta repozitórium nyitólapján megtalálható. A különféle lekérdezések segítségével speciális keresések indíthatók. A régi könyvek és a tudóshagyatékokra vonatkozó találatok a Gyűjtemény almenü alatt jelennek meg. Poór Imre bőrgyógyász professzor hagyatékában általános orvostudományi és speciális, bőrgyógyászati témájú muzeális könyvek találhatóak. A gyűjtemény főként latin és német nyelvű műveket tartalmaz. Lekérdezés alkalmával¹³ a Letöltések grafikus megjelenítési módjában figyelemmel követhető, hogy a könyvek havi letöltési statisztikája 2020-ban nem került nyolcvan alá, ami meglepő egy ilyen különleges kollektív esetén. A Tevékenységek áttekintése a Tételek (246), a Letöltések (2,954) a Teljes szöveg (100%) és a Nyílt hozzáférés (100%) segítségével lehetséges. Érdekes az is, hogy az első három legtöbbet letöltött mű német nyelvű.

9. Összegzés

Az EPrints szoftvert alapvetően nem speciális dokumentumok metaadatolásához fejlesztették ki. Ugyanakkor munkánk során sikerült olyan űrlapokat és azon belül adatelemeket létrehozni, melyek alkalmazásával az eddig felépített rendszeren belül kialakult egy olyan repozitóriumrész, melyben különleges gyűjteményeket is el tudunk helyezni, segítve ezzel az oktatók és kutatók munkáját.

13 SZTE Miscellanea repozitórium Poór Imre hagyaték – Statisztika, hozzáférés 2021. 05. 20.
http://misc.bibl.u-szeged.hu/cgi/stats/report/collection/col_hagyatek_poor/

Szöveghasonlósági vizsgálatok automatizálása

Kiss Margit
BTK Irodalomtudományi Intézet, XIX. századi osztály
kiss.margit@abtk.hu

Palkó Gábor
ELTE BTK TI Digitális Bölcsészeti Tanszék, Digitális Örökség Nemzeti Laboratórium,
BTK Irodalomtudományi Intézet
palko.gabor@abtk.hu

Szakács Béla Benedek
BME VIK, MA hallgató
skyraiderfighter@gmail.com

A stilometriai elemzések a digitális bölcsészeti kutatásokban központi szerepet töltenek be: a szövegalkotás egyéni jellemzőinek a feltárásában segítenek úgy, hogy a meghatározó jegyek kiszűrése révén tulajdonságok tízezrei alapján hasonlíthatók össze a kérdéses szövegek. E számítógéppel végzett összetett statisztikai elemzőeljárások elvégzése közel sem egyszerű művelet. Különböző elemzőszoftverek (Websty, JGAAP, Stylene stb.) érhetőek ma már el, azonban azt a feladatot, amely a megfelelő elemzési eredmények eléréséhez szükséges, a felhasználónak komoly nyelvészeti-statisztikai háttértudással és empirikus kísérletek révén kell elvégeznie.

Tanulmányunkban felvázoljuk a stilometriai elemzés különböző alkalmazási területeit és eddigi eredményeit. Mint hazai innováció, bemutatunk egy a BME Méréstechnika és Információs Rendszerek Tanszék fejlesztésében, az ELTE Digitális Bölcsészeti Tanszék, illetve a BTK Irodalomtudományi Intézet együttműködésében kialakított webszolgáltatást, a *Shtylo*-t,¹ amelyet a Digitális örökség nemzeti laboratórium üzemeltet. Az alkalmazás korszerű újításai révén több szempontból is megkönnyíti a stilometriai elemzést végzők munkáját. Újdonsága, hogy webes környezetbe ágyazva a felhasználó mentesül a szoftver telepítése és a számítási feladatok elvégzéséhez szükséges, igen jelentős gépi erőforrás biztosítása alól. Egyedülálló módon, a felhasználótól elvárt tudásigényes feladatot automatikus paraméterezéssel támogatja: a heurisztikus paraméterbeállítás elemzi a vizsgálandó korpuszokat, és javaslatot tesz az adekvát beállításokra; ezen túlmenően a már meglévő paraméterek több elemzés automatikus futtatásával még optimálisabbá tehetők a pontosabb eredmények elérése érdekében.

Ma a stilometria megnevezés a stílus statisztikai alapú vizsgálatát jelenti, amely különböző tudományterületeket integrál: helye van benne a nyelvészetnek, irodalomtudománynak, filológiának, stilisztikának, statisztikának és az informatikának. A szó és a diszciplína megalkotója Wincenty Lutosławsky, aki a módszert Platón dialógusainak a

1 <https://github.com/szakacs/shtylo>



kronologizálásához alkalmazta.² Feladata az adott szerzőre jellemző megkülönböztető jegyeknek, az ún. szerzői ujjlenyomatnak a feltárása. Tipikus alkalmazási területe a szerzői idiolektus vizsgálata, az anonim vagy vitatott szerzőségű szövegek azonosítása,³ az egyéni nyelvezet alakulása, formálódása, korszakolás,⁴ csoporthoz tartozás vizsgálata,⁵ műfaji jelleg, de akár a nyelvi szempontból megmutatkozó hatás elemzése is.⁶ Az irodalom- és nyelvtudományi kutatásokon kívül törvényszéki eljárások során alkalmazzák,⁷ az utóbbi időben azonban más tudományterületeken is hozott eredményeket.⁸

Már a 19. században is felfigyeltek a szerzői jelleg statisztikai alapú megközelítésére,⁹ azonban a jelentős áttörést mégis a 20. század hozta el az informatika térhódításával, amelynek révén lehetővé vált a szövegek stilisztikai jegyeinek a mérése, eredmények összevethetősége, értékelése.¹⁰

A mai modern stilometria esetében, a számítógép azokon a területeken nyújt segítséget, amelyeken több szempontú, átfogó összehasonlítások szükségeltetnek: a nyelvi modellek vizsgálatakor a szövegalkotás, kifejezésmód egyéni jellemzőinek a feltárásában úgy, hogy a szerzőt meghatározó jegyek kiszűrésére törekszik. Mivel a kézi összehasonlítás csak rövidebb terjedelmű szövegeken egy-egy jellemző esetében lehet sikeres, digitális bölcsészeti eszközökkel jellemzően nagy terjedelmű szövegkorpuszok vizsgálatára nyílik már lehetőség. Ehhez a gépi feldolgozásra alkalmas jegyek kigyűjtését és összehasonlítását kell elvégeznünk a számítógép segítségével. Mindaz, amit a számítógépes nyelvészet egy adott szövegről meg tud határozni, az egy-egy tulajdonság lehet az összehasonlítás során: például teljes szóstatisztikák, szavak gyakorisága, szófaja, karakter n-gram száma, szóhossz, különféle gyakorisági jellemzők a teljes szövegre stb. A korszerű vizsgálatokban a legszélesebb körben elterjedt a minimum 100 leggyakoribb szó elemzése, ezt követi a mondatosság, a szóhosszúság, a hangsúlyos és hangsúlytalan szótagok váltakozása, a szókészlet gazdagsága, a leggyakoribb funkciószavak, a

- 2 Lutosławski Wincenty, *The Origin and Growth of Plato's Logic: With an Account of Plato's Style and of the Chronology of his Writings* (London: Longmans, 1897), <https://archive.org/details/originandgrowth00lutogoog/page/n44>.
- 3 Patrick Juola, „The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions,” *Digital Scholarship in the Humanities*, 30, 1. sz, 30, (2015), 100–113.
- 4 Dirk Van Hulle and Mike Kestemont, „Periodizing Samuel Beckett's Works: A Stylochronometric Approach,” *Style* 6, 2. sz. (2016), 172–202.
- 5 Sean G. Weidman and James O'Sullivan, „The Limits of Distinctive Words: Re-evaluating Literature's Gender Marker Debate,” *Digital Scholarship in the Humanities* 33, 2. sz. (2018): 374–390.
- 6 Regula Hohl Trillini and Sixta Quassdorf, „A 'Key to all Quotations'? A Corpus-Based Parameter Model of Intertextuality,” *Literary and Linguistic Computing* 25, 3. sz. (2010), 269–286.
- 7 Pelin Canbay, Ebru Akcapinar Sezer and Hayri Sever, „Deep Combination of Stylometry Features in Forensic Authorship Analysis,” *International Journal of Information Security Science* 9, 3. sz. , (2020) 154–163.
- 8 Haixia Liu, Raymond H. Chan and Yuan Yao, „Geometric tight frame based stylometry for art authentication of van Gogh paintings,” *Applied and Computational Harmonic Analysis* 41, 2. sz., (2016), 590–602.
- 9 Thomas Corwin Mendenhall, „The Characteristic Curves of Composition,” *Science* 9, 214. sz. (1887): 237–249.
- 10 John Burrows, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*, Oxford: Clarendon Press, 1987.

központosítás, a kollokációk, bizonyos betűsorozatok gyakoriságának a vizsgálata.¹¹ Ennek megfelelően ezeket a gépi feldolgozásra alkalmas jegyeket kell kigyűjteni, és a különböző szövegeket ez alapján összehasonlítani. A számítógép tulajdonságok tízezrei alapján képes erre, s a nagy mennyiségű jellemzőknek a vizsgálatára statisztikai módszereket alkalmazunk. Ezek a statisztikai elemző módszerek rendkívül bonyolult és összetett eljárások, amelyek megfelelő kiválasztása függ attól, hogy konkrétan mire vagyunk kíváncsiak az adott szöveg esetében. Hiszen ami az egyik esetben sikeres attribúciós eljárás lehet, az nem feltétlen eredményes a másikban: meg kell határozni, hogy adott feladatra, adott szövegekre milyen tulajdonságokat kell vizsgálni, s ehhez mi az adekvát módszer. Az elemzéshez reprezentatív korpusz összeállítása rendkívül fontos feladat, hiszen tudnunk kell, hogy az elemzés során miből adódik a szövegek közti különbség. Az elemzendő szövegek mérete ugyancsak fontos paraméter: bizonyos távolságmértékek erre nagyon érzékenyek, és emiatt torzíthatnak. A sikeres stilometriai elemzés tehát közel sem triviális feladat, az elméleti háttér alapos ismerete mellett az empirikus módon szerzett tudás szintűgy alapvető fontosságú.

A hazai innovációként fejlesztett *Shtylo*¹² webalkalmazás abban nyújt segítséget, hogy azok számára is megkönnyítse a stilometriai elemzések elvégzését, akiknek komolyabb nehézséget okoz ennek az elemzőmódszernek az alapos megismerése és számítógépes futtatása.

A *Shtylo* webalkalmazás alapját a *stylo* jelenti, egy R nyelven írt könyvtár, amelyet Maciej Eder és társai készítettek és publikáltak 2016-ban¹³. A stilometriai elemzés során végzett műveletek mind az ő implementációjukra támaszkodnak. Ez egy gazdag könyvtár, sok funkcióval, de a használata az R nyelv ismeretét és az ilyen típusú programok futtatásához szükséges környezetet igényli. Dobi Jan Sándor erre a problémára adott választ, amikor létrehozta a *Shiny* elnevezésű R nyelvű framework használatával ennek a szoftvernek az első változatát, amely már webes környezetben adott hozzáférést a *stylo* funkcióihoz.¹⁴ Mi ezt az alkalmazást dolgoztuk át és bővítettük ki a *Wizard* és *Analyzer* funkciókkal.

A *Shtylo* elsődleges előnye, hogy bárhonnán elérhető egy böngészővel, és képes url-ekről korpuszokat betölteni, hogy aztán azokon különböző elemzéseket lehessen elvégezni. Az elemzésekhez használt paraméterezés egy egyszerű *JSON* formátumú szöveggént kimásolható és újra felhasználható más korpuszoknál. Ezen felül pedig hozzáférést biztosít a *stylo* fő funkciójához, egy részletesen paraméterezhető elemzéshez, annak eredményének elmentéséhez, valamint a *Wizard* és *Analyzer* felületekhez, amelyek között a paraméterezés szabadon mozgatható. Mivel az elemzések meglehetősen számításigényes műveletek, ezért ezek a háttérben futnak, hogy a felhasználó közben használni tudja a felületet egyéb célokra.

11 Maciej Eder, „Style-Markers in Authorship Attribution: A Cross-Language Study of the Authorial Fingerprint,” *Studies of Polish Linguistics* 6, 1.sz. (2011), 99–114.

12 <https://github.com/szakacs/shtylo>

13 Maciej Eder, Jan Rybicki, Mike Kestemont “Stylometry with R: a package for computational text analysis.” *R Journal*, 8(1) (2016), 107–21.

14 Dobi Jan Sándor, Mészáros Tamás és Kiss Margit „Shtylo : stilometriai elemzések webes támogatása”, in Magyar Számítógépes Nyelvészeti Konferencia, (14)., 2018., 423–436.



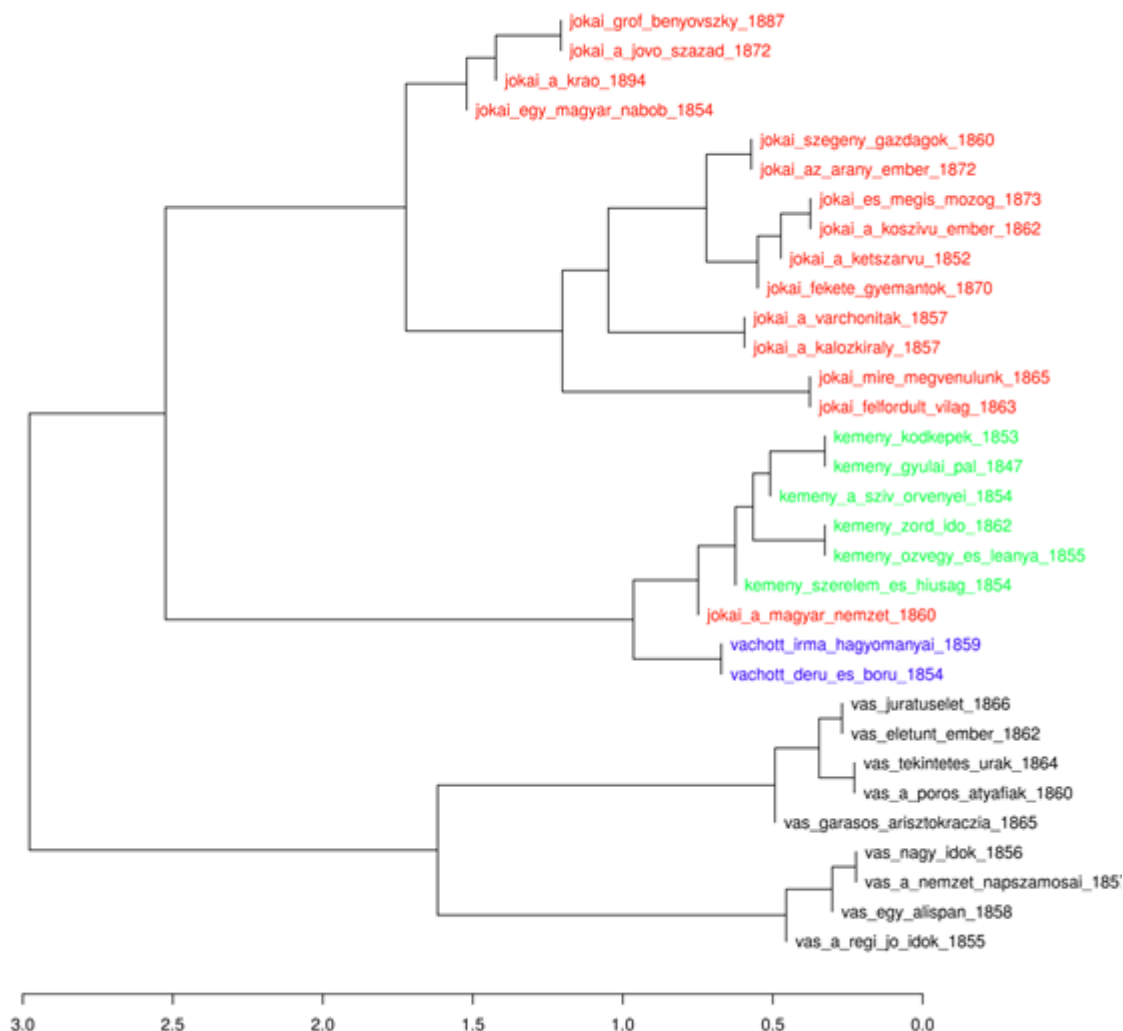
Maga a *Wizard*, hogy minél gyorsabb legyen, heurisztikus módon elemzi a szövegeket, szabályok segítségével, nem végez rajtuk tényleges stilometriai elemzést, csak alapvető statisztikai adatokat használ fel: a szó- vagy karakter-n-gramok eloszlását, a szövegek számát, a szövegek hosszát, és a szövegek szerzők közötti megoszlását. Pár fontosabb paramétert állít be vagy szűkít le: azt, hogy szó- vagy karakter-n-gram legyen használva, hogy mekkora legyen az n, a Most Frequent Words (a feature-ök hány százaléka legyen felhasználva) és a Culling (a szövegek hány százalékában kell jelen lennie a feature-nek, hogy használva legyen), és egyéb paraméterek. Ezek után a felhasználóra van bízva, hogy végighaladjon rajtuk, és módosítsa őket igényei szerint. A webes felületen láthatja a hozzájuk fűzött magyarázatokat is, amelyek segítenek a paraméterek elemzésében.

Az *Analyzer* esetében a cél az volt, hogy minél könnyebben csoportosíthatóvá tegye a megadott szövegeket. A működésének alapelve, hogy a szimulált lehűtés nevezetű lokális keresési algoritmus segítségével próbál egy olyan paraméter-kombinációt találni, amivel a megadott, különböző szerzőjű szövegek egymástól a lehető legtávolabb helyezkednek el. Ehhez minden lépésben megváltoztatja valamelyik paramétert egy egységgel (ez paramétertől függ, hogy mennyi). Mivel sok paraméter van, ezért csak a legfontosabbakat vizsgálja: ezek az elemzéshez használt n-gram típusa, hossza, a Most Frequent Words, a Culling és a távolságmérték típusa. A keresés az ezen paraméterek lehetséges értékeiből álló térben történik. Használhat előre megadott kezdőpontot ebben a térben, ami lehet akár a *Wizard* kimenete is, vagy van lehetőség arra, hogy több véletlenszerűen kiválasztott közül a legjobbat használja. Az algoritmus egy előre beállított lépésszám után áll meg. Jóval lassabb a *Wizard*-nál, mivel minden lépéskor lefut egy teljes stilometriai elemzés, viszont jelentősen pontosabb is. A használatával előállított paraméterezést aztán fel lehet használni olyan korpuszokon, ahol ismeretlen szerzőjű szövegek is vannak, és azok helyzetét lehet viszonyítani az ismert szerzőjű szövegekhez.

Az alábbi kísérlettel mutatjuk be az itt említett eljárások (*Wizard*, *Analyzer*) hatékonyságát. Az ELTE Digitális Bölcsészeti Tanszéken létrehozott regénykorpuszból¹⁵ összesen 32 regényt elemeztünk négy XIX. századi szerzőtől. Úgy válogattuk ki a műveket, hogy a keletkezési idejük közel essen egymáshoz – tehát a szerzőség és ne az adott időszak legyen a megkülönböztető elem. Az első elemzés a *Shtylo* alapbeállításával futott.¹⁶

¹⁵ <https://regenykorpusz.elte-dh.hu/>

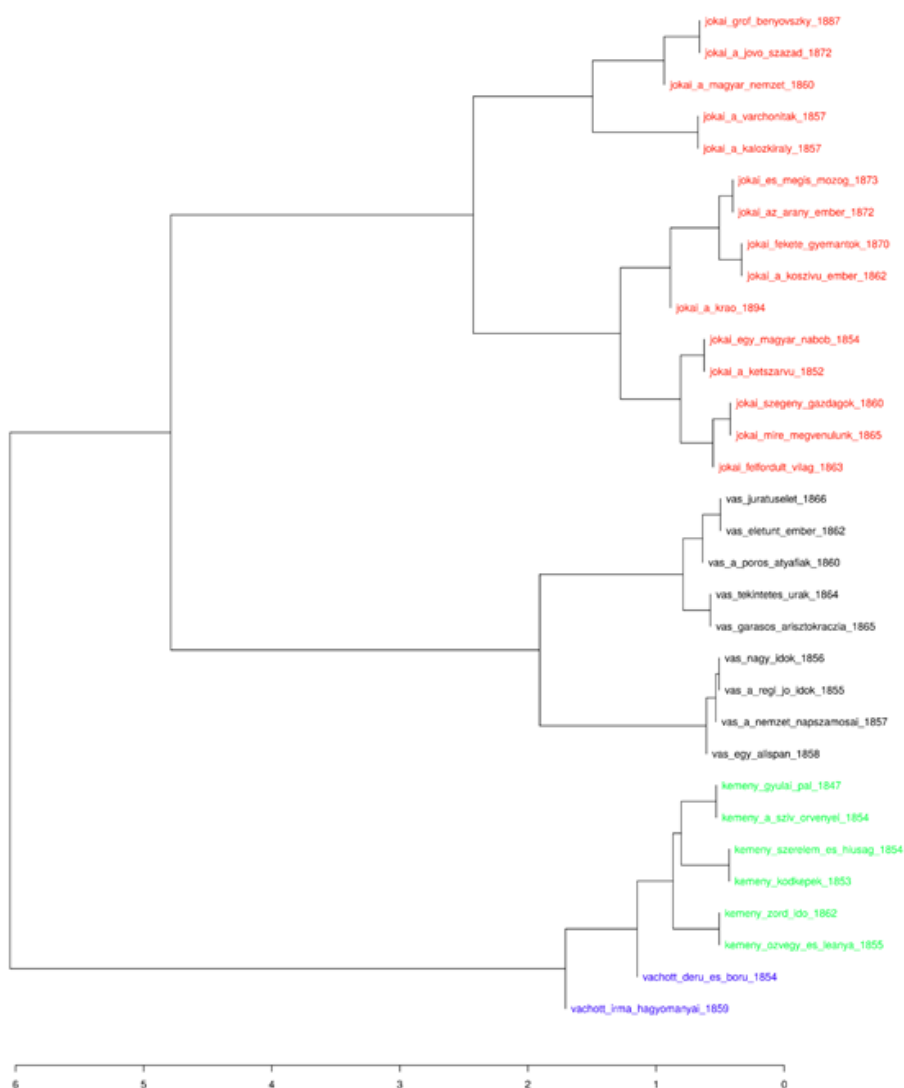
¹⁶ filetype: plain, language: English, encoding: yes, features: c, NgramInput: 1, CaseCheckbox: no, MfwMinimumInput: 100, MfwMaximumInput: 100, MfwIncrementInput: 100, MfwFreqRankInput: 1, CullingMinimumInput: 0, CullingMaximumInput: 0, CullingIncrementInput: 20, CullingListCutoffInput: 5000, CullingPronoun: no, Statistics: CA, StatisticsConsensus: 0.5, Scatterplot: labels, ScatterplotMargin: 2, ScatterplotOffset: 3, PcaFlavour: classic, ClusteringHorizontal: yes, Distances: dist.delta, SamplingMethod: no.sampling, SamplingNumber: 10000, OutputPlotHeight: 10, OutputPlotWidth: 10, OutputPlotFont: 10, OutputPlotLine: 1, OutputPlotColour: colors, OutputPlotDefault: no, OutputPlotTitles: no



Itt azt láthatjuk, hogy nagyrészt elkülönülnek az egyes szerzők művei, ám Jókai egyik regényét az elemző más klasztercsoporthoz sorolja, és az azonos szerzők művei is meglehetősen szétszóródnak egymástól.

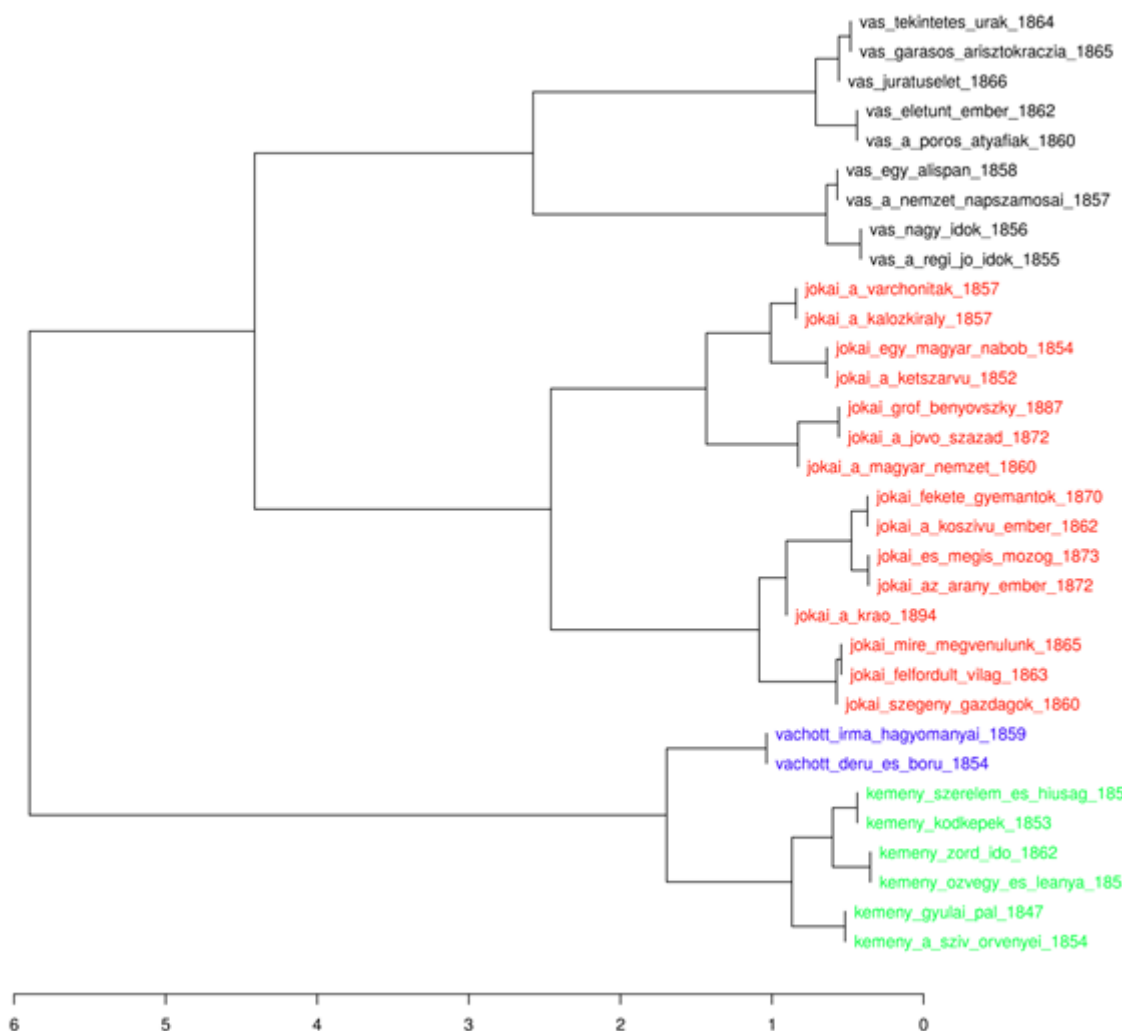
A következő kísérletben a *Wizard* futtatása után az alábbi eredményt kaptuk:¹⁷

17 filetype: plain, language: hungarian, encoding: yes, features: c, NgramInput: 3, CaseCheckbox: no, MfwMinimumInput: 67, MfwMaximumInput: 67, MfwIncrementInput: 100, MfwFreqRankInput: 1, CullingMinimumInput: 96, CullingMaximumInput: 96, CullingIncrementInput: 20, CullingListCutoffInput: 5000, CullingPronoun: yes, Statistics: CA, StatisticsConsensus: 0.5, Scatterplot: labels, ScatterplotMargin: 2, ScatterplotOffset: 3, PcaFlavour: classic, ClusteringHorizontal: yes, Distances: dist.delta, SamplingMethod: normal.sampling, SamplingNumber: .na, OutputPlotHeight: 16, OutputPlotWidth: 13, OutputPlotFont: 10, OutputPlotLine: 1, OutputPlotColour: colors, OutputPlotDefault: no, OutputPlotTitles: no



Ebben az esetben már egyértelműbben kirajzolódik a négy elkülönülő szerző csoportja, ugyanakkor az is látszik, hogy Vachott egyik regénye az elemzés alapján közelebbi rokonságban van Kemény Zsigmond szövegeivel, mint a másik általa írt művel. Ezt követően az *Analyzert* 10 iterációban futtattuk.¹⁸

18 filetype: plain, language: hungarian, encoding: yes, features: c, NgramInput: 3, CaseCheckbox: no, MfwMinimumInput: 67, MfwMaximumInput: 67, MfwIncrementInput: 100, MfwFreqRankInput: 1, CullingMinimumInput: 96, CullingMaximumInput: 96, CullingIncrementInput: 20, CullingListCutoffInput: 5000, CullingPronoun: yes, Statistics: CA, StatisticsConsensus: 0.5, Scatterplot: labels, ScatterplotMargin: 2, ScatterplotOffset: 3, PcaFlavour: classic, ClusteringHorizontal: yes, Distances: dist.delta, SamplingMethod: normal.sampling, SamplingNumber: .na, OutputPlotHeight: 16, OutputPlotWidth: 13, OutputPlotFont: 10, OutputPlotLine: 1, OutputPlotColour: colors, OutputPlotDefault: no, OutputPlotTitles: no



A négy szerző regényei ebben az elemzési eljárásban határozottan elkülönülnek egymástól, ugyanakkor az azonos szerzőhöz tartozó művek nem szóródnak külön ágakba, hanem sokkal szorosabb tömböket alkotnak, mint a korábbi kísérletekben. Az is tisztán látszik, hogy a négy szerző olyan különálló tömböket alkot, amelyek nem keverednek más szerzők klasztereivel.

A fenti kutatások és fejlesztések szakmai háttérét a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézet, valamint az ELTE BTK TI Digitális Bölcsészet Tanszék együttműködésével létrejött Stílometriai Kutatócsoport biztosítja, az infrastruktúrát pedig a Digitális örökség nemzeti laboratórium nyújtja. A *Shtylo* eszközre épülő szolgáltatás megvalósításához a technológiai know-how-t a BME VIK Méréstechnika és Információs Rendszerek Tanszék oktatója és hallgatói biztosították, nélkülük nem jöhetett volna létre a szolgáltatás.



A szolgáltatás nyitott minden érdeklődő számára, de célcsoportként a stilometriai kutatások iránt érdeklődő kutatókat és egyetemi hallgatókat határoztuk meg, azokat méghozzá, akik bár használni szeretnék a számítógépes stilsztika digitális eszközeit, de - híján a programozói kompetenciának - eddig ki voltak zárva ezen eszközök használati köréből. A limitált számítógépes infrastruktúrával, illetve tudással rendelkező kutatók számára különösen a fent bemutatott, a hiperparaméterek beállításokat automatizáló eszközök jelentenek nagy segítséget, ezek képezik a *Shtylo* újdonságát az eredeti R csomag, a *stylo*-hoz képest. Azonban még az automatizált paraméterezés sem teszi magától értetődővé, még kevésbé tudományos szempontból körültekintően argumentálttá a stilometriai kalkulációkat. Ehhez olyan képzésekre van szükség, ahol szerzőazonosításban vagy más stilometriai műveletekben technológiailag járatos, gyakorlott, de a szakirodalmi kontextust is ismerő kutatók bemutatják az eszközök helyes használatát, konkrét elemzéseken keresztül. Az elkövetkező években a Stilometriai Kutatócsoport ilyen képzések elindítását tervezi, személyes jelenlétre épülő kurzusok mellett eLearning anyagok formájában is.

A kvantitatív stilometriai kalkuláció nyelvfüggő, a magyar nyelv nyelvstatisztikai vizsgálata egészen más technológiákat követel meg, mint a stilometriai elemzések fősdórába tartozó világnyelvekre épülő elemzések. Ezért nagyon fontos, hogy a kalkulációk eredménye mögötti hiperparamétereket, és a paraméterezés számszerű eredményeit publikáljuk, hogy a jövőbeli, feltételezhetően egyre nagyobb számú egyedi kutatási eredmény, szerzőazonosító algoritmusra támaszkodó érvelés egy már létező és jól dokumentált kutatási térbe illeszkedjen, növelve az eredmények ellenőrizhetőségét. Eddig nagyon kevés ilyen eredményt publikáltak magyar nyelvű anyagok vizsgálatára építve. Ezért is fontos, hogy miközben mérésekkel alátámasztott érvelések segítségével erősödjön meg a magyar stilometria szakirodalma, folyamatosan a nálunk előbbre járó diskurzusok összefüggésében helyezzük el a kutatásainkat.

A Stilometriai Kutatócsoport ezen dolgozik, részt veszünk az International Journal of Digital Humanities szerzőazonosítással foglalkozó különszámának szerkesztésében, illetve a COST Action „Distant Reading for European Literary History” nemzetközi projektben, amelynek keretei között olyan korpuszokat hozunk létre és publikálunk, amelyek eleve a stilometriai vizsgálatokra vannak optimalizálva.

A szolgáltatás további fejlesztésének rövid távú céljai között szerepel a felhasználói fiókok kialakításának lehetősége, lehetőség szerint EduID alapú azonosítás használatával, valamint a korpuszok kezeléséhez felhő alapú tárhely szolgáltatás integrálása. Hosszabb távon a továbbra is az R programnyelv környezetében működő *stylo* funkcionalitását a digitális bölcsészetben előre törő, hovatovább egyeduralmukodóvá váló, így a kutatók szélesebb köre számára könnyebben elérhető *Python* nyelvi környezetbe fogjuk átültetni. Ezen munka első lépéseit a jelen kötetben megjelenő Jókai stilometriai elemzés írja le. Célunk továbbá, hogy a szerzőazonosítás módszertanát az irodalmi szövegeken túli szövegműfajokra is kiterjesszük, kézenfekvő az eszközök felhasználása webaratással létrejövő korpuszokra.

Távlati célunk olyan új, a *stylo* algoritmusainak frissebb technológiák integrálása a stilometriai kutatásba, mint a vektortér technológia, illetve a mélytanulás eszközeinek alkalmazása. Ezzel a munkával párhuzamosan haladunk a Maciej Eder vezette, és Európa egyik legfontosabb számítógépes stilisztikai műhelynek számító krakkói „Computational Stylistics Group” tevékenységével.



Az ELTE DH Regénykorpusz és lehetőségei

Bajzát Tímea Borbála

ELTE DH

timimimi@student.elte.hu

Szemes Botond Bálint

ELTE DH

boboelte@student.elte.hu

Szlávich Eszter

ELTE DH

szlavich.eszter@btk.elte.hu

The corpus of the Hungarian novel created at the ELTE Department of Digital Humanities offers new methodologies for the philological research and “distant reading” approaches by providing a digitized, annotated and searchable database of freely accessible novels from the Hungarian literary history. The database fits organically into the international collection of the ELTeC COST Action Project (European Literary Text Collection <https://www.distant-reading.net/eltec/>), since the first 100 novels of the database are part of the Hungarian sub-corpus of that collection. Beside the description of the corpus and the aspects of the selection, the paper also reports in detail on the possibilities of the quantitative analysis of the novels. In doing so, we want to present what kind of knowledge of the Hungarian literary history can be produced by applying statistical and linguistic approaches, and what role these methods can play in the process of the interpretation of the texts. Through the visualized tendencies, a new history of the style of the Hungarian prose can be outlined, while the peculiarities of some texts in relation to others can lead to the description of the poetics of the given authors and their novels.

Keywords: distant reading, corpus linguistics, literary corpus, Hungarian novels



1. Bevezetés

A tanulmány célkitűzése az ELTE BTK Digitális Bölcsészet Tanszék által létrehozott és fejlesztett Regénykorpusz¹ projektum² összegző bemutatása. Az ELTE BTK Digitális Bölcsészet Tanszék által szolgáltatott Regénykorpusz olyan folyamatosan bővülő, innovatív kezdeményezés,³ amely egyedülként biztosítja magyar nyelvű regények annotált, kereshető elérését, ezáltal alkalmas a nagy mennyiségű adatokon végzett korpuszalapú és korpuszvezérelt szövegvizsgálatokra. A Regénykorpusz aktuálisan 100 magyar nyelven írt regényt tartalmaz, összesen 81 szerzőtől, mérete megközelítőleg 7 000 000 token. A Regénykorpusz ezen gyűjteménye szervesen illeszkedik abba a nemzetközi COST Action kutatási együttműködésbe, amely a Distant Reading for European Literary History⁴ nevet viseli. A kutatási projektum célkitűzései közé tartozik egy kutatói hálózat megteremtése a projektben részt vevő kutatóintézetek között. Az együttműködés célja az, hogy olyan források és eszközök váljanak széles körben hozzáférhetővé, melyek elősegítik azt, hogy az európai irodalomtörténet-írás és irodalomelméleti paradigma, valamint az ezekhez fűződő fogalmi apparátus újraértelmezhetővé váljon. A projektum a kutatás módszertanában a távoli olvasás metódusát érvényesíti. Valamint célja az, hogy adatvezérelt, koherens, számítógépes eljárást alakítson ki és tegyen tesztelhetővé az irodalmi szövegek komparatív elemzésére.⁵ A projekt keretében jött létre, és dinamikusan fejlődik az European Literary Text Collection (ELTeC)⁶, amely a projektben részt vevő európai nyelvek regénykorpuszait tartalmazza,⁷ ennek része a jelen tanulmányban bemutatandó Regénykorpusz is.

2. A magyar regény korpusza

A Regénykorpusz jelenlegi státuszában az ELTeC, vagyis a European Literary Text Collection gyűjteményének részét képezi, ezáltal mind a gyűjteménybe kerülő szövegek kiválasztási módszere, mind pedig az anyagon alkalmazott kódolási séma ezen

1 <https://regenykorpusz.elte-dh.hu/?lang=hu-HU> (utolsó elérés: 2021.05.26.)

2 A projektum létrehozásán és fejlesztésén dolgozó kutatók és az ellátott feladatkörük a következők: Dr. Palkó Gábor (projektvezető és az alkalmazott TEI XML specifikáció készítője), Fellegi Zsófia (az alkalmazott TEI XML specifikáció készítője), Takács Emma (jelölőnyelvi kódolás), Véték Bence (jelölőnyelvi kódolás), Dr. Kundráth Péter (a Regénykorpusz lekérdező felületének létrehozása), Dr. Horváth Péter (a Regénykorpusz lekérdező felületének létrehozása), Szemes Botond (a lekérdező funkciók bővítése), Szlávich Eszter (lekérdező funkciók bővítése) és Bajzát Tímea Borbála (jelölőnyelvi kódolás, metaadatolás, lekérdező funkciók bővítése).

3 A projektet a Felsőoktatási Intézményi Kiválósági Program támogatta, jelenleg a Digitális Örökség Nemzeti Laboratórium keretei között végezzük a korpusz fejlesztését.

4 [COST Action CA16204](https://www.cost.eu/Action/CA16204) (utolsó elérés: 2021.05.26.)

5 Vö. Christof Schöch, Maciej Eder, Arias Rosario, François Pieter, Antonija Primorac, *Foundations of Distant Reading. Historical Roots, Conceptual Development and Theoretical Assumptions around Computational Approaches to Literary Texts*. 2020, hozzáférés: 2021.05.26. <https://dh2020.hcommons.org/>

6 Carolin Odebrecht, Lou Burnard, Christof Schöch, *COST Action Distant Reading for European Literary History (CA16204)*, 2021, hozzáférés: 2021.05.26. doi: <https://doi.org/10.5281/zenodo.4662444>

7 A projekthez lásd még: <https://www.distant-reading.net/eltec/> (2021.05.23), amelynek magyar nyelvű adatbázisa elérhető: <https://github.com/COST-ELTeC/ELTeC-hun> (2021.05.23).

projektum előírásaihoz illeszkedik.⁸ Ennek alapelveit követve törekedtünk a korpusz változatosságának maximalizálására, tehát a különféle terjedelmű és kanonizáltságú művek, valamint a különböző nemű szerzők arányos eloszlására.

A Regénykorpusz gyűjteményét olyan művek képezik, amelyeknek az első nyomtatott kiadásuk az 1840-től az 1920-ig tartó periódusra datálható az Országos Széchényi Könyvtár internetes katalógusa szerint.⁹ Ezen 80 évből álló időszak további négy alperiódusra oszlik fel (lásd 1. ábra), tehát az alkorpusz a 19. századi és 20. század eleji magyar regényirodalomból nyújt mintát a vizsgálatokhoz. Ugyan a szövegek kiválasztásánál ezen alperiódusokat vettük figyelembe, de a keresőfelületen az első kiadás évszáma alapján végezhetünk szűréseket. Az alperiódusok pontos mérete a következő: A T1 alkorpusz 22 regényt, a T2 21 regényt, a T3 27 regényt, a T4 pedig 30 regényt tartalmaz. A regények között nem szerepelnek fordítások, tehát mindegyike magyar nyelven íródott. A periódusoknak megfelelő alkorpuszokra azonban nem csupán mennyiségi megkötést alkalmaztunk, hanem minden algyűjteménynek minimum 10%-át kellett kitenniük a női szerzők által írt műveknek, ami így alkorpuszonként legalább három női szerző által írt szöveget eredményezett. A változatosság maximalizálása miatt pedig a teljes gyűjteményre érvényes volt az a szabály, hogy szerzőismétlődés legfeljebb tizenegyszer fordulhatott elő és ugyanattól a személytől legfeljebb három regény kerülhetett beválogatásra.

Az ELTeC által szabott kritériumoknak eleget téve csak olyan szöveget vettünk fel a korpuszba, amely legalább 10 000 szó terjedelemben íródott. A terjedelmi kategóriákat tekintve rövid prózának címkéztünk minden olyan művet, amelynek mérete 10 000 és 49 999 token közé esett, közepes méretűnek számítottak azon szövegek, amelyek 50 000 és 99 999 közötti szövegszót tartalmaztak és a hosszú regények kategóriába eső műveknek pedig a 100 000 szó felettiek számítottak. A teljes gyűjteményre vonatkozóan minimum 20, az előbbieket alapján hosszúnak számító regény került be a korpuszba (a Regénykorpuszban összesen 22 hosszú regény található).

A kanonikusságot tekintve a válogatási kritérium az volt, hogy a gyűjtemény minimum egyharmadát kell azon szövegeknek képezniük, amelyek magas kanonicitásúnak számítanak. Az ELTeC előírásai szerint azok a művek tartoznak ebbe a kategóriába, amelyek 1979 után minimum 2 új kiadással rendelkeznek, tehát a kanonikusság meghatározása ebben a kritériumrendszerben alapvetően a kiadástörténethez rendelődik.

8 A kiválasztási kritériumok forrásaként lásd: https://distantreading.github.io/sampling_proposal.html (hozzáférés: 2021.05.23.)

9 Lásd http://nektar1.oszk.hu/librivation_hun.html (hozzáférés: 2021.05.23.)



T1 (1840–1860)			T3 (1880–1900)		
22 db			27 db		
női szerző által írt	férfi szerző által írt		női szerző által írt	férfi szerző által írt	
3 db	19 db		4 db	23 db	
alacsony kanonicitás	magas kanonicitás		alacsony kanonicitás	magas kanonicitás	
14 db	8 db		21 db	6 db	
rövid terjedelmű	közepes terjedelmű	hosszú terjedelmű	rövid terjedelmű	közepes terjedelmű	hosszú terjedelmű
6 db	7 db	9 db	20 db	5 db	2 db
T2 (1860–1880)			T4 (1900–1920)		
21 db			30 db		
női szerző által írt	férfi szerző által írt		női szerző által írt	férfi szerző által írt	
6 db	15 db		7 db	23 db	
alacsony kanonicitás	magas kanonicitás		alacsony kanonicitás	magas kanonicitás	
17 db	4 db		18 db	12 db	
rövid terjedelmű	közepes terjedelmű	hosszú terjedelmű	rövid terjedelmű	közepes terjedelmű	hosszú terjedelmű
9 db	9 db	9 db	12 db	10 db	8 db

1. ábra. A Regénykorpusz algyűjteményei

A Regénykorpuszba csak szabadon elérhető szövegeket használtunk fel, a szövegek elsődleges forrása a Magyar Elektronikus Könyvtár,¹⁰ de ahhoz, hogy a gyűjtemény megfelelhessen az ELTeC által támasztott válogatási kritériumoknak, kettő regény a Google Books szabadon hozzáférhető adatbázisából származik.¹¹ A leválogatásnál elsősorban arra törekedtünk, hogy olyan szövegekkel dolgozzunk, amelyek RTF formátumban elérhetők a MEK felületén, mert ezek olyan jó minőségben tárolják a munkaanyagot a számítógépes feldolgozás számára, hogy további munkákat nem igényelnek a kódolás során, azonban filológiai szempontból további kérdések fogalmazhatók meg velük kapcsolatban. Azonban a MEK-ről vételezett RTF dokumentumok önmagukban nem bizonyultak elegendőnek a kritériumok teljesítéséhez, így a mintavételezést kibővítettük a MEK-es és a Google Books-os anyag kétrétegű PDF-ben tárolt dokumentumaira is, amelyeken újra OCR-t (optikai karakterfelismerést) végeztünk el az ABBYY FineReader 14 szoftver alkalmazásával, majd a tipikus OCR hibákat kézzel javítottuk. A regények metaadatait (az első kiadás éve, kiadások száma) az OSZK katalógusának internetes keresőjéből és a Magyar Országos Közös Katalógusból (MOKKA)¹² gyűjtöttük össze.

A korpuszba kerülő szövegek alapvető kódolási formátuma a TEI XML jelölőnyelv,¹³ amely mind az ember, mind pedig a gép számára olvasható metanyelv. Előnye, hogy eszköz- és rendszerfüggetlen, valamint a kódolt szövegtestek együttesen tárolhatók azok metaadataival. Ezen keresztül olyan irányelvek gyűjteménye, amely segítségével lehetővé válik a strukturált szöveg és információ megjelenítése a böngészőben, illetve más szövegformátummá konvertálható a feldolgozott anyag, ezen kívül a felhasználás célkitűzéseinek teljesítésére alkalmas annotációval láthatjuk el a szövegeket.¹⁴ Az ELTeC projektum specifikus standardizációt alkalmaz a regények kódolásához, amely

10 <https://mek.oszk.hu/> (hozzáférés: 2021.05.23.)

11 <https://books.google.hu/> (hozzáférés: 2021.05.23.)

12 <http://www.mokka.hu/> (hozzáférés: 2021.05.23.)

13 Lásd <https://tei-c.org/about/history/> (hozzáférés: 2021.05.23.)

14 Kalcsó Gyula, „A TEI-XML felhasználása magyar nyelvű korpuszok építésében”, in Boda István, Mónos Katalin szerk., *MANYE XX. Az alkalmazott nyelvészet ma: Innováció, technológia, tradíció*, (Debrecen: MANYE, Debreceni Egyetem), 67–68. 2011.

olyan specifikációja a TEI XML-nek,¹⁵ mely lehetővé teszi a fejlécben a projektum szempontjából releváns metaadatok jelölését (lásd 2. ábra) az XML fejlécben.

```
35 </respStmt>
36 <bibl type="printSource"><title>Egy régi udvarház utolsó gazdája</title><author>Gyulai Pál</author><publisher>Interpopulart</publisher><pubPlace>Budapest</pubP
37 <ref target="http://mek.oszk.hu/00600/00666/00666.pdf"/></bibl>
38 <idno>ISBN 963 613 127 9</idno>
39 </bibl>
40 <bibl type="firstEdition">
41 <title>Egy régi udvarház utolsó gazdája : Regény</title>
42 <author>Gyulai Pál</author>
43 <date>1857</date>
44 </bibl>
45 </sourceDesc>
46 </fileDesc>
47 <encodingDesc n="eltec-0">
48 <p/>
49 </encodingDesc>
50 <profileDesc>
51 <langUsage>
52 <language ident="hu">Hungarian</language>
53 </langUsage>
54 <textDesc>
55 <authorGender xmlns="http://distantreading.net/eltec/ns" key="M"/>
56 <size xmlns="http://distantreading.net/eltec/ns" key="short"/>
57 <canonicity xmlns="http://distantreading.net/eltec/ns" key="high"/>
58 <timeSlot xmlns="http://distantreading.net/eltec/ns" key="I1"/>
59 </textDesc>
60 </profileDesc>
```

2. ábra. A TEI XML specifikációja a <header>-ben, amely a metaadatok tárolására alkalmas

Ahhoz, hogy az ELTE Digitális Bölcsészeti Tanszék által összeállított Regénykorpusz kereshető legyen, akár például a morfológiai kódok alapján, tehát teljesítse azokat az elvárásainkat, amelyeket egy annotált korpusz felé támasztunk, szükség volt a szöveg elemeinek lemmatizációjára és morfológiai, valamint szófaji elemzésére. Ezek eléréséhez az MTA Nyelvtudományi Intézetben fejlesztett e-magyar automatikus elemzőlánc emtsv verzióját alkalmaztuk¹⁶, úgy, ahogy a szintén a Tanszéken fejlesztett Verskorpusz projektum esetében is.¹⁷ Az e-magyar segítségével így lehetővé vált a szövegek tokenizálása, lemmatizálása, morfológiai és szófaji elemzése is. A következő (3.) fejezetben útmutatót adunk a Regénykorpuszban¹⁸ való keresés lehetőségeinek használatához.

3. Keresés a korpuszban

A Regénykorpuszhoz elérhető egy nyilvános online lekérdező felület is, amely funkcióival és arculatával illeszkedik az ELTE Digitális Bölcsészeti Tanszék többi szolgáltatása közé (az ELTE DH elérhető, kereshető szövegkorpuszai: Verskorpusz és Cikk-kereső).

15 Vö. Lou Burnard, Christof Schöch, Carolin Odebrecht, In Search of Comity: TEI for Distant Reading, 2019, <https://doi.org/10.5281/zenodo.3552489> (utolsó elérés: 2021.05.23.) 65–72., <https://github.com/COST-ELTeC/ELTeC-hun> (utolsó elérés: 2021.06.23.)

16 Indig Balázs, Sass Bálint, Simon Eszter, Mittelholcz Iván, Kundráth Péter, Vadász Noémi, „emtsv – egy formátum mind felett”, in Berend Gábor, Gosztolya Gábor, Vincze Veronika szerk., XV. Magyar Számítógépes Nyelvészeti Konferencia. (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet), 235–247. 2019. Mittelholcz Iván, „emToken: Unicode-képes tokenizáló magyar nyelvre”, in Vincze Veronika szerk., XIII. Magyar Számítógépes Nyelvészeti Konferencia. (Szeged: Szegedi Tudományegyetem, Informatikai Intézet) 61–69. 2017.

17 Horváth Péter, „Az ELTE Verskorpusz automatikus annotációs eljárásai révén nyerhető kvantitatív adattípusok”, in Simon Gábor, Tolcsvai Nagy Gábor szerk., *Nyelvtan, diskurzus, megismerés*, (Budapest: Eötvös Kiadó), 313–331, 2020.

18 <https://regenykorpusz.elte-dh.hu/> (utolsó elérés: 2021.06.23.); <https://github.com/ELTE-DH/regenykorpusz> (utolsó elérés: 2021.06.23.)



A korpusz létrehozásáról és a kereső részletes leírásáról a Súgó menüpontban lehet tájékozódni. A keresőfelület elérhető angol nyelven is.

A regénykorpusz honlapján elérhető részletes keresőfelület funkcióit három fő kategóriába lehet sorolni: a felület egyrészt alkalmas alkorporuszok létrehozására, másrészt lehet rajta tokenekre és tokenkapcsolatokra keresni, és végül vannak funkciók az adatok feldolgozási módjaira is.

The screenshot displays a search interface for a corpus. It is organized into three main sections, each highlighted with a colored border:

- Metadata Filters (Yellow border):** Includes dropdown menus for 'Szerző:' (Author), 'Műcím:' (Title), 'Szerző neve:' (Author name), 'Terjedelem:' (Scope), and 'Kanonikusság:' (Canonicity). It also features date range inputs for 'Keletkezés ideje:' (Date of origin) and buttons for 'Szóalak' (Word form), 'Szóösszetétel' (Word structure), 'Szófaj' (Word class), and 'Morfológia' (Morphology).
- Token Search (Green border):** Includes a text input for 'Tartalom (tokenek):' (Content (tokens)), a dropdown for 'Tokenek kapcsolata:' (Token relationship) set to 'minden elem ugyanabban a regényben' (all elements in the same novel), and a numeric input for 'Tokenek max. távolsága:' (Token max. distance) set to 1.
- Context Search (Blue border):** Includes dropdowns for 'Kontextus típusa:' (Context type) set to 'szavak' (words), 'Kontextus mérete:' (Context size) set to 'bekezdés' (paragraph), and 'Találatok száma / oldal:' (Number of hits / page) set to 20.

At the bottom, there are buttons for 'Keresés' (Search), 'Mentés' (Save), and 'Szűrők törlése' (Clear filters).

3. ábra. A korpusz keresőfelülete

A keresőfunkciók első nagyobb csoportja a szövegek metaadatait szűri, azaz a szövegtestek körét szűkíti le a kereséshez: bevonható és/vagy kizárható egy-egy műcím vagy szerző; szűrhetjük a szövegeket a szerzők neve alapján; megadható a keletkezési időszak intervallumosan; valamint a regények terjedelme is specifikálható egy háromfokozatú skálán. Fontos és újszerű funkció a Regénykorpuszban az irodalmi kanonikusság címkézése, így szűrhetővé válnak a kevésbé vagy jobban kanonikus regények.

A tokenekre, illetve tokenkapcsolatokra való szűrés lehetővé teszi konkrét szóalakok vagy szótövek keresését. Az annotált korpuszok azonban alkalmasak arra is, hogy különböző morfológiai jellemzők alapján úgy keressünk különböző elemekre vagy elemkapcsolatokra, hogy a konkrét szóalakot nem adjuk meg. Mivel az egyes tokenek több szempontú annotációval rendelkeznek, szűrhetők a szóalakok, azaz a szövegbeli előfordulásuk alapján, szótövek alapján, illetve megadható a szófaj és a morfológiai jellemzők is. A kívánt szempontoz tartozó jellemzőket legördülő menüben választhatjuk ki. Több tokenre való keresés esetén megadható a tokenkapcsolat és a tokentávolság is, hogy egyes nyelvi szerkezetek teljes köre kinyerhető legyen a korpuszból.

A harmadik nagyobb csoport a találatok megjelenítéséhez, az eredmények feldolgozásához kapcsolódik. Egyrészt megadható, hogy a találatokat mekkora kontextusban szeretnénk lekérni, illetve tetszőlegesen választható, hogy mekkora egységet szeretnénk egy oldalon megjeleníteni (5–500 találat/oldal). Végül a *Mentés* gomb legördülő menüjén kiválaszthatjuk, hogy a találatok listáját, gyakorisági listát, statisztikát, vagy a kiválasztott regények metaadatait szeretnénk-e menteni. Az így előállított listák (tsv formátum) minden statisztikai alapon nyugvó digitális bölcsészeti kutatás kiindulópontját jelentik; a Regénykorpusz egyik fontos jellemzője, hogy ezeket pár kattintással, komolyabb technikai ismeretek nélkül elérhetővé teszik a felhasználók számára.

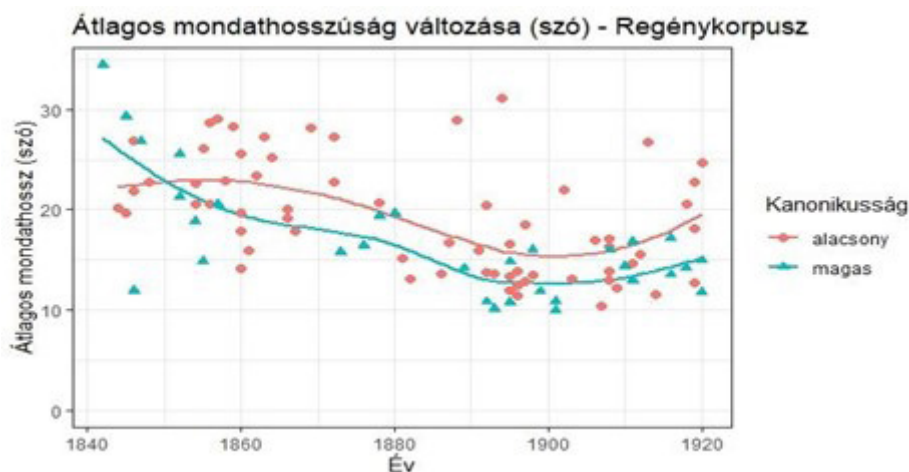
4. Két példa a keresőfelület által létrehozott adatok irodalomtudományos hasznosíthatóságára

A keresőfelület segítségével létrehozott gyakorisági listák nemcsak a legtöbb nyelvstatisztikai megközelítés alapját jelentik, hanem fontos tanulságokkal szolgálhatnak az irodalomtudomány és az irodalomtörténet-írás számára is. Segítségükkel ugyanis a szövegek nyelvi szerveződésének mélystruktúráiba nyerhetünk bepillantást, valamint olyan összefüggéseket tárhatunk fel, amelyek az olvasás folyamatában reflektálatlanok maradnak. Hiszen ezek a keresések a nyelvi működés egy más szintjére vonatkoznak, mint amelyek a befogadás során a figyelem előterébe kerülnek. Ezáltal egyrészt az egyes szövegekre vonatkozó értelmezéseink egészülhetnek ki új szempontokkal, másrészt olyan irodalom- és stílustörténeti folyamatokat tehetünk láthatóvá, amelyek a nagy mennyiségű adatból kiindulva az irodalomtörténet-írás egy reprezentatívabb megvalósulását ígérhetik, amennyiben az így felvázolt folyamatok nem csak egy szűk kánonhoz tartozó szerzők történetére vonatkoznak, hanem az irodalmi termelés egészére kívánnak rálátást biztosítani. A feltárt összefüggések sokszor valóban új tudás létrehozásához járulnak hozzá, azaz újraírhatják eddigi, kizárólag az olvasás tapasztalatán nyugvó megállapításainkat. Más esetekben a statisztikai elemzés eredményei alátámasztják korábbi történeti, vagy a szövegek belső szerveződésére vonatkozó elképzeléseinket – ám ekkor ugyanolyan fontos és hasznos a kereséseket elvégezni, hiszen a korábbi hipotézisek és intuíciók ezáltal statisztikailag alátámasztható, adathozható visszaigazolást nyernek.

Az alábbi egyszerű példák a Regénykorpusz keresőfelületén lekérdezett adatok feldolgozását és vizualizációját foglalják magukban, és az eredmények irodalomtudományos hasznosíthatóságát hivatottak szemléltetni. A példák egyszerűsége jól mutatja, hogy még a meglehetősen banális szempontok mentén létrehozott adatok is a regények elrendezését, összehasonlítását és sajátosságaik kiemelését segíthetik elő.

A 4. ábrán a 100 regény mondatainak a szavak számában megadott átlagos hosszúsága látható egy idővonal mentén. Mivel a keresés lehetővé teszi a kanonikusság szerinti csoportosítását, így az ábrán a magas és alacsony kanonikusságú szövegek összehasonlítása is látható. Az ábrán tisztán kirajzolódik a mondatok hosszúságának csökkenése a 19. század során, ami a magyar prózahagyomány stíláriális átalakulásán túl a

sajtó szerepének megerősödésével, az új íróeszközök elterjedésével, valamint az írás és az olvasás oktatására vonatkozó iskolai reformokkal hozható összefüggésbe.¹⁹ Ezeknek a folyamatoknak az elemzése, valamint a mondatok rövidülésének stílustörténeti értékelése már eddig is az irodalomtörténet-írás részét képezték,²⁰ ám a kvantitatív módszerrel elért eredmények egyrészt adatolható módon igazolják vissza az eddigi belátásokat, másrészt az ábra – mint előre nem értelmezett viszonyrendszereket bemutató képi reprezentáció – lehetőséget teremt a különböző területek (stílus-, sajtó-, oktatás-, médiatörténet) összekapcsolására is a kirajzolódó tendenciák értelmezése során. Érdekes kiemelni továbbá, hogy a magas és alacsony kanonikusságú szövegek egyaránt követik ezt a tendenciát – érdekes azonban, hogy míg a 19. század első harmadában a magas kanonikusságú szövegekhez tartozik egy inkább hosszú mondatos prózastílus, addig a század többi részében ez az arány megfordul. Ennek a megállapításnak az érvényességét egy nagyobb korpuszon végzett kutatás igazolhatja: a Regénykorpusz adatbázisának bővülése ennek lehetőségét is megteremtheti. Ez azért is lenne különösen fontos, mert a kanonikus és nem kanonikus szövegek összehasonlító elemzése (azaz annak a kérdésnek a vizsgálata, hogy a kánonképződés során kiválasztott szövegek rendelkeznek-e sajátos, elkülöníthető textuális jellemzőkkel a különböző történeti korokban) ezidáig az irodalomtörténetírás perifériájára szorult.



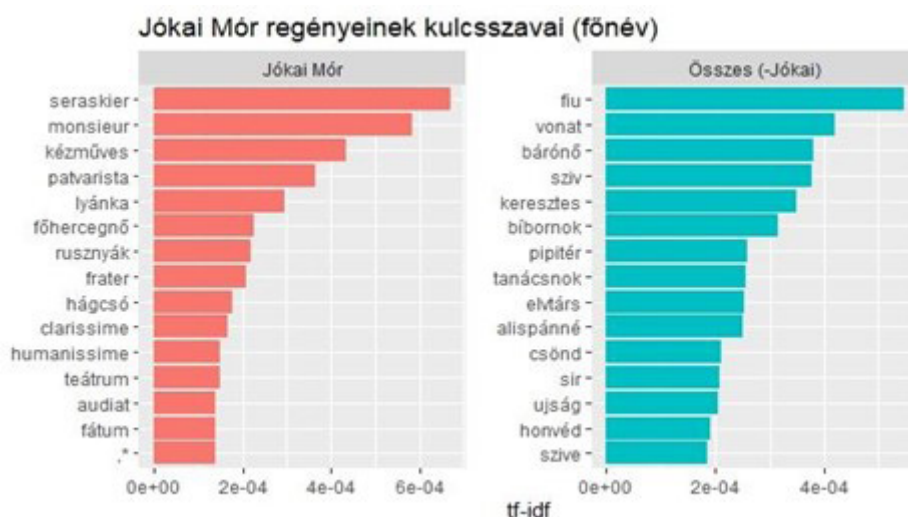
4. ábra. Átlagos mondatösszehosszúság a Regénykorpuszban

Az 5. ábra a Jókai-alkorpusz, valamint a korpusz további 97 regényének összehasonlítását mutatja. Ehhez a Regénykorpusz felületén lekérdezett gyakorisági listákat az ún. „kifejezés gyakoriság–fordított dokumentum gyakoriság” (term-frequency-inverse document frequency, TF-IDF) módszerével hasonlítottuk össze. A módszer lényege, hogy a korpuszok egymáshoz képesti kulcsszavait tudjuk feltárni: kulcsszónak azok a kifejezések minősülnek, amelyek az egyik korpuszban magas számban, míg a

¹⁹ A mondatösszehosszúságok változásának részletesebb elemzését lásd: Szemes Botond, Mondatösszehosszúság és irodalomtörténet. 100 magyar regény szövegstatistikai elemzése, *Literatura* (2020):3: 335–367.

²⁰ Pl. Herczeg Gyula, *A XIX. századi magyar próza stílusformái*, Budapest, Tankönyvkiadó, 1981.

másikban ritkán fordulnak elő – és viszont.²¹ Az ábra Jókai Mór tárgyalt regényeinek egy karaktertípusára és a hozzá tartozó nyelvhasználatra hívja fel a figyelmet: a franciás kifejezések a korpusz többi regénye esetében nem tekinthetők jellemzőnek. Az egyes karakterekhez kötődő francia nyelvhasználat már az olvasás tapasztalatában is reflektálttá válik, ugyanakkor annak kimutatásához, hogy ez mennyiben tekinthető a Jókai-próza sajátosságának, elengedhetetlen a fent vázolt módszer alkalmazása.



5. ábra. Jókai Mór regényeinek kulcsszavai

5. Összefoglalás és kitekintés

A tanulmány célkitűzése az ELTE BTK Digitális Bölcsészet Tanszék által létrehozott és fejlesztett Regénykorpusz projektum jelenlegi stádiumának és felhasználási lehetőségeinek összefoglaló bemutatása volt. A tanulmányban igyekeztünk felvázolni a korpuszban való keresés lehetőségeit, illetve bemutatni néhány egyszerű példán keresztül azt, hogy milyen potencialitással rendelkezik a gyűjtemény és annak keresőfelülete akár az irodalomtörténeti kérdések megválaszolásában. A Regénykorpusz esetében fontos hangsúlyozni azt, hogy a projekt korántsem tekinthető lezárt munkának. A jövőben tervezzük a korpuszban elérhető szövegek bővítését, s ezzel együtt azt, hogy lazítjuk a kiválasztási kritériumokat (például felvesszünk olyan műveket is, amelyek 1840 előtt, vagy 1920 után kerültek kiadásra), ezáltal a felület alkalmas lesz akár diakrón vizsgálatokhoz való felhasználásra. Ezzel párhuzamosan tervezzük a keresési lehetőségek bővítését is. Jelenleg azon dolgozunk, hogy elérhetővé váljon egy tagmondatkapcsolat-felismerő funkció a felhasználói felületen, ami többek között mondatstilisztikai kutatások számára teremthet majd alapot.

²¹ Vö.: Shahzad Qaiser, Ramsha Ali, „Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents” *International Journal of Computer Applications* (0975 –8887) Volume 181. (2018):1. Eun-Soon You, Gun-Hee CHOI, Seung-Hoon KIM, „Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels”, *Journal of the Korea Society of Computer and Information*, 20/2, (2015): 121–129.



6. Bibliográfia

- Lou Burnard, Christof Schöch, Carolin Odebrecht, *In Search of Comity: TEI for Distant Reading*, 2019, hozzáférés: 2021.05.26. <https://doi.org/10.5281/zenodo.3552489>
- Eun-Soon You, Gun-Hee Choi, Seung-Hoon Kim, „Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels”, *Journal of the Korea Society of Computer and Information*, 20/2, (2015): 121–129.
- Horváth Péter, „Az ELTE Verskorpusz automatikus annotációs eljárásai révén nyerhető kvantitatív adattípusok”, In *Nyelvtan, diskurzus, megismerés, szerkesztette Simon Gábor, Tolcsvai Nagy Gábor*, 313–331. Budapest: Eötvös Kiadó, 2020.
- Indig Balázs, Sass Bálint, Simon Eszter, Mittelholcz Iván, Kundráth Péter, Vadász Noémi, „emtsv – egy formátum mind felett”, In *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette Berend Gábor, Gosztolya Gábor, Vincze Veronika, 235–247. Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet. 2019.
- Kalcsó Gyula, „A TEI-XML felhasználása magyar nyelvű korpuszok építésében”, In *MANYE XX. Az alkalmazott nyelvészet ma: Innováció, technológia, tradíció*, Boda István, Mónos Katalin szerkesztette, 65–72. Debrecen: MANYE, Debreceni Egyetem, 2011.
- Mittelholcz Iván, „emToken: Unicode-képes tokenizáló magyar nyelvre”, In *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Vincze Veronika szerkesztette, 61–69. Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2017.
- Carolin Odebrecht, Lou Burnard, Christof Schöch, *COST Action Distant Reading for European Literary History (CA16204)*, 2021, hozzáférés: 2021.05.26., <https://doi.org/10.5281/zenodo.4662444>
- Christof Schöch, Maciej Eder, Arias Rosario, François Pieter, Antonija Primorac, *Foundations of Distant Reading. Historical Roots, Conceptual Development and Theoretical Assumptions around Computational Approaches to Literary Texts*, hozzáférés: 2021.05.26. <https://dh2020.hcommons.org/>
- Shahzad Qaiser, Ramsha Ali, „Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents”, *International Journal of Computer Applications (0975–8887) Volume 181*. (2018):1.
- Szemes Botond, Mondathosszúság és irodalomtörténet. 100 magyar regény szövegstatistikai elemzése, *Literatura* (2020):3: 335–367.

Automatikus kézírás-felismertetés Kiss József levelezésén

Szűcs Kata Ágnes

Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ (Budapest)

szucs.kata@dbk.pim.hu

Abstract

The digital edition of the József Kiss correspondence is a pilot project of the Centre for Digital Humanities, Petőfi Literary Museum. In addition to the processing of the personal and professional letters of the 19th-century writer, poet, and editor of the literary journal *A Hét* (The Week), the project is to explore the possibilities offered by the Transkribus software.

Handwritten Text Recognition is an emerging field of the digital humanities. The paper will discuss this artificial intelligence-based technology and our experiences in creating a Hungarian model. The best result has a 6,94% character error rate. Apart from the practical experience gained in testing, the paper discusses the possibilities of implementing HTR in public collections.

Keywords: HTR, handwritten text recognition, Transkribus, correspondence, 19th-century literature

1. Bevezető

Az alábbi esettanulmány a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ Handwritten Text Recognition (HTR), azaz a kézírás-felismertetéssel kapcsolatos eddigi munkáját, és a Transkribus használatát hivatott bemutatni. Rövid elméleti áttekintés után, ismertetem a témában szerzett tapasztalatainkat, milyen eredményeket kaptunk a modellek tanításakor, és hogyan sikerült javítani rajtuk.

A modellépítés alapját a PIM Kézirattárában folyamatos feldolgozás alatt álló Kiss József-levelezés képezi. A 493 leltári tételből álló korpusz 1683 darab személyes vagy szakmai levelet jelent, melyeknek címzettje vagy írója Kiss József, a 19–20. század fordulóján élt költő és lapszerkesztő. Az ő nevéhez kötődik *A Hét* című hetilap alapítása, mely csomópontot jelentett a kialakulóban lévő polgári értelmiség képviselőinek, mintegy alapot teremtve a későbbi irodalmi folyóirat, a *Nyugat* számára is. A levelezés kiadása hiánypótló és irodalomtörténeti szempontból is jelentőséggel bír. Emellett lehetőséget ad arra, hogy a digitalizált kéziratok gyűjteményi kezelésére, digitális források kiadására alakítsunk ki alternatívákat.

2. Automatikus kézírás-felismertetés (HTR)

A kézírás felismerését célzó technika sokáig együtt fejlődött az OCR-rel (optikai karakterfelismerés), ahol a szkennelt dokumentumok nyomtatott szövege válik gép által olvashatóvá. Az OCR technológiában az egyes karakterek képezik a felismerés



alaját, melyeket előre megadott mintákkal hasonlít össze. A HTR különálló kutatási területté fejlődött a 2000-es évek óta, a kézírások különbözősége és a feladat számítási komplexitása miatt.¹

Az egyik fő különbség a két technológia között, hogy a HTR egy szegmentált sor szövegében lévő összes karakter felismerésére fókuszál.² A gépi tanulással támogatott technológia képes a vizuális jegyek elsajátítására (így nem kell külön mintákat létrehozni), a neurális háló segítségével pedig több egymást átfedő szövegsor képéből képes karakter valószínűséget számítani.³⁴ Természetesen a HTR technológiát nem csak kézzel írt szövegeken lehet alkalmazni.

3. Transkribus

Szegmentálás, szöveg átírása

Az ingyenesen letölthető és használható szoftver megkönnyíti a kéziratokkal való munkát. Az átírás sorról sorra történik a digitális faksimile folyamatos jelenlétében. Verziókövető rendszerrel van ellátva a program, tehát minden mentés bármikor visszaállítható. Felhő alapú, ezért a megfelelő elővigyázatosságok mellett egy gyűjteményen egyszerre többen is dolgozhatnak párhuzamosan. Sokrétű címkézési és metaadatolási rendszer van beépítve. A fájlokat többféle kimeneti formátumban lehet exportálni (.pdf, .tei, .docx, txt, .xlsx, .zip, stb.) A Transkribus emellett lehetőséget biztosít az automatikus kézírás-felismertetésre és a kézírást felismerő modell létrehozására is.

Magyar nyelven még nem készült nyilvánosan elérhető modell, így azt a program segítségével kezdtük el építeni.⁵ Ehhez a szkennelt kéziratképeket fel kell tölteni a Transkribus szervereire, majd előállítani egy minimum 5–15000 szót tartalmazó, átírással rendelkező korpuszt. A nyomtatott szövegek esetén kevesebb is elengedő lehet. Az átírást a Transkribuson belül is létre lehet hozni, de meglévő átírat akár utólag is hozzárendelhető a képekhez (Text2Image funkció).⁶

Az átíráshoz a képeket szegmentálni kell. A digitális képfeldolgozásban a szegmentálás egy kép több szegmensre (pixelhalmazokra, más néven képjektumokra) történő

- 1 Katuščak Dušan, „Automated Transcription of Handwritten Text: READ and TRANSKRIBUS (An Experiment with Transcribing Letters of Andrej Kmet)”, 2019. október 20.
- 2 Puigcerver Joan, „Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?”, in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 01, 2017, 67–72, <https://doi.org/10.1109/ICDAR.2017.20>.
- 3 Dietrich Felix, „OCR vs. HTR or “What Is AI, Actually?””, *READ-COOP* (blog), elérés 2021. június 2., <https://readcoop.eu/insights/ocr-vs-htr/>.
- 4 Puigcerver Joan, „Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?”, in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 01, 2017, 67–72, <https://doi.org/10.1109/ICDAR.2017.20>.
- 5 „Public Models in Transkribus”, *READ-COOP*, elérés 2021. május 26., <https://readcoop.eu/transkribus/public-models/>.
- 6 Vö.: „How To Use Existing Transcriptions to Train a HTR-Model with the TextToImage-Tool”, *READ-COOP*, elérés 2021. június 7., <https://readcoop.eu/transkribus/howto/how-to-use-existing-transcriptions-to-train-a-handwritten-text-recognition-model/>.

felosztásának folyamata, célja a kép ábrázolásának egyszerűsítése.⁷ A folyamat során olyan területekre osztjuk a képet, melyek meghatározott koordináták alapján felismerhetővé teszik a sorok és bekezdések helyzetét a neurális háló számára. Három réteget jelölünk ki a képen poligonok formájában (Text Region, Line és a Base Line). A szegmentálás manuálisan és automatikusan⁸ (ill., a kettőt keverve) is elvégezhető. Míg az előbbi pontosabb, de időigényesebb, az utóbbinál szükség van az utólagos korrekcióra.

A szegmentálás után megkezdődhet a szövegek átírása is, ami ily módon a sorok szintjére lebontva eleve összekötetésben van a hozzá tartozó képpel.⁹

A kézírás-felismertetés szempontjából fontos átíráskor a lehető legpontosabb, betű szerinti átírásra törekedni, és a karaktereket következetesen ugyanúgy megadni.¹⁰ Az Unicode határain belül lehetőség van speciális karakterek beszúrására is.

HTR modell építése

Az átírással rendelkező képeket két csoportba osztjuk, melynek során egy Training set (TS) és egy Validation set (VS) jön létre. Lehetőség van ezek automatikus válogatására is, ami 2%, 5%, vagy 10% VS-eket hoz létre. A TS a rendelkezésre álló fájlok kb. 90%-át teszi ki.

A program a TS-be került fájlokban mesterséges intelligencia segítségével azonosítja a sorokban látható írásképet az átírt szöveggel. Egy tanulási ciklus (epoch) során végigmegy a TS összes fájlján és vizuális jegyek alapján kitalálja, hogy a szegmentált sorok milyen karakterekből állnak össze. Ezután ellenőrzi magát az általunk megadott átírásra hagyatkozva. A tanulási folyamat több ilyen ciklusból áll össze, és a következő epoch-ba már az előzőből elsajátított tudással kezd bele. Az epoch-ok száma a dokumentumok minősége szerint változtatható (max. 250).

Végül a másik csoporton, a VS-en teszteli le magát a mesterséges intelligencia. Ennél a csoportnál csak a TS-en elsajátítottakra és a képen látható vizuális jegyekre hagyatkozik. A VS-ben lévő fájlokban egyszer megy végig, majd ellenőrzi magát az emberi intelligencia által készített átírás alapján. Az itt kapott hibaérték (CER on Validation set) azt jelzi, hogyan teljesít a modell egy ismeretlen szövegen. Az eredményességet tovább lehet növelni Base Model beépítésével. Ilyenkor egy másik HTR modellből már elsajátított tudást építünk be.

7 Srinivasan G N, „Segmentation Techniques for Target Recognition” 1, 3 (2007): 7.

8 Elforgatott rájegyzések, szokatlan oldaltörések (pl. borítékok, képeslapok) esetében többet téveszt.

9 A Transkribus lehetőséget biztosít a szavak szintjén történő szegmentálásra is, de mivel a funkció használatát egy rövid tesztidőszakot követően elvetettük, a továbbiakban erről nem lesz szó

10 Például egy német szövegnél a sárfesz s-t egyféleképpen, ß karakterrel, vagy dupla s-sel jelöljük, nem keverve. A hiányzó szövegrészeket pedig jobb üresen hagyni.



Az eddigi legjobb eredményt a KEZ17_Kiss József kézírása_5 nevű modellel sikerült elérni, ahol a CER on Validation Set értéke 6,94% volt. Ez azt jelenti, hogy a modell által készített átírásban könnyen javítható hibák találhatók (pl. ékezetek, egy-egy betűtívesztés).

4. Tapasztalataink

4.1 Első fázis

Egy modell kialakításakor fontos, hogy a korpuszban jobbra egy kéz által írt szövegek legyenek, melyek korban közel keletkeztek egymáshoz, illetve, hogy a forrás típusa is egyező legyen (pl. napló, számadáskönyv, levelezés stb.).¹¹ Először létrehoztunk egy csak Kiss József kézírásából és egy mindenki máséból álló korpuszt, ahova folyamatosan kerültek be újabb átírt anyagok a neurális háló számára.

Eleinte kevés kapaszkodót találtunk az eredmények értékelésére. A modellek finomhangolására nézve saját magunknak kellett egy viszonyrendszert kialakítani. Először is felmértük a rendelkezésre álló anyag összetételét. Erre különböző csoportokat alakítottunk ki, felcímkézve és táblázatos formában nyilvántartva a levelek nyelvét, a kézírás eszközét (ceruza vagy toll), illetve, hogy géppel írt vagy nyomtatott a dokumentum. Úgy véltük ezek olyan faktorok, amelyek befolyásolhatják a kézírás-felismertetést végrehajtó neurális háló viselkedését. A gépi feldolgozás során az algoritmus figyelembe veszi az adott nyelvben előkerülő karakterek számát és azok egymáshoz viszonyított előfordulási gyakoriságát is. Másrészt a nyomtatott és géppel írt betűk más vizuális jegyekkel rendelkeznek, mint a kézzel írtak.

homokozó, próba 2		kiss_jozsef_masolat, Kiss József kézírása				kj 2		kj3	
validation set	training set	validation set	training set	validation set	training set	validation set	training set	validation set	training set
CER	61	CER	194	CER	15	CER	22	CER	229
9,23%	0,95%	12,66%	1,90%	17,29%	1,32%	9,72%	1,89%		
0 átlót a tinta	2 átlót a tinta	0 átlót a tinta	2 átlót a tinta	0 átlót a tinta	0 átlót a tinta	0 átlót a tinta	2 átlót a tinta	0 átlót a tinta	2 átlót a tinta
0 boríték	1 boríték	0 boríték	27 boríték	0 boríték	1 boríték	0 boríték	3 boríték	0 boríték	33 boríték
3 ceruza	23 ceruza	5 ceruza	63 ceruza	4 ceruza	34 ceruza	4 ceruza	72 ceruza	0 ceruza	72 ceruza
0 címzés	1 címzés	4 címzés	6 címzés	0 címzés	0 címzés	0 címzés	1 címzés	0 címzés	15 címzés
0 firka	1 firka	0 firka	0 firka	0 firka	0 firka	0 firka	0 firka	0 firka	2 firka
0 géppírat	1 géppírat	0 géppírat	1 géppírat	0 géppírat	0 géppírat	0 géppírat	0 géppírat	0 géppírat	0 géppírat
0 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás	1 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás	0 nem KJ kézírás
0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	0 gép. firka	10 gép. firka
1 námet	1 námet	0 námet	0 námet	0 námet	0 námet	0 námet	0 námet	0 námet	1 námet
0 nyomtatott kézírás	0 nyomtatott kézírás	1 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás	0 nyomtatott kézírás
3 toll	31 toll	12 toll	94 toll	11 toll	82 toll	14 toll	99 toll		

1. ábra Az első modellek

Az első modellépítések eredményei a folyamatosan bővülő szövegtörzs miatt nem voltak teljes mértékben összehasonlíthatók, de rengeteg tapasztalatot szereztünk, és kijelölték a további tesztelések irányát.

4.2 Második fázis

Ekkorra a Kiss József által írt levelek nagy részét feldolgoztuk, ez kb. 370 felhasználható oldalt, több mint 20.000 szót jelentett. Az elképzelés szerint egyféle anyagon többféleképpen teszteltük a modelleket.

¹¹ Vö.: <https://readcoop.eu/transkribus/resources/how-to-guides>

	KJ_keze_valogatás_1	KJ_keze_valogatás_2
Base Model	kj3	nincs
Train set	aut_ts_2	aut_ts_2
pages	302	302
lines	4170	4170
words	18941	18941
Validation set	aut_vs_2	aut_vs_2
pages	33	33
lines	558	558
words	2681	2681
Eredmények		
CER on Train Set	2,81%	6,11%
CER on Validation Set	6,13%	11,49%
Nr. of Epochs	50	50
Összes fájl	335	335

	KJ_keze_osszes_1	KJ_keze_osszes_2
Base Model	kj3	nincs
Train set	aut_ts_1	aut_ts_1
pages	330	330
lines	4399	4399
words	20039	20039
Validation set	aut_vs_1	aut_vs_1
pages	36	36
lines	459	459
words	2120	2120
Eredmények		
CER on Train Set	3,05%	6,54%
CER on Validation Set	5,82%	10,70%
Nr. of Epochs	50	50
Összes fájl	366	366

2. ábra A Válogatás gyűjtemény eredményei az Összes gyűjteményhez képest

Először megbizonyosodtunk arról, hogy a tollal, ceruzával és nyomtatottan írt Kiss József levelek külön kezelését megszüntethetjük. A *Válogatás* gyűjteményben csak Kiss József kézzel írt, magyar nyelvű levelei szerepeltek, ezzel párhuzamosan egy másikban az összes általa írt levél. A válogatott gyűjteményen futtatott modell eredménye nem jobban, hanem 1–2%-kal rosszabbul teljesített a másikhoz képest (Vö.: 1. ábra CER on Validation Set sorai). A válogatás minősége fontos, mindkét set-nek megfelelően kell reprezentálnia a korpusz egészének összetételét.

Model name	Selection	Selection type	Base model	SUM	Train Set (pages)	lines	words	Validation set (pages)	lines	words	Epoch	Set	CER on Train	CER on Validation Set	Collection ID	Collection name
KJ_keze_osszes_1_auf1	minden átlírt oldal (10%)	Kiss József kézírása_3		306	300	4399	20039	36	459	2120	50		3,05%	5,82%	79300	KJ_keze_osszes_1
KJ_keze_osszes_2_auf1	minden átlírt oldal (10%)	nincs		306	300	4399	20039	36	459	2120	50		6,54%	10,70%	79300	KJ_keze_osszes_1
KJ_keze_osszes_3_manu1	minden átlírt oldal	KJ_keze_osszes_1		369	306	4413	20014	33	475	2264	50		2,26%	4,26%	79300	KJ_keze_osszes_1
KJ_keze_osszes_4_manu1	minden átlírt oldal	Kiss József kézírása_3		369	306	4413	20014	33	475	2264	50		2,58%	8,38%	79300	KJ_keze_osszes_1
KJ_keze_osszes_5_manu1	minden átlírt oldal	KJ_keze_osszes_1		369	306	4413	20014	33	475	2264	100		1,77%	6,08%	79300	KJ_keze_osszes_1
KJ_keze_osszes_6_auf1	minden átlírt oldal (10%)	KJ_keze_osszes_1		306	300	4399	20039	36	459	2120	50		2,13%	6,59%	79300	KJ_keze_osszes_1

3. ábra Második fázisban keletkezett HTR modellek

Ebben a fázisban a Base Model (BM) használatával is kísérleteztünk, két modell között mindig csak egy dolgot megváltoztatva. A KJ_keze_osszes_4 modellnél az automatikus besorolást felváltotta egy manuális, szélesebb körű válogatás. Az eredményeken mégis az látszik, hogy az összes próbálkozás közül ennek a legmagasabb a karakter hiba értéke. Egy másik alapmodellel ugyanez az összeállítás nagyon jó eredményt hozott. Az epoch-ok számát duplázva, a KJ_keze_osszes_5 modellnél szintén rosszabb eredményeket kaptunk. És végül, a hatos modellnél az automatikus TS és VS válogatást alapul véve, egy másik alapmodell szintén rosszabb eredményeket hozott. Ezek alapján az számít, hogy melyik alapmodellt melyik korpuszon használjuk; a legjobb eredményt a hasonló adatokon tanult BM adja.

Az első két esetben a Base Modellel rendelkező modell kb. kétszer jobban teljesített. A többinél azonban már nem ennyire egyértelmű a kimenetel. Ennek oka a BM használatban rejlik. A folyamatos dokumentumfeldolgozás miatt a Base Model és a Training data (VS + TS) között átfedések keletkeztek, ami hamis pozitív eredményt és túltanulást is okozott ezeknél a modelleknél. Tehát, a második fázis legjobb eredményű modellje (4,26%) egy teljesen új próbaszövegen lefuttatva valójában rosszabbul (8–10%-os pontossággal) teljesített.



4.3 Harmadik fázis

A befolyásoltság (bias)¹² kizárására a harmadik fázisban került sor. Három, az alábbi ábrán zöld árnyalatokkal jelölt szövegtörzset hoztunk létre, és a Base Modell használatát egyelőre megszüntettük, azt a teljes vegyes kézírásmodell tanításakor alkalmazzuk újra.

Model name	Selection	Selection type	Train Set	Validation set	Epoch	CER on Validation Set	CER on Train Set	Collection ID	Collection name
KEZ_12_KI_keze_osszes_7	6. adag kivételével	manu	434	48	25	11,71%	15,09%	93264	KEZ12 Kévéllő (6. adag kivételével)
KEZ_12_KI_keze_osszes_8	6. adag kivételével	manu	434	48	50	8,73%	6,62%	93264	KEZ12 Kévéllő (6. adag kivételével)
KEZ_12_KI_keze_osszes_9	6. adag kivételével	manu	434	48	75	8,19%	4,38%	93264	KEZ12 Kévéllő (6. adag kivételével)
KEZ_12_KI_keze_osszes_10	6. adag kivételével	manu	434	48	100	8,21%	3,45%	93264	KEZ12 Kévéllő (6. adag kivételével)
KEZ_15_friss_kj_kziras_1	külön tesztkorpusz	manu	406	42	75	7,45%	4,35%	95196	KEZ15 Friss KJ kézírása modell
KEZ_15_friss_kj_kziras_2	külön tesztkorpusz	manu	406	42	100	7,76%	3,87%	95196	KEZ15 Friss KJ kézírása modell
KEZ_15_friss_kj_kziras_3	külön tesztkorpusz	aut	405	45	75	14,19%	3,81%	95196	KEZ15 Friss KJ kézírása modell
KEZ_15_friss_kj_kziras_4	külön tesztkorpusz	aut	405	45	75	13,81%	4,01%	95196	KEZ15 Friss KJ kézírása modell
KEZ_15_friss_kj_kziras_5	külön tesztkorpusz	aut	405	45	75	12,16%	4,74%	95196	KEZ15 Friss KJ kézírása modell
KEZ_15_friss_kj_kziras_6	külön tesztkorpusz	aut	405	45	100	13,65%	2,78%	95196	KEZ15 Friss KJ kézírása modell
KEZ17_Kiss József kézírása_1	teljes KJ kézírás (TK nélkül)	manu	455	42	75	7,39%	4,75%	95196	KEZ15 Friss KJ kézírása modell
KEZ17_Kiss József kézírása_2	teljes KJ kézírás (TK nélkül)	manu	455	42	100	7,15%	3,81%	95196	KEZ15 Friss KJ kézírása modell
KEZ17_Kiss József kézírása_3	teljes KJ kézírás (TK nélkül)	manu	455	42	125	7,11%	2,99%	95196	KEZ15 Friss KJ kézírása modell
KEZ17_Kiss József kézírása_4	teljes KJ kézírás (TK nélkül)	manu	455	42	150	7,01%	2,58%	95196	KEZ15 Friss KJ kézírása modell
KEZ17_Kiss József kézírása_5 / József Kiss' handwriting 19-20th century	teljes KJ kézírás (TK nélkül)	manu	455	42	200	6,94%	2,13%	95196	KEZ15 Friss KJ kézírása modell

4. ábra A harmadik fázisban létrehozott HTR modellek

A korpuszokat a szöveg mennyisége és a válogatás folyamata különböztette meg egymástól. Az elsőben (kez_12) a második fázisban alkalmazott felosztást és szövegtörzset meghagytuk, de új modellépítés történt (base modell nélkül). A legjobb hibaszázalék-eredmény 8% körüli lett. A második csoporton belül (kez_15) elkülönítettünk egy tesztkorpuszt, és a fennmaradó anyagból manuálisan alakítottunk ki TS-t és VS-t. A megfelelő mennyiségű epoch után 8% alá ment a CER értéke. Az automatikus válogatás a második csoporton csak rosszabb eredményeket hozott. A harmadik csoport (kez_17) abban tér el az eddigiektől, hogy az eredeti felosztás mellett a tesztkorpusz anyaga is bekerült a végleges TS-be. Itt értük el a legjobb, 6,94%-os eredményt, bias nélkül.

5. Javítási lehetőségek

Minél több adat áll a neurális háló rendelkezésére, annál jobb eredményt fogunk kapni. Egy bias nélküli Base Model használatával felgyorsul a training process, és kevesebb training data is elég igen jó eredmények eléréséhez. Fontos, hogy a BM szöveganyaga hasonlítson a felismertetésre váró szöveghez, és hogy minden típusú forrás reprezentálva legyen.

Szótárak beépítésével a szövegben előforduló specifikus kifejezéseket lehet megtanítani a MI segítségével, mellyel tovább lehet javítani a modell eredményességén. Ehhez fel kell venni a kapcsolatot a Transkribus csapatával.

¹² A statisztikai bias során az eredmények várható értéke eltér a becslés alapjául szolgáló valódi kvantitatív paramétertől. A minta torzított/elfogult, ha nem reprezentatív, ha a vizsgált tulajdonság megoszlása a mintában nem ugyanaz, mint az alapsokaságban. Ennek három oka lehet: (1) a mintavételi eljárás elfogult, (2) a minta elégtelen, (3) túl messzire megy az általánosítás. Margitay Tihamér, *Az érvelés mestersége: érvelések elemzése, értékelése és kritikája* (Budapest: Typotex, 2007) (11.5. Statisztikus következtetések és hibáik).

A túltanulás elkerülése végett az epoch-ok számát a tanulási görbe változásának figyelembevételével azon a ponton kell megállítani, ahol az a legalacsonyabban van.

Problémákat okozhat a HTR modell tanulása során a szegmentálás pontatlansága is. Érdeemes tehát ellenőrizni, és manuálisan javítani a szegmentálást, figyelembe véve, hogy a rendszer a baseline és a kijelölt sor alapján próbálja megtalálni a szöveges információkat.

A különböző modellek egymásba építése esetén pedig a TS és VS adatai nem szabad, hogy keveredjenek egymással. Így elkerülhető a fals pozitív eredmény a karakterhiba-értékben.¹³

6. Legújabb eredmények – 6,94%



5. ábra A Ground Truth-ként elmentett átírás és a HTR által elvégzett átírat közötti különbségek

A Transkribus-ban összehasonlítható az egyes szövegek különböző verziója. Látható, hogy a 6,94%-os CER esetében a tévesztések nem értelemzavaróak, többnyire gyorsan javítható ékezetekről, írásjelekről, kis-nagybetű tévesztésről van szó. A későbbiekben az 5% alatti eredmények elérése a cél.

7. Következő lépések

A Transkribus szerves részét képezi a PIM-ben őrzött kéziratos hagyatékok, eddig publikálatlan kéziratok digitális feldolgozása. Ahogy jelenleg az átírást, a jövőben az automatikus kézírás-felismertetést is szeretnénk a workflow részévé tenni. A Kiss József-projekt ideális az igények felmérésére, a fejlesztések és eszközök tesztelésére. Az egyik kimeneti formátum, melyet használunk a kétrétegű PDF. Előnye, hogy a Transkribusból közvetlenül állítható elő és azonnal publikálható a PIM online katalógus felületén (PIM OPAC).¹⁴ A másik, még tesztelés alatt álló formátum a TEI XML, amely egy következő publikálási stádiumban lesz elérhető. Az átírás végén ez szintén exportálással nyerhető

¹³ Ha olyan fájlok kerülnek a VS-be, melyeken a mesterséges intelligencia a TS-en tanult, akkor A CER on Validation Set jó százalékos eredményt mutat, de az nem egyezik a modell valós teljesítményével teljesen ismeretlen szövegen.

¹⁴ A PIM Kéziratgyűjteményében található Kiss József-levelezés: [https://resolver.pim.hu/gyujtemeny/levelek/"Kiss József 1843-1921"](https://resolver.pim.hu/gyujtemeny/levelek/).



ki a Transkribusból, viszont megfelelő megjelenítő felület fejlesztése szükséges hozzá, ahol fontos, hogy jól érvényesüljön a szöveg és a faksimilék sor szintű összeköttetése, s az átírt szöveget olvasva látszik a kontextus is.

Megkezdődött egy vegyes kézírásos alapuló modell építése. Ezután elérjük azt a szövegmennyiséget, hogy megtörténhet az első magyar nyelvű kézírásmodell publikálása is. Hosszútávon a PIM Kézírtárában lévő egyéb írók kézírásából álló modellek kialakításával, és ezek egymásba építésével egy általánosabb modellt szeretnénk megalkotni, amely a 19–20. századi magyar kézírások automatikus felismertetését teszi majd lehetővé.

Bibliográfia

- Dietrich, Felix. „OCR vs. HTR or “What Is AI, Actually?””. *READ-COOP* (blog). Elérés 2021. június 2. <https://readcoop.eu/insights/ocr-vs-htr/> .
- „How To Use Existing Transcriptions to Train a HTR-Model with the TextToImage-Tool”. *READ-COOP*. Elérés 2021. június 7. <https://readcoop.eu/transkribus/howto/how-to-use-existing-transcriptions-to-train-a-handwritten-text-recognition-model/> .
- Katušćak, Dušan. „Automated Transcription of Handwritten Text: READ and TRANSKRIBUS (An Experiment with Transcribing Letters of Andrej Kmeť)”, 2019. október 20.
- Margitay, Tihamér. *Az érvelés mestersége: érvelések elemzése, értékelése és kritikája*. Budapest: Typotex, 2007.
- „Public Models in Transkribus”. *READ-COOP*. Elérés 2021. május 26. <https://readcoop.eu/transkribus/public-models/>.
- Puigcerver, Joan. „Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?” In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 01:67–72, 2017. <https://doi.org/10.1109/ICDAR.2017.20> .
- Srinivasan, G. N., Shobha, G. „Segmentation Techniques for Target Recognition” *INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS*, 1, 3 (2007): 75.

19. századi magyar regényeken végzett kísérletek a magyar nyelvű szerzőazonosítás leghatékonyabb távolságméréseinek megtalálására

Timári Mária

ELTE DH

mariazimanyi@btk.elte.hu

Bajzát Tímea Borbála

ELTE DH

timimimi@student.elte.hu

Palkó Gábor

ELTE DH

palko.gabor@btk.elte.hu

In the field of computational stylistics it is a widespread assumption that there exists a unique pattern of a person's language use and this is a so-called authorial fingerprint (Baayen et al. 2002). Identifying the authorial fingerprint can become a base of using quantitative text similarity studies for authorship attribution. Although the metaphor of the "fingerprint" may give the false impression that this pattern can be read from the author's texts in an objective way. The modelling of this "fingerprint" is in fact a creative digital humanities task which constructs a pattern that is based on a selection and combination of dozens of linguistic features that can be interpreted statistically and that can only be interpreted in comparison with other authorial texts.

Considering the size of the corpora and the complexity of text characteristics and similarity calculations, the method cannot be done without the use of computer algorithms. There are already some softwares (e.g. R-Stylo, JGAAP, Websty) that apparently offer an accessible way to researchers for analyzing texts, (however these methods limit the process of searching for patterns), while on the other hand it is possible to implement the calculations using custom program codes.

In our research we endeavour to find the most efficient measures in authorship attribution for Hungarian texts.

1. Bevezetés

A számítógépes stilsztika területén széles körben elfogadott nézet, hogy léteznek az egyes ember nyelvhasználatát jellemző egyedi mintázatok, szerzői „ujjlenyomatok”¹, amelyek feltérképezése megalapozhatja a kvantitatív szöveghasonlósági vizsgálatok felhasználását szerzőazonosítás céljaira.

Az „ujjlenyomat” metaforája ugyanakkor azt a téves képzetet keltheti, hogy a szerzőre jellemző ésszámszerűsíthető mintákaszövegekből objektív módon volnának kiolvashatók.

1 Harald et al., 2002.



Az „ujjlenyomat” megalkotása valójában egy olyan kreatív digitális bölcsészeti feladat, amely a szövegstatistikailag értelmezhető több tucatnyi nyelvi jellemző, majd az ezekre épülő különféle hasonlósági számítások szelekciója és kombinációja alapján konstruál egy, mindig csak más szerzői szövegekkel való összehasonlítás horizontján értelmezhető „mintázatot”.

Mindez – tekintettel a szövegtörzsek méretére, a szövegjellemzők és a hasonlósági számítások komplexitására – ma már elképzelhetetlen számítógépes algoritmusok használata nélkül. Léteznek (látszólag) kész megoldásokat nyújtó, a kutatók számára könnyebben elérhető, de a mintázatkeresés folyamatát leszűkítő szoftveres megoldások (pl. R-Stylo, Websty, JGAAP), másrésztől lehetséges a számítások implementálása egyedi programkódok segítségével is.

Kutatásunkban arra vállalkozunk, hogy 19. századi magyar regényeken vizsgálva – előtérbe helyezve Jókai Mór szövegterének vizsgálatát – megtaláljuk azokat a beállításokat, illetve távolságmértékeket, melyek a magyar nyelvű szerzőazonosítási és klaszterezési eljárások során a legpontosabb eredményekkel szolgálnak.

2. A vizsgálatokhoz felhasznált (al)korpuszok

A legelső és a legfontosabb lépés annak irányába, hogy felvázolhassuk azon nyelvhasználati és stilisztikai relevanciákat, amelyek feltehetően jellemzik Jókai Mór írói ujjlenyomatát, a vizsgálati korpuszok létrehozása. Ezen feladat nem nélkülözheti a széleskörű és irányított mintavételt Jókai Mór életművéből, illetőleg kiemelt fontosságú a referenciaanyag összeállítása is, amely segítségével komparatív méréseket tudunk végezni, s ezek tükrében döntést hozni abban, hogy a 19. századi magyar prózairodalomra vonatkozóan mely paraméterbeállítások lesznek azok, amelyek valószínűsíthetően képesek eredményt felmutatni a szerzőazonosítás céljából elvégzett vizsgálatokban.

Az első kialakított vizsgálati korpuszunk a Jókai korpusz volt, amely összeállítása során törekedtünk arra, hogy Jókai Mór életművéből minél teljesebb gyűjtemény kerüljön bele a korpuszba. Az elsődleges forrása a Jókai Mór által írt regényeknek az Unikornis Kiadó Jókai sorozata volt.² Mivel ezen kiadássorozat nem tartalmazta Jókai Mór teljes életművét, ezért bővítenünk kellett a forrásainkat, a korpuszba kerülő többi regény forrása így a Révai Testvérek kiadó kiadványai lettek.³ Végül a két kiadássorozat alapján összeállított korpusz anyagát összevetettük a Magyar Országos Közös Katalógus (MOKKA)⁴ listájával (a katalógusban Jókai Mór névvel indítottunk keresést). Az így kapott lista alapján a korpuszunkat bővítettük azokkal a prózai munkákkal, amelyek a két kiadássorozatban nem szerepeltek, viszont szabadon hozzáférhetőek voltak interneten, tehát a Magyar Elektronikus Könyvtár gyűjteményéből vételeztük ezen szövegeket.⁵ Összesítve a Jókai korpuszunk 66 darab regényt, 6 darab novellát és 14 darab más prózakötetet tartalmazott.

2 *Jókai Mór munkái - gyűjteményes díszkiadás.* Unikornis Kiadó. Budapest. 1992–2004.

3 *Jókai Mór összes művei.* Révai Testvérek Irodalmi Intézet. Budapest. 1894–1898.

4 <http://www.mokka.hu/> (hozzáférés: 2021.05.24.)

5 <https://mek.oszk.hu/> (hozzáférés: 2021.05.24.)

A referencia korpusz összesen 132 regényt tartalmaz, a szövegek forrása a Magyar Elektronikus Könyvtár, ennek a korpusznak egy része az ELTE DH Regénykorpusz gyűjteményében is megtalálható.⁶ A referencia korpusz összeállításánál az volt a kiindulópontunk, hogy olyan szabadon elérhető regényeket keressünk, amelyek első kiadása az 1846 és 1907 közötti időszakra esik, mert ezen időszak párhuzamos a Jókai Mór életmű keletkezési idejével. A 132 regény összesen 56 szerzőtől származik. Mind a Jókai korpusz mind pedig a referencia korpusz szövegeinek tárolási- és munkaformátuma a TXT formátum.

Azonban önmagában ez a két korpusz nem bizonyult alkalmasnak a mérések beállítására azért, mert a referencia korpusz teljes anyagának jelentős része olyan szövegkiadásokból áll, amelyek a 19. századi helyesírási és szövegkiadási hagyományokat tükrözik, míg a Jókai korpuszba kerülő művek zöme (Unikornis Kiadás sorozata például) a helyesírást tekintve a mai sztenderdet követi, ez pedig a nyelvtisztítási mérések eredményeit s azok hitelességét alapvetően megkérdőjelezheti. Ezen hipotézis tesztelésére további alkorpuszokat hoztunk létre. Az első volt egy szűkített regénykorpusz, amely Jókai Mór három regényét tartalmazta, illetve négy szerzőtől további 12 regényt. Olyan szerzőket választottunk ebbe az alkorpuszba, akik életműve kanonikusként tartható számon a magyar irodalom- és szövegkiadás történetben, így feltételezhető volt az, hogy a MEK-en olyan szövegkiadásai is elérhetők, amelyek már a XX. század második felére datálhatók, így a helyesírásukat tekintve nem különböznek jelentős mértékben a Jókai Mór szövegek modernebb kiadványaitól. Az így kiválogatott szerzők és műveik a következők: Mikszáth Kálmán (*Beszterce ostroma*;⁷ *A beszélő köntös*;⁸ *Új Zrínyiász*⁹), Kemény Zsigmond (*Férj és nő*;¹⁰ *Gyulai Pál*;¹¹ *Az özvegy és leánya*¹²), Gárdonyi Géza (*Az isten rabjai*;¹³ *A világjáró angol*;¹⁴ *A lámpás*¹⁵), Eötvös József (*Magyarország 1514-ben*;¹⁶ *A karthauzi*;¹⁷ *A falu jegyzője*¹⁸). Jókai Mórtól pedig *Az arany ember*,¹⁹ *A lőcsei fehér asszony*²⁰ és a *Szomorú napok*²¹ című regény került a szűkített korpuszba.

A nagyobb korpusz, amelyben a vételezett szövegek kiadási évétől eltekintettünk, 69 szöveget tartalmazott. 3 Jókai Mór által írt művet és 66 másik regényt 32 szerzőtől a

6 <https://regenykorpusz.elte-dh.hu/> (hozzáférés: 2021.05.24.)

7 Mikszáth Kálmán, *Beszterce ostroma*, Arcanum, Budapest, 1998.

8 Mikszáth Kálmán, *A beszélő köntös*, Arcanum, Budapest, 1998.

9 Mikszáth Kálmán, *Új Zrínyiász*, Arcanum, Budapest, 1998.

10 Kemény Zsigmond, *Férj és nő*, Unikornis Kiadó, Budapest, 1996.

11 Kemény Zsigmond, *Gyulai Pál*, Szépirodalmi Könyvkiadó, Budapest, 1967.

12 Kemény Zsigmond, *Az özvegy és leánya*, Európa, Budapest, 2000.

13 Gárdonyi Géza, *Az isten rabjai*, Szépirodalmi Könyvkiadó, Budapest, 1974.

14 Gárdonyi Géza, *A világjáró angol*, Rozsnyai Kiadó, Budapest, 1905.

15 Gárdonyi Géza, *A lámpás*, Irod. és Műv. Kiadó, Bukarest, 1957.

16 Eötvös József, *Magyarország 1514-ben*, Szépirodalmi Könyvkiadó, Budapest, 1972.

17 Eötvös József, *A karthauzi*, Franklin társulat, Budapest, 1911.

18 Eötvös József, *A falu jegyzője*, Szépirodalmi Könyvkiadó, Budapest, 1971.

19 Jókai Mór, *Az arany ember*. Unikornis Kiadó, Budapest, 2001.

20 Jókai Mór, *A lőcsei fehér asszony*, Unikornis Kiadó, 2001.

21 Jókai Mór, *Szomorú napok*, Unikornis Kiadó, 2001.



referencia korpuszunkból. Az így létrehozott két alkorpusz lehetőséget biztosított arra, hogy teszteljük azt, hogy mi történik akkor a szerzőazonosításra alkalmazott kódok és paraméterbeállítások alkalmazásánál, amikor a vizsgált szövegek kiadásának évét (és ezen keresztül a helyesírási hagyományt) nem vesszük figyelembe.

A korpuszok létrehozásánál kiemelkedően fontos szemponttá vált a duplumok szűrése, hiszen például a Jókai Mór kötetek korpuszba való illesztésénél előfordult az, hogy ismétlődő szövegeket azonosítottunk. A célunk az volt a duplumszűrésnél tehát, hogy minden szöveg lehetőleg csak egyszer forduljon elő a vizsgálatra kialakított szöveggyűjteményeinkben.

Az ismétlődő szövegek felbukkanásának alapvetően két oka volt. Az első, hogy különböző novellák, elbeszélések nem egy, hanem akár több kötetnek is részét képezték, vagyis ilyenkor az egész novella, illetve elbeszélés egynél többször szerepelt a korpuszban. Ezekben az esetekben a duplumszűrés segítségével megkerestük az ismétlődő szövegeket, s csupán egyszer hagytuk meg őket a korpuszban. A másik lehetőség a szövegek ismétlődésére az intratextualitás, azaz a szerző szövegei közti kapcsolat volt, ugyanis jellemzően az útleírások és életrajzi írások tartalmaztak olyan szövegrészt, melyek egy-egy regényben, elbeszélésben is felbukkantak. Ezek ugyanakkor jórészt csupán egy-egy bekezdésnyi szöveget jelentettek, s olykor nem is teljesen pontosan voltak átvéve, némi módosítás történt rajtuk. Ez utóbbi esetekben tehát a korpuszban hagytuk az ismétlődő részeket, egyrészt, mert az intratextualitást nem hagyhatjuk figyelmen kívül teljes mértékben a szerzőségi vizsgálatban, másrészt, mivel a szűkített korpuszból kihagytuk az útleírásokat, valamint az önéletrajzi írásokat, tekintve, hogy a referenciakorpusz is jobbára csak fikciós műveket tartalmaz.

3. A kutatás módszere

A korpuszok összeállítása után a vizsgálatot a Stylo-tól ellépve, pythonban végeztük. A Stylo csomag egy R-ben - mely egy programozási nyelv és szoftverkörnyezet statisztikai számításokhoz és ábrázoláshoz - kialakított függvénykönyvtár, mely a stilometriai elemzések során gyakran használt funkciókat tartalmazza, beleértve a szerzői hozzárendelést és a műfaji felismerést.

A python egy általános célú, programozási nyelv, melyet az 1980-as évek végén, eredetileg oktatási célra alakítottak ki. Szövegelemzési célokra is kiválóan alkalmazható, így ma a digitális bölcsészetben gyakran használt programnyelv.

A Mike Kestemont által készített, ám 5 éve nem frissített, python programnyelvre átírt Stylo implementációt, a Pystyl-t, valamint egy teljes dokumentációjú kódot, melyben a szöveg és a végrehajtható kód váltja egymást, a Pydeltát felhasználva lefuttattunk a korpuszokon több száz beállítást, hogy lássuk, melyek a leghatékonyabbak a 19. századi magyar szövegek klaszterezésére, vagyis adattömbök homogén csoportokba sorolására, klasszifikálására.

A Stylo-tól való ellépés egy igen fontos lépés volt, méghozzá több szempontból. Először is azért, mert ezt megelőzően csupán manuálisan, a különböző paramétereket egyesével beállítva végeztünk hasonló vizsgálatokat a Stylo-ban.

Továbbá, bár a Stylo kódjaihoz is hozzá lehet férni, úgy véltük, az ellenőrizhetőség, a megismételhetőség és az átláthatóság kedvéért célszerű ezeket a kódokat pythonba implementálni, s ott végezni ezeket a vizsgálatokat. A nyílt tudomány elveinek mentén a futtatott kódokat, valamint a korpuszokat természetesen közzé tesszük.

Harmadrészt az indokolta a Stylo-tól való ellépést, hogy olyan beállításokat is kívántunk alkalmazni, melyeket a stylo nem tud. A Stylo-ban ugyanis nincs lehetőség az általunk vizsgált egyik távolságmérték, a Burrows delta használatára, továbbá a Feature-ek keverése sem megoldható.

Feature alatt azokat a megszámlálható jellemzőket értjük, melyeket automatikusan kivonunk a korpuszból, majd ezeket változóként használjuk a további elemzéshez. A stylo alapbeállításában ezek a korpuszokban leggyakrabban előforduló szövegelemek, azaz az MFW lista. Ezek lehetnek akár szó vagy karakter n-gramok. Az n-gram a számítógépes nyelvészetben olyan összefüggő/folyamatos szekvencia, amely n elemből áll (tehát például szó-bigram esetében ez azt jelenti, hogy 2 egymást követő lexéma).

Bizonyos esetekben azonban van értelme személyre szabott funkciók halmazát használni, pl. azokat a szavakat, amelyek érzelmekkel társulnak, vagy mondjuk a funkciószavak meghatározott részhalmazát, erre ugyanakkor a Styloban szintén nincs lehetőség.

Az előbbieken említett MFW rövidítés a „Most Frequent Words” szóhármast takarja, habár pontosabb lenne szavak helyett szövegelemről beszélni. Az MFW mint paraméter azért bír kiemelt jelentőséggel, mert nagymértékben hatással lehet az elemzés kimenetelére, és mindemellett a megfelelő érték megtalálása is összetett feladat, például függ a szövegek hosszától.

Szintén fontos paraméter a **Culling**. Ez egy százalékos érték. Az itt megadott szám jelentése, hogy az adott szónak/szövegelemnek a korpuszban található szövegek hány százalékában kell előfordulnia ahhoz, hogy felhasználásra kerüljön az elemzés során. Ez a paraméter, az MFW-hez hasonlóan nehezen megjósolható, habár a szövegek hossza és a szövegek szóhalmazai közti átfedés elég jól meghatározza az optimális mértéket.

Az előzetes, manuálisan végzett vizsgálatok eredményei alapján úgy tűnt, a 19. századi magyar szövegek esetén a legpontosabb eredményeket a Cosine (Koszinus) Távolság adja, szó-bigrammokon futtatva, 1000-es MFW listával, Culling és Sampling nélkül. Habár n-dimenziós koordináták közti távolságok kiszámítására rendkívül sokféle módszer létezik, az előzetes eredményekből, továbbá Christof Schöch egy korábbi vizsgálatából kiindulva három távolságmértéket választottunk ki a vizsgálatához.²²

22 Baayen et al., 2017

Lásd még a kutatás során felhasznált kódokhoz <https://github.com/christofs/stylometry-bibliography> (hozzáférés: 2021.05.27.)



Burrows Deltája nagy előrelépést jelentett az eddig használatos távolságmértékekhez képest, és mind a mai napig az egyik leggyakrabban használt távolságmérték. Gyakran használják más módszerekkel való összehasonlítás céljára.

$$\Delta_{(A,B)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - \mu_i}{\sigma_i} - \frac{B_i - \mu_i}{\sigma_i} \right|$$

1. ábra

Eder Deltája egy módosított verziója a Burrows Deltájának. A gyakoribb szavak súlyozását megnöveli valamelyest, a ritkébbakét pedig lecsökkenti.

$$\Delta_{(A,B)} = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{A_i - B_i}{\sigma_i} \right| \times \frac{n - n_i + 1}{n} \right)$$

2. ábra

A Koszinusz Távoóság is egy gyakran használt, jól ismert mértékegység. Két N dimenziójú vektor közötti hasonlóság mértékét jelenti a vektorok közötti szög koszinuszának megállapításával.

$$1 - \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}}$$

3. ábra

Ezekkel a távolságmértékekkel dolgoztunk tehát, és a lehető legtöbb beállítást futtattuk le, még hozzá egy kisebb és egy nagyobb korpuszon is, a már említett 15 valamint 69 szöveget tartalmazó korpuszokon. A következő fejezetben (4.) az ábrán különböző beállításokkal való futtatás eredményei láthatók.

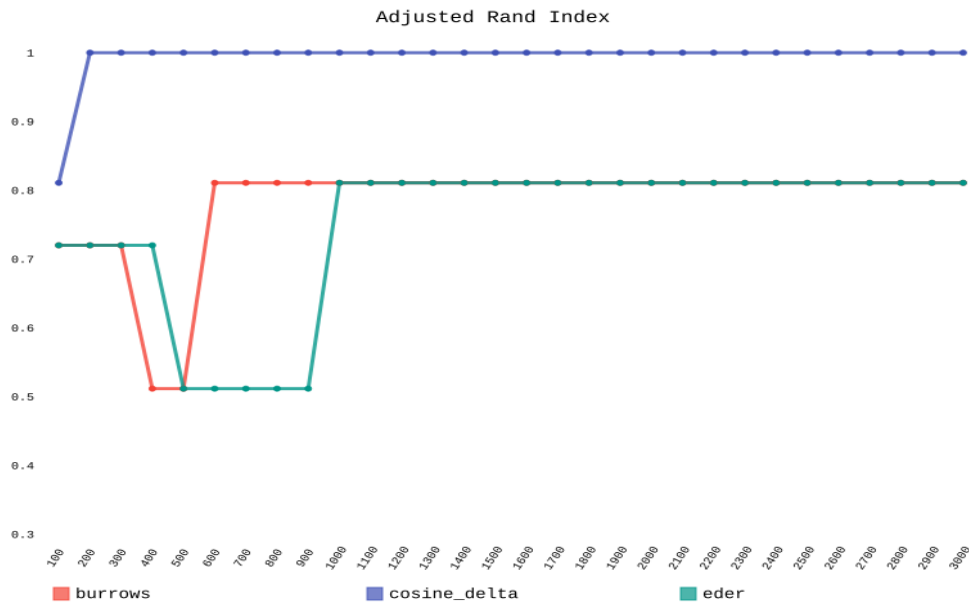
4. A vizsgálatban tesztelt paraméterbeállítások és függvények eredményei²³

Az ábrákon a függőleges tengely 0-tól 1-ig van beosztva, s az 1 jelenti a maximum értéket, vagyis azt, hogy a függvény maximális pontossággal tudta meghatározni a szövegek szerzőjét, és az összes művet hibátlanul csoportosította a szerzőség alapján.

A vízszintes tengelyen az MFW (Most Frequent Word) lista értékének növekedése látható, 100-as kezdőértékkel, egészen 3000-ig, 100-as lépésközzel. Ez azt jelenti, hogy minden egyes beállításról 30 különböző MFW listán való lefutás eredményét tudjuk megmondani. Az adott szám azt jelenti, hogy hány leggyakoribb szövegelemét tartalmazza a lista a korpusznak.

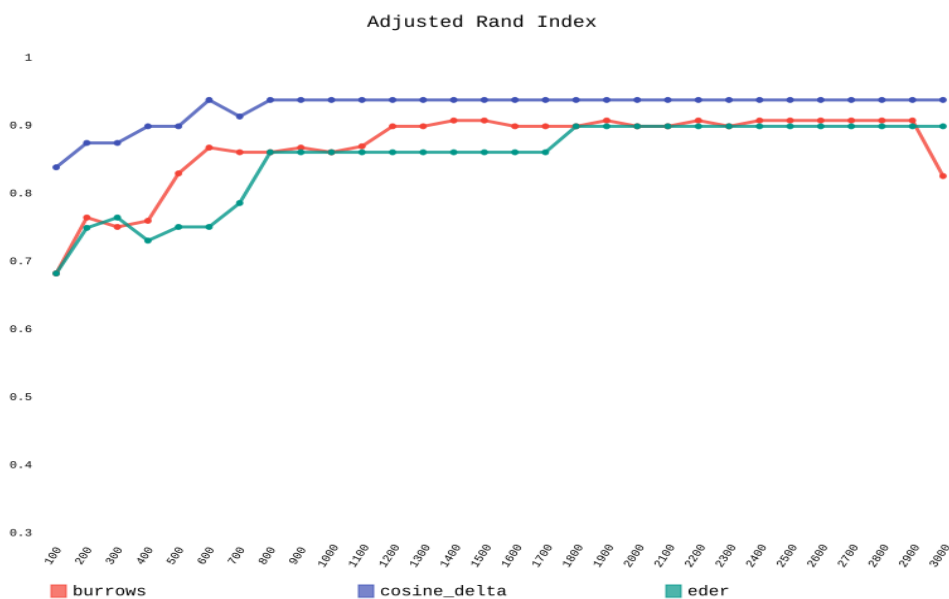
²³ A kutatáshoz alkalmazott kódok pythonba történő implementációját Nemeskey Dávid végezte el.

Ezen az ábrán egy olyan lefutás eredménye látható, mely a 15 szöveget tartalmazó korpuszon ment végbe, karakter-trigrammokon, 10%-os cullinggal.



4. ábra

Szembevetendő, hogy a Koszinusz Távolság már 200-as MFW listánál eléri maximumértékét (4. ábra), s ez után a vizsgált MFW intervallumon belül képes maximális pontossággal a szerzők csoportosítására, míg a másik két függvénynek 300, illetve 400 MFW-nél a hatékonysága csökken, s csak 600, illetve 1000 MFW esetén ível fel újra, azonban még ekkor sem érik el azt a teljesítményt, amit a Koszinusz.

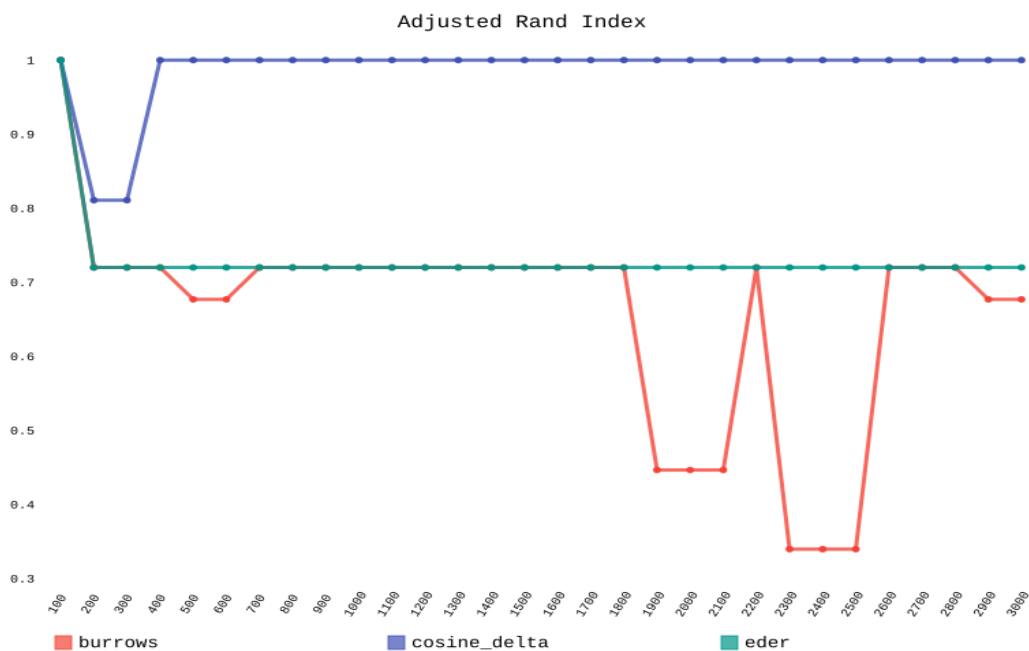


5. ábra

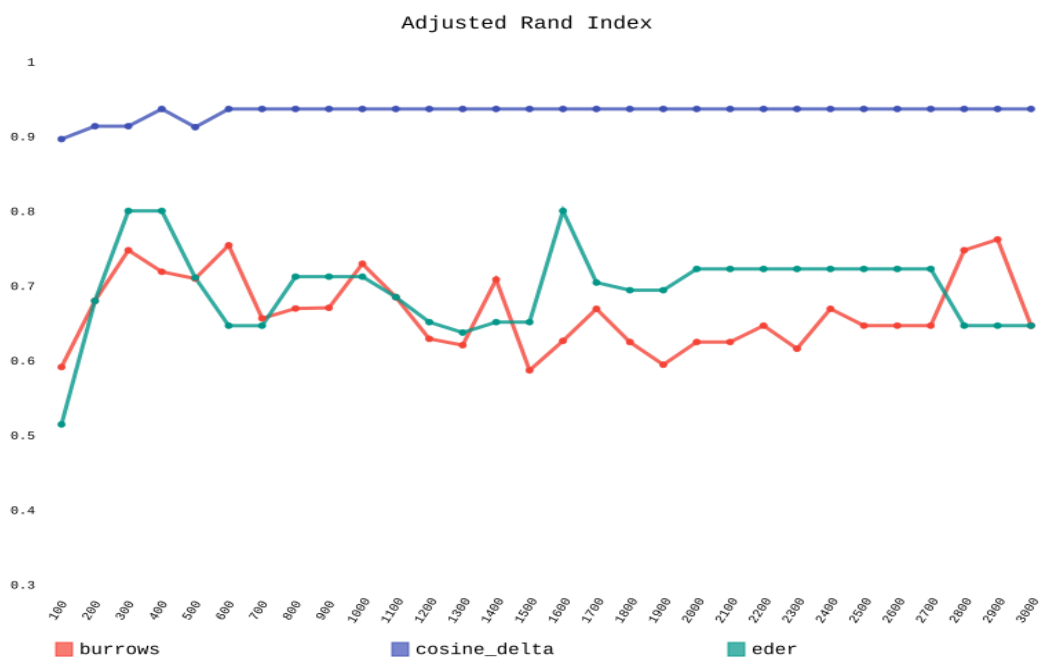


Az 5. ábrán ugyanezen paraméterek eredményei láthatóak a 69 szövegből álló korpuszon futtatva.

Megjegyzendő, hogy míg a 15 szövegen végzett vizsgálat során igen, addig a 69 szövegen végzett vizsgálat során még a Koszinusz Távolság sem érte el a maximum hatékonyságot, azaz nem tudta hibátlanul klaszterezni a szövegeket, azonban ahogy a többi ábrán is látszik, a másik két függvényhez képest így is kiemelkedő teljesítményt nyújtott.



6. ábra



7. ábra

A 6.-7. ábra a szó-bigrammokon, szintén 10%-os Cullinggal futtatott eredményeket mutatja mind a 15 szöveges, mind a 69 szöveget tartalmazó korpusz esetén. Látható az, hogy a szó-bigrammok esetén szintén a Koszínusz Távolság függvény az, amely az általunk alkalmazott paraméterbeállítások mentén a legmagasabb pontossággal képes klasszifikálni a vizsgált szöveganyagot.

5. Összefoglalás

Vizsgálatunk eredményeiből az látszik, hogy a Koszínusz Távolság bizonyult a leghatékonyabb távolságmérésnek a 19. századi magyar nyelvű szövegek szerzőazonosítási, illetve klaszterezési eljárásai során. A különböző beállítások közül a 3-as és 4-es karakter n-gramok, valamint az 1-es és 2-es szó n-gramok adták a legpontosabb eredményeket. Ez utóbbiakat azonban fenntartással kezeljük abból a megfontolásból, hogy a magyar nyelv esetén a szó n-gramok használata torzíthatja a vizsgálatot, mivel a regény tematikája helyeződik előtérbe. Ennek kiküszöbölésére szolgál majd az a további vizsgálatunk, mely során az e-magyar²⁴ emtsv verziójának segítségével a szövegek szemantikai tartalmú szavait (a főneveket, melléneveket, valamint az igéket) poszttag-ekre (morfológiai elemzést tartalmazó címkékre) cseréljük, s az így kapott szövegeken futtatjuk újra a leghatékonyabbnak bizonyult beállításokat. Így egy tematikafüggetlen módszerrel egészítenénk ki ezen kutatásunkat.

Szintén további kutatási terveink közé tartozik, hogy Jókai Mór szövegeinek nyelvezetéről egy szélesebb körű stilometriai elemzést készítsünk, mely a Jókaira jellemző nyelvstatisztikai alapú mintázatok feltérképezését jelenti, vagyis megkíséreljük megkonstruálni Jókai Mór szerzői „ujjlenyomatát”. Ehhez kapcsolódik az a szerzőséggel foglalkozó vizsgálat, amelyet a Jókai Kritikai Kiadást készítő kutatócsoporttal együttműködve végzünk. A vizsgálat során stilometriai módszerek alkalmazásával teszünk kísérletet arra, hogy a kutatók által kétes hitelűnek tekintett szövegekről eldöntsük, vajon a konstruált „ujjlenyomat” alapján a biztos szerzőségű Jókai szövegekhez mennyire állnak közel a nyelvstatisztika perspektívájából.

6. Források és bibliográfia

6.1. Források

A tanulmányban idézett képletekhez lásd:

<http://tdk.bme.hu/VIK/DownloadPaper/Szovegelemzesi-modszerek-automatizacioja>
(utolsó elérés: 2021.05.27.)

A kutatás során felhasznált kódok: <https://github.com/christofs/stylometry-bibliography>

Jókai Mór összes művei. Révai Testvérek Irodalmi Intézet. Budapest. 1894–1898.

Jókai Mór munkái – gyűjteményes díszkiadás. Unikornis Kiadó. Budapest. 1992–2004.

²⁴ Indig et al., 2019.



6.2. Bibliográfia

Harald Baayen, Hans van Halteren, Anneke Neijt, Fiona Tweddle, “An experiment in authorship attribution” Conference Paper (JADT 2002: 6es Journées internationales d’Analyse statistique des Données Textuelles) 2002.

Indig Balázs, Sass Bálint, Simon Eszter, Mittelholcz Iván, Kundráth Péter, Vadász Noémi, „emtsv – egy formátum mind felett”, In *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette Berend Gábor, Gosztolya Gábor, Vincze Veronika, 235–247. Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet. 2019.

Stefan Evert, Fotis Jannidis, Thomas Proisl, Isabella Reger, Steffen Pielström, Christof Schöch, Thorsten Vitt, “Understanding and explaining Delta measures for authorship attribution”, *Digital Scholarship in the Humanities*, 2017(2): ii4–ii16. hozzáférés: 2021.05.27. doi: [10.1093/llc/fqx023](https://doi.org/10.1093/llc/fqx023)

A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben

Mihály Eszter
Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ
mihaly.eszter@dbk.pim.hu

Közgyűjteményi feladatok kéziratok források esetén

A kéziratok forrásokat őrző közgyűjtemények hagyományos feladatait a digitális technológiák fejlődése és rohamos elterjedése miatt egyre sürgetőbb újragondolni, átformálni, kibővíteni. A hagyományos eljárások, mint a források leírása, metaadatolása, gyűjteményezése már nem látszik elegendőnek, illetve ezek módszere is korszerűsítésre szorul. Az archiválás, hosszútávú megőrzés, illetve a kereshetővé és hozzáférhetővé tétel szintén mást jelent digitális közegben, mint az analóg világban.

A források digitális másolatainak létrehozásának általános gyakorlatával az is kérdésessé vált, hogy a közzététel mely szintjét nevezhetjük már publikációnak. Ha egy kézirat forrás digitális másolatát közzétesszük, az publikációnak számít-e?¹

Mindenesetre ma már alapvető feladatnak tekinthetjük, hogy a közintézményekben őrzött kéziratok források valamilyen módon hozzáférhetőek legyenek anélkül, hogy a fizikai példányt a kezünkbe kéne vennünk. Ez viszont azt jelenti, hogy a szövegkiadás legalább is bizonyos szintjei a közgyűjtemények hatáskörébe kerültek. A GLAM szektor kéziratok forrásokat őrző szereplőinek ezen új feladatkör kihívásaival kell szembesülniük, és válaszokat adniuk.

A kéziratok források közzététele több szinten lehetséges:

1. A kéziratok digitális másolatának közzététele

Mivel az archiváláshoz is mindenképp szükséges egy jó minőségű digitális faksimile készítése, nem jelent sok többletmunkát az ún. master fájlból egy közzétételre alkalmas (kisebb felbontású, de jól használható) verzió létrehozása, amely automatikus konverzióval megoldható. A szkennelés elvégzésén túl csak az intézmény online katalógusfelületére való feltöltés elvégzése van szükség (és természetesen előzetesen az adott forrás gyűjteményezésére).

Így a gyűjteményi rekord adatai, tehát a kéziratok metaadatai kereshetővé, szűrhetővé válnak (az adott katalógusfelület lehetőségei szerint), a felhasználó pedig azonnal megkapja olvasásra a kézirat digitális másolatát.

Ez a szint az analóg világban a hasonló kiadásnak feleltethető meg, bár az online keresőfelület ehhez számos egyéb funkciót ad hozzá.

2. A kéziratok digitális másolatának és átiratainak közzététele

Az előző szinthez képest ez az eljárás egy lépéssel továbbmegy, nem csak digitális faksimilét készít, hanem a szövegek átirását is elvégzi. Ez természetesen idő-

1 Erre a kérdésre itt nincs lehetőség részletesebben kitérni, a téma külön tanulmány tárgyát képezi.



és erőforrásigényes munka, amelyre nem biztos, hogy mindenhol van kapacitás. A jövőben viszont a nyomtatott szövegek automatikus gépi felismertetéséhez (OCR) hasonlóan egyre több lehetőség nyílik a kéziratos dokumentumok gépi szövegfelismertetésére is (ld. később), amely nagyban csökkenteni fogja a digitális szövegek létrehozásához szükséges erőforrásigényeket.

Az eredmény pedig nagy kilátásokat rejt magában, hiszen a kéziratok teljes szövegállománya kereshetővé válik, nem csak a metaadataik, amely a kutatók, felhasználók előtt hihetetlen távlatokat nyit meg, illetve létrejön az alapvető nyersanyag egy digitális forráskiadáshoz. (Ezen a szinten készülhet betűhű, betű szerinti, illetve normalizált átírat is a projekt céljainak megfelelően.)

Az átírt szöveg és a digitális fakszimile együttes megjelenítésére célszerű formátum a kétrétegű PDF, ahol a felhasználó az eredeti kézirat fotóját látva kereshet a kép mögött rejtve tárolt szövegátiratokban. Ehhez szükséges a szkennelt kéziratképek ún. szegmentálása is, amelynek során zónákra osztjuk a képet, s a zónák koordinátáinak használatával lesz összekötve a kép az átírt szöveggel.

A kétrétegű PDF-ek szintén feltölthetőek az adott intézmény meglévő online katalógusfelületére az előző pontban leírtak szerint, s itt kell lehetőséget nyújtani az ún. teljes szövegű, tehát a gyűjteményi rekordokhoz kapcsolt PDF-ek tartalmában való keresésre².

3. Digitális forráskiadás

Ezen a szinten hozhatunk létre egy textológiai-filológiai szempontból forráskiadásnak tekinthető publikációt. Az előző pontban született nyersanyagon tovább munkálkodva, vagy azt kihagyva és újonnan létrehozva születhet meg a kézirat annotált, adatgazdagított, részletesen metaadatolt kiadása. Erre a célra alkalmas a TEI XML³ formátum, amely nemzetközi szinten is a szövegek annotálására leginkább elfogadott és elterjedt jelölőnyelv. Az ún. címkék (tagok) és tulajdonságaik (attribútumok) használatával lehetőség nyílik szintaktikai (pl. bekezdés) és szemantikai (pl. név) szövegegységek beazonosítására, megjelölésére, ennél fogva lekérdezhetővé, kereshetővé és különböző módokon megjeleníthetővé válnak. A TEI fejlécben (TEI Header) részletes metaadatokat rögzíthetünk, illetve szerkesztői megjegyzésekkel, külső adattárakra való hivatkozásokkal gazdagíthatjuk a szövegközlést. A TEI XML publikációjára külön platform kialakítása szükséges⁴, amely kiaknázza az annotált, részletesen feldolgozott szövegekben rejlő megjelenítési, keresési és egyéb lehetőségeket.

2 Természetesen lehet átírást készíteni egyéb formátumokban is, pl. sima szöveggént (txt), Word-ben, de figyelembe kell vennünk, melyek a kitűzött céljaink, szeretnénk-e a digitális fakszimilét összekapcsolni az átírt szöveggel, illetve kívánunk-e bizonyos alapvető textológiai jelenségeket már ezen a szinten is jelölni, annotálni (pl. törlések).

3 Text Encoding Initiative: <https://tei-c.org/>

4 Illetve megoldást nyújthat egy közösségi platform használata. Többek között erre a célra tervezzük a Digital Humanities Platform (dHUpla) felületét.

A digitális forráskiadás textológiai-filológiai munkájának mélysége természetesen nagyon különböző mértékű lehet, a projekt céljainak megfelelően kell meghatározni, sőt a feldolgozást további egymásra épülő fázisokra lehet bontani (pl. az alapvető textológiai annotációk elhelyezésétől a tulajdonnevek beazonosításáig). Így már az első fázis elvégzésével a szövegek publikálhatóvá válnak, és a további feldolgozás során csak frissíteni kell őket.

A forráskiadás alapját képezheti egy későbbi kritikai (tudományos⁵) kiadásnak is, amely már nem feltétlenül a forráskiadást készítő intézmény feladatkörébe tartozik, de adott feltételek mellett természetesen ott is megoldható.

Minden GLAM szektorhoz tartozó intézménynek fel kell mérnie a saját lehetőségeit, kapacitását, és ahhoz mérten választani a fenti lehetőségek közül.

Miért jó a digitális szövegkiadás?

A digitális szövegkiadásoknak tulajdonképpen ugyanazok az elsődleges előnyei, amelyek a hátrányai is. Rugalmasak, ennélfogva állandóan változnak, így folyamatosan javíthatóak is. Esetükben nincs terjedelemhatár, bármeddig bővíthetőek, tehát nem szükséges válogatásokat eszközölni - hacsak nem humánerőforrás-hiány miatt.

Abszolút előnyük viszont az analóg kiadásokhoz képest, hogy annotálhatóak, összeköthetőek egyéb tudástárakkal (pl. névterek, bibliográfiák, tezasaurusok), több formában és folyamatosan publikálhatóak, verziózhatóak, illetve a legkülönbélebb műveletek végezhetőek bennük (pl. keresés, szűrés, adatvizualizáció). Egészen új módszerekkel válnak kutathatóvá a szöveges tartalmak (ld. "distant reading"⁶), illetve az eredeti források archiválására is alkalmasak.

A digitális szövegkiadások ettől függetlenül nem kell, hogy kizárják a nyomtatott szövegkiadások lehetőségét, sőt. A tendenciák inkább azt mutatják, hogy nagyon is tudnak egymás mellett létezni, más-más felhasználói igényeket kielégítve. A jövő afelé halad, hogy a digitális szövegkiadások lehetnek a nyomtatott kiadások alapjai, s nem fordítva.

Új feladatok, problémák

Milyen új feladatokkal, problémákkal kell megküzdenie egy közgyűjteménynek, ha – valamilyen formában – vállalja a kéziratos forrásai közlését?

Először is össze kell hangolnia a digitalizálást végző műhely, a kéziratoskat őrző gyűjteményi tár munkáját, illetve a digitális bölcsészeti feladatokat. Minden esetben meg kell találnia a közös nevezőt, amely az összes fél igényeinek megfelel. Ez korántsem

5 A kritikai kiadás, illetve tudományos kiadás (digital scholarly edition) terminológiai kérdései szintén egy külön tanulmány tárgyát képezik.

6 https://en.wikipedia.org/wiki/Distant_reading



egyszerű, már csak a meglévő eszközök, hosszú idők során kialakult hagyományok és protokollok fényében sem⁷.

Alapvetően a közgyűjteményi és a digitális bölcsészeti szempontok egyeztetéséről van szó, amely elengedhetetlen a közös munka elvégzéséhez. Ha egy közgyűjtemény kéziratos források közzétételére vállalkozik, elkerülhetetlenül beengedi a digitális bölcsészeti feladatokat az intézmény hatáskörébe, és integrálnia kell azokat.

Ugyanígy, bizonyos digitális bölcsészeti eszközöket is be kell építenie a meglévő közgyűjteményi infrastruktúrába, illetve adott esetben visszafelé, egyes digitális bölcsészeti eszközökbe integrálhatja a már meglévő közgyűjteményi infrastruktúra elemeit (ld. később).

Kritikus probléma a felmerülő többletfeladatokhoz szükséges humánerőforrás biztosítása (pl. szövegek átírása). Amennyiben van rá lehetőség, mindenképpen kézenfekvő és célravezető a kéziratos anyagok őrző gyűjteményi tár dolgozóinak betanítása, és bevonása a digitális bölcsészeti feladatokba. Ezenkívül feltétlenül szükséges az intézmény különböző területein dolgozó informatikusokkal való folyamatos együttműködés is.

Pilot projekt

A fentiekben felvázoltak miatt a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában működő Digitális Bölcsészeti munkacsoporttal elindítottunk egy pilot projektet, ahol igyekeztünk a felmerülő kérdésekre választ adni, módszereket, protokollokat, munkafolyamatokat kialakítani, megfelelő eszközöket találni, a digitális forráskiadás architektúrájához teszt-modulokat fejleszteni. Az alábbiakban ezeket fogom vázlatosan bemutatni.

A pilot projekt Kiss József⁸ levelezésének digitális forráskiadása.

-
- 7 Jó példa levelezéskiadás esetén a digitális objektum egységének meghatározása. A PIM Kézirattár a gyűjteményezés során általában teljes levelezési irányokat ír le, amely egy személy másik személynek írott összes levelének metaadatait tartalmazza. A szövegfeldolgozás során az alapegységnek viszont nyilvánvalóan a levélnek kell lennie. Össze kellett tehát hangolni a szkenneléstől a kiadásig, mikor mi legyen az alapegység, a létrejött digitális objektumokat hogyan kapcsoljuk össze a gyűjteményi rekordokkal, és hogyan jelenítsük meg azokat.
- 8 [Kiss József \(1843-1921\)](#), A Hét című hetilap alapítója „olyan szerkesztői tehetség, akit csak az egy Osvát Ernő múlt felül. Lapja rövidesen orgánuma lett mindazoknak, akikben az új polgári szellem irodalmi formát öltött, de hordozta a régebbi nemzedékek kritikusabb elméit is.” – írja Szerb Antal [Szerb Antal: *A magyar irodalom története*. Budapest, Magvető, 1972 (ötödik kiadás), 420. o.]. A Hét című folyóirat a Nyugat elődjének tekinthető, amelyre az irodalomtörténeti kutatások eddig méltatlanul kis figyelmet fordítottak. Kiss József szerkesztőségi és egyéb levelezésének közzététele eddig feltáratlan közeget tesz hozzáférhetővé mind a szakmai közönség, mind az olvasók számára.

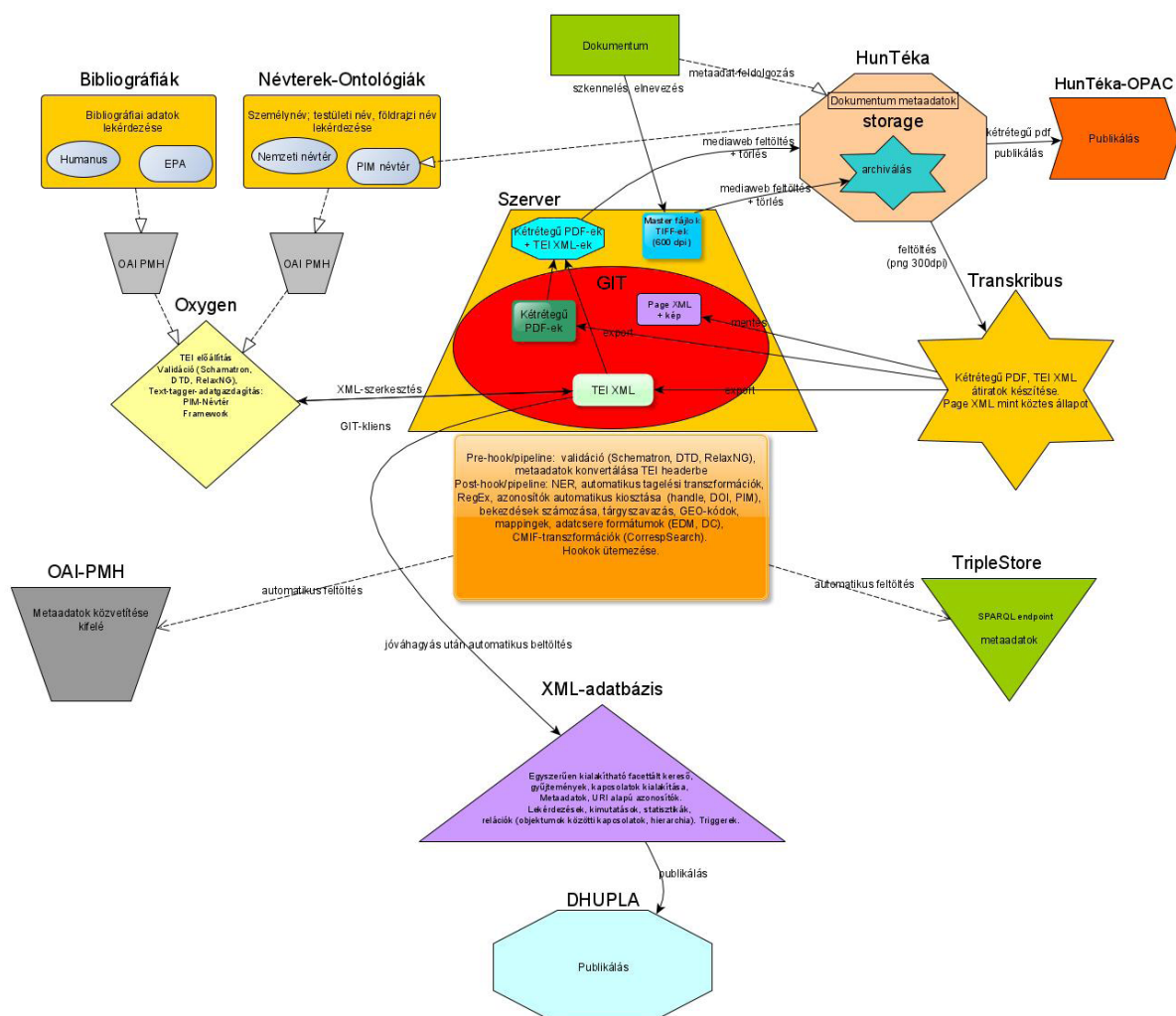
Első lépések

Megtettük a szükséges első lépéseket:

- szkennelés szabályainak kialakítása
- névkonvenció meghatározása (fájlok elnevezése, strukturálása)
- nyilvántartások, kimutatások készítése a projekt forrásairól
- ütemezés
- content management környezet kialakítása
- eszközök kiválasztása
- szerepek kiosztása
- workflow megtervezése

Workflow, infrastruktúra

A kialakított infrastruktúra, s ez alapján a munkafolyamat egyes lépéseinek meghatározása a következő ábrán látható:



A szkennelés-szövegátírás-publikálás egyszerű hármasa helyett láthatóan sokkal összetettebb rendszerre volt szükség, ugyanakkor mégsem olyan bonyolult, mint amilyenek első ránézésre tűnik. A középpontban egy verziókövető rendszer áll,

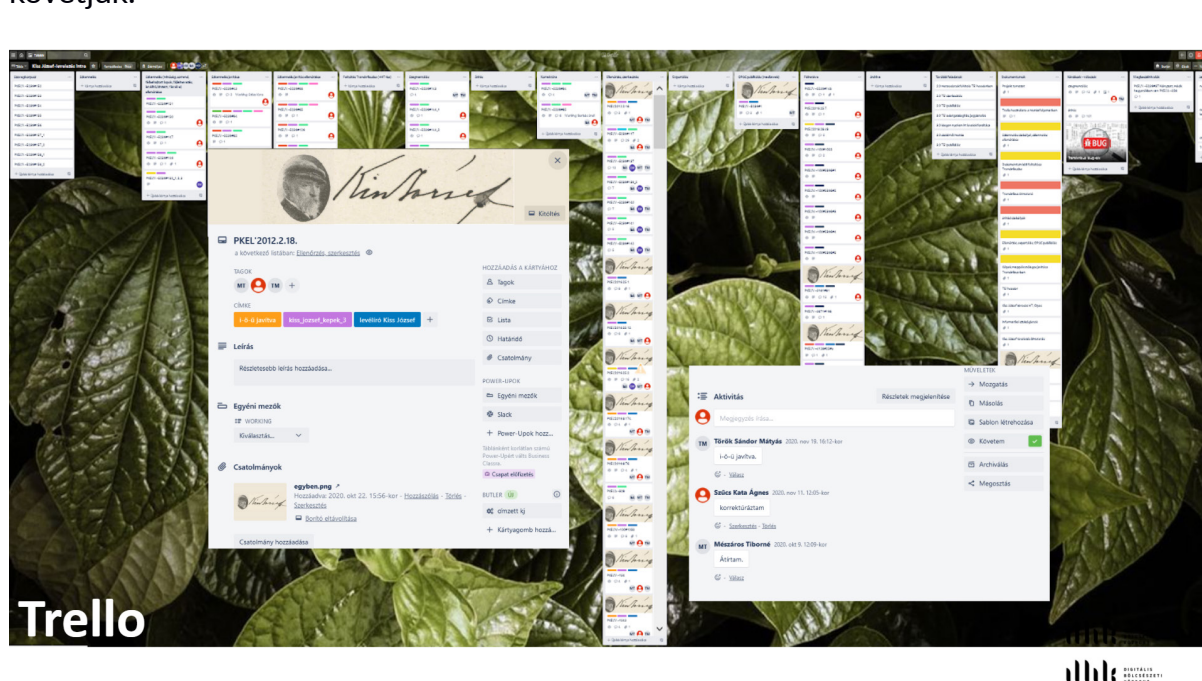


amelyben a tárolt anyagok minden változása visszakövethető, s innen indul ki minden automatizált folyamat, ezzel áll kapcsolatban közvetlenül vagy közvetve az összes használt eszköz.

Kétszintű szövegzóklést tűztünk ki célul, előbb kétrétegű PDF-ek formájában a PIM Opac felületén⁹, majd TEI XML-ek formájában egy új, erre a célra fejlesztett platformon¹⁰, amely az első fejezet 2-es és 3-as pontjának felelnek meg. (A TEI XML-ek publikációja több fázisra van osztva a feldolgozás mélységének megfelelően, hogy minél előbb publikussá válhassanak a szövegek. A névterekkel, bibliográfiákkal való összekapcsolás például csak a harmadik fázisban történik.) Beépítettük a rendszerbe a metaadatok automatikus továbbításának lehetőségeit, illetve a különböző lekérdezési módszerekhez szükséges eszközöket is.

Content management

Az egyes dokumentumok munkafolyamatban történő haladását, a Trello¹¹ felületén követjük.



Az ún. Kanban-módszer¹² testreszabása eredményeképpen minden oszlop a workflow egy állomásának felel meg, s minden dokumentumnak van egy külön kártyája. A kártya tartalmazza az összes szükséges információt: ki dolgozott már a dokumentummal, jelenleg épp ki dolgozik rajta, illetve minden egyéb a projekt szempontjából szükséges metaadatot, amelyek szűrhetőek is. Ezenkívül hozzászólások formájában lehet megjegyzéseket írni, kérdéseket megvitatni az adott dokumentummal kapcsolatban.

9 <https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/>"Kiss József 1843-1921"

10 dHUpa: Digital Humanities Platform - hamarosan elérhető a demo a www.dhupla.hu oldalon

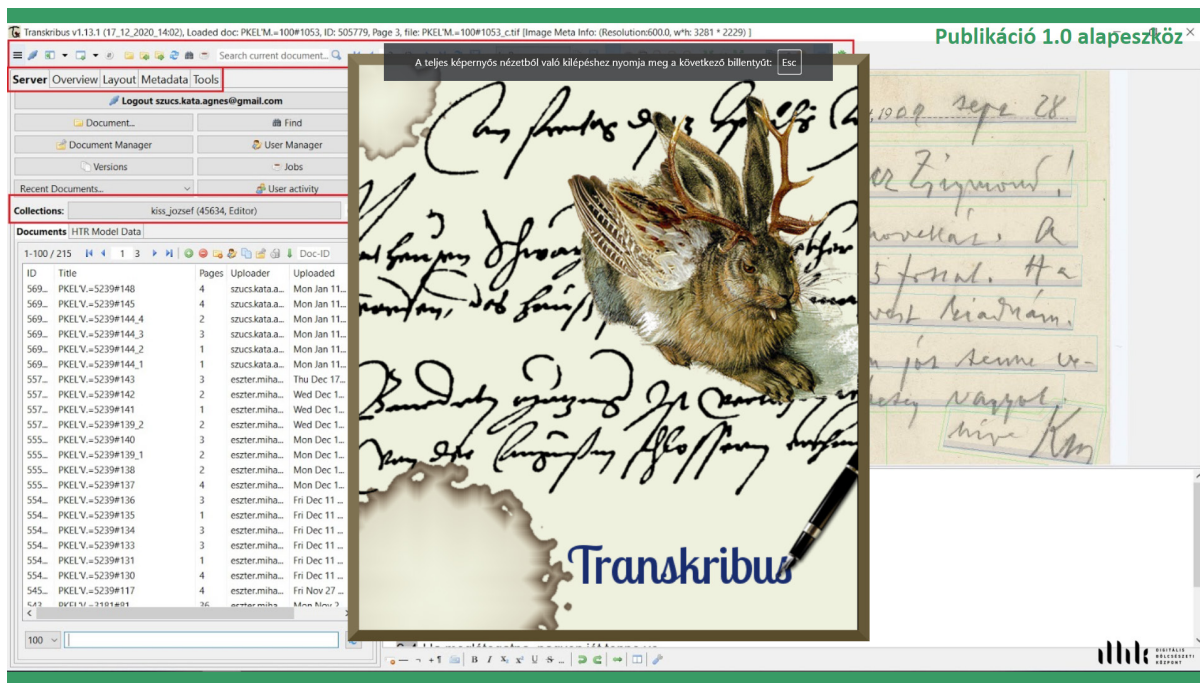
11 <https://trello.com/>

12 <https://en.wikipedia.org/wiki/Kanban>

Publikáció 1.0 - kétrétegű PDF

Szerkesztő program

A kétrétegű PDF fájlok előállításához a Transkribus¹³ programot használjuk.



A programot kifejezetten kéziratos dokumentumok átírására fejlesztették, felhasználóbarát felülettel rendelkeznek, és nem mellesleg automatikus kézirásfelismertető, ún. HTR (Handwritten Text Recognition) modellek létrehozására is alkalmas. Az átírt dokumentumok többféle formátumban exportálhatók, mint például a PDF és a TEI XML.

A programmal lehetséges a saját gépen való szerkesztés is, de az összes funkció kihasználásához a Transkribus szerverére is fel lehet tölteni a dokumentumokat, amely szintén verziókezelővel biztosítja a tárolásukat.

Eszközök integrációja

A számos felhasználói funkció teljes igénybevétele miatt a projekt során a szerveren dolgozunk, ezért egyik modulfejlesztésként a Transkribus-gyűjtemények automatikus napi mentésének megoldását tűztük ki célul, amely mára meg is valósult: a PIM saját szerverének verziókezelő rendszerébe naponta mentés készül a Transkribus szerveréről.

A Transkribus programot integráltuk a meglévő közgyűjteményi infrastruktúrába is: az ún. Mediaweb eszközbe, amely eddig a gyűjteményi rekordokhoz tartozó médiatartalmak feltöltésére szolgált, beépítettük a Transkribus-ba való feltöltés funkcióját is.

¹³ <https://readcoop.eu/transkribus/?sc=Transkribus>



FELTÖLTÉS

Quito Média Modul

Tár: PIM - Kéziratanyag (Petőfi Irodalmi Múzeum) Csak a megadott Tár-ban keressen kapcsolatot

Hozzáférés: Belső Csak a belső felhasználók érik el.

Publikus kapcsolat: igen nem

Fájlnév leképezés: PIM mapping PIM mapping_DESCRIPTION

Média típus: Automatikus felismerés

Importálási stratégia mind egyezők kézi Egyező elemek importálása.

DBK digitalizálás:

Feltöltés a transzkribus rendszerbe:

Transzkribus gyűjtemény:

Archiválás

Archív könyvtár: Dátum:

Ezzel jelentősen leegyszerűsödött a munkafolyamat kezdeti szakasza, tudniillik az archiváláshoz előállított szkennelt kéziratfotók gyűjteménykezelő-rendszerbe (Huntéka) való feltöltésekor a digitális fakszimilék egy gombnyomással feltölthetők a Transzkribusba is (a rendszer automatikusan konvertálja azokat a megfelelő felbontásra és formátumba), s azonnal megkezdődhet az átírásuk.

Publikációs, szűrési, keresési, rendezési lehetőségek

A kétrétegű PDF-ek publikálásával viszonylag gyors munkafolyamat valósult meg, az Opac-on kapcsolt médiarekordként való közzététel szinte semmilyen fejlesztést nem igényelt. (A PDF-ek megjelenítésével és a találati listákkal kapcsolatban merültek fel kisebb testreszabás-jellegű feladatok.) Bebizonyosodott tehát, hogy ebben a formátumban jól megoldható a szövegek integrációja a közgyűjteményi rendszerbe. A PIM Opac-ján külön URL-en elérhető a teljes Kiss József-levelezés¹⁴, és a levelezés azon része, amely már PDF formátumban is hozzáférhető¹⁵.

Az online katalógusfelületen eddig is lehetőség nyílt teljes szövegű keresésre, illetve metaadatok szerinti szűrésre is. Az ún. facetek segítségével (ld. alábbi képen bal oldalt) szűkíthetjük a találatokat az egyes adatmezők szerint, és a találati listánkat különböző szempontok szerint rendezhetjük is.

14 [https://resolver.pim.hu/gyujtemeny/levelek/"Kiss József 1843-1921"](https://resolver.pim.hu/gyujtemeny/levelek/)

15 [https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/"Kiss József 1843-1921"](https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/)

Mihály Eszter: A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben

ÚJ KERESÉS SÚGÓ KIÁLLÍTÁS ESEMÉNY MÚZEUM TANULÁS

PETŐFI IRODALMI MÚZEUM

Gyűjtemények Névtér Díjak Adattárak Böngészés

Műtárgytípus
Levél (113)

Alkotó/Közreműködő
Kiss József (51)
Ady Endre (1)
Apponyi Albert (1)

Anyag
papír (4)

Keletkezés/Megjelenés helye
Budapest (39)
H. n. (26)
Pest (8)

Levél címzettje
Kiss József (62)
Finicky Mihály (8)
Ismeretlen (4)

Copyright © Monguz Kft 2017 Powered by Quito Monguz

A teljes szövegű keresés esetén láthatjuk a rekordokon belüli találatokat:

ÚJ KERESÉS SÚGÓ KIÁLLÍTÁS ESEMÉNY MÚZEUM TANULÁS

PETŐFI IRODALMI MÚZEUM

Gyűjtemények Névtér Díjak Adattárak Böngészés

4 találat

Szűrők elrejtése

Műveletek

Rendezés: Alkotó: növekvő Nézet: Listás Találatok oldalanként: 24

Keresési előzmények
4. ((alkotó/közrem))
3. (alkotó/közrem)
2. (alkotó/közrem)
1. Teljes szöveg = r

Műtárgytípus
Levél (4)

Alkotó/Közreműködő
Ady Endre (1)
Fedák Sári (1)
Gara Illés (1)

Keletkezés/Megjelenés helye
Budapest (1)
H. n. (1)

1. Ady Endre (1877-1919)
Címzett: Kiss József (1843-1921)
Levél
Budapest, [1907 nov. 22.]
Petőfi Irodalmi Múzeum

Leltári példány (1) Média (1) Részletes...

BÁTYÁM: 1 TALÁLAT A TELJES SZÖVEGBEN

1. ...Mélyen Tisztelt Szerkesztő Ur, kedves Józsi bátyám, itt küldöm a cikket, a címet tessék föl

2. Fedák Sári (1879-1955)
Címzett: Kiss József (1843-1921)
Levél
H. n., Budapest, 1904. márc. 27-1906. jan. 23. és k. n.

Copyright © Monguz Kft 2017 Powered by Quito Monguz



Egy adott rekordra kattintva csak az azon belüli találatokat látjuk:

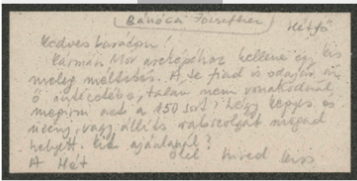
ÚJ KERESÉS SÚGÓ KIÁLLÍTÁS ESEMÉNY MÚZEUM TANULÁS

PETŐFI IRODALMI MÚZEUM
Gyűjtemények Névtér Díjak Adattárak Böngészés

https://resolver.pim.hu/bib/PIM1225902 Letöltés

Címkés MARCXML LIDO

Megnyitás



Levélíró: Kiss József (1843-1921)
Címzett: Bánóczi József (1849-1926)
Terjedelem: 1 f
Megjegyzés: Az eredeti levél ceruzairásos másolata?
Nyelv: Magyar
Dokumentumtípus: Levél

Múzeum	Gyűjtemény	Hagyatékneve	Raktári jelzet	Leltári szám
PIM	PIM - Kéziratanyag	Kiss József-hagyaték	V. 5239/1	V. 5239/1

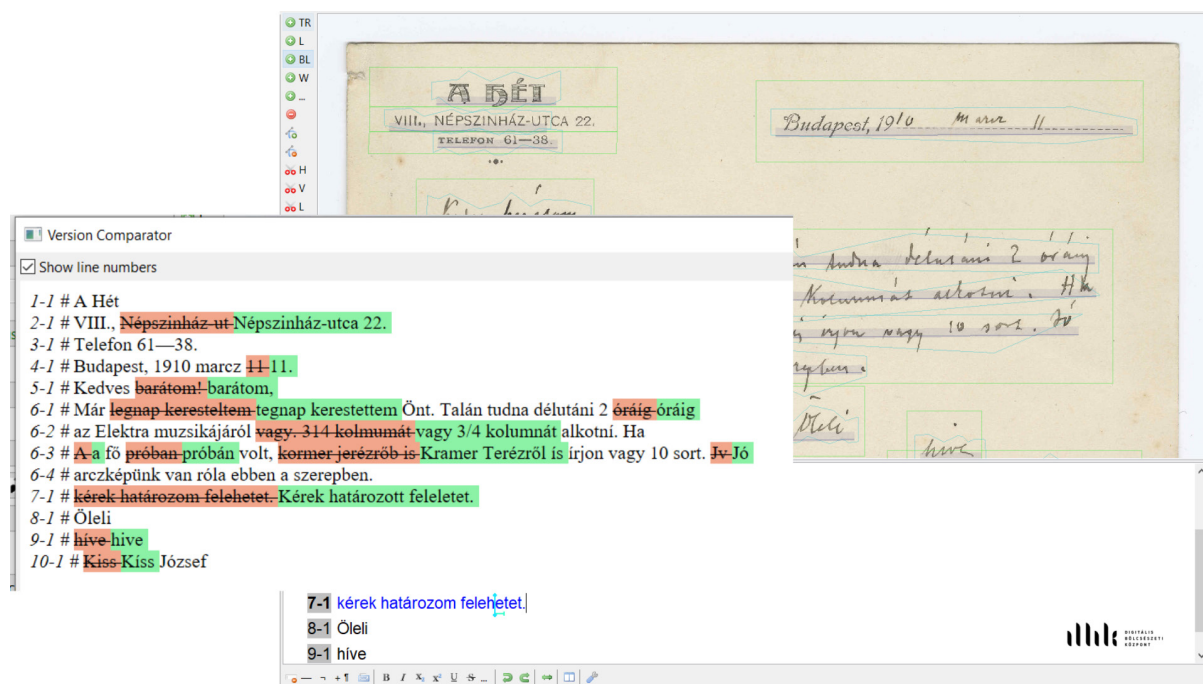
Találatok a teljes szövegben (3 találat):

1. oldal - állíts rabszolgát magad helyett. Kit ajánland? **Ölel híved Kiss A Hét**

A PDF megnyitásakor pedig magán a kézirattfotón is ki vannak emelve a találatok:

Automatikus kézírásfelismertetés

A Transkribus program bizonyos mennyiségű kézirat manuális átírása után lehetőséget nyújt automatikus kézírásfelismertetésre használandó HTR-modellek építésére¹⁶. Ennek mikéntjéről a kötetben külön tanulmányunk számol be, itt csak az eredményeket villantom fel. A Kiss József kézírásából épített modell segítségével 6,94%-os hibaszázalékkal ismertettük fel automatikusan a még át nem írt leveleket. A hibák többnyire a kis- és nagybetűk, rövid-hosszú ékezetek, illetve a pontuáció hibás felismeréséből fakadtak.



Ez azt jelenti, hogy az automatikus gépi kézírásfelismertetés után minimális kézi javítás szükséges egy átírat elkészítéséhez, amely a jövőben rendkívüli mértékben felgyorsítja a szövegfeldolgozást.¹⁷

Publikáció 2.0 - TEI XML

Szerkesztő program

A TEI XML-ek szerkesztéséhez az Oxygen XML Editor¹⁸ programot használjuk. Ezen belül ki lehet alakítani egy ún. frameworköt, amely az adott projekt(ek) igényeire van szabva, annak minden szempontját figyelembe véve. A szerkesztők ebben a nézetben gombok használatával tudnak annotálni, validálni, különféle automatizált műveleteket elvégezni. A framework kiterjedt textológiai-filológiai jelölésrendszer alkalmazására

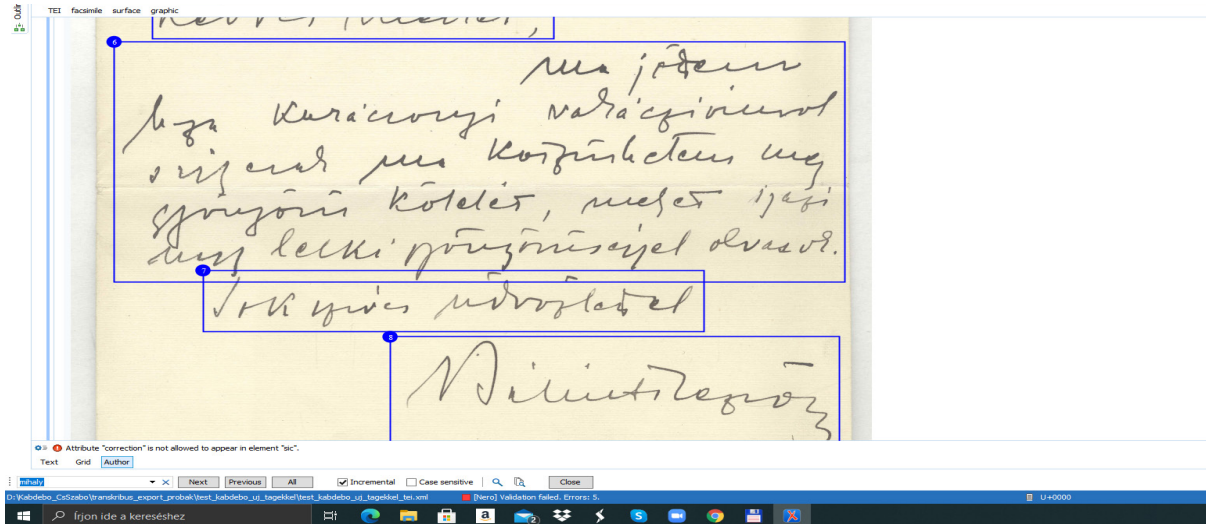
¹⁶ <https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus/>

¹⁷ A HTR modellek futtatása a Transkribus programban már fizetős funkció, amit bele kell kalkulálni a projekt költségvetésébe. Ennek fejében viszont készen kapunk egy eszközt az automatikus kézírásfelismertetés eléréséhez, a betanító szövegek előállításán túl csak a modell paraméterezését kell elvégeznünk. Az egyes modellek egymásba is építhetők, így egyre általánosabb érvényű HTR-modell hozható létre az adott nyelven.

¹⁸ https://www.oxygenxml.com/xml_editor.html

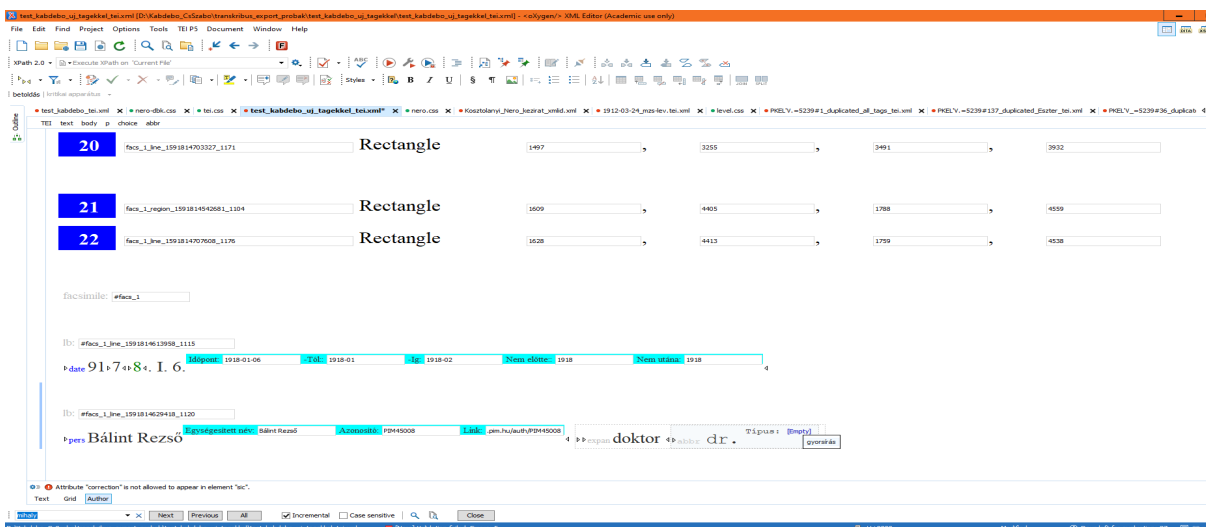


nyújt lehetőséget anélkül, hogy a szerkesztőnek a technológiai háttérrel ismernie kéne. Az adatgazdagítás, metaadatok részletes kitöltése, külső adatbázisok bekapcsolása, legkülönbözőbb annotációk alkalmazása, a text-image linking (kép és szöveg összekötése) mind felhasználóbarát módon megoldható ebben a környezetben.



A szerkesztőfelületen a szöveg és a benne lévő annotációk megjelenítését saját igényeink szerint alakíthatjuk, akár úgy is, hogy a szerkesztő a majdani publikált formát (is) láthassa. Meghatározhatunk template-fájlokat, amelyek egy új dokumentum készítéséhez szükséges alapvető elemeket tartalmazzák, ún. sémafájlokat (DTD, RELAX NG, Schematron stb.) alkalmazhatunk, amelyek a nemzetközi szabvány mellett saját szabályrendszerünk szerint is ellenőrzik a dokumentumokat, beépíthetünk automatikus műveleteket (pl. azonosító kiosztás, metaadatok beemelése, illetve továbbítása stb.).

Az Oxygen rendelkezik ún. Git-klienssel is, amelynek segítségével a szerkesztők közvetlenül a verziókezelő rendszerbe dolgozhatnak, illetve megoldható a szerkesztőségi környezetből való közvetlen publikáció is.



Eszközök integrációja

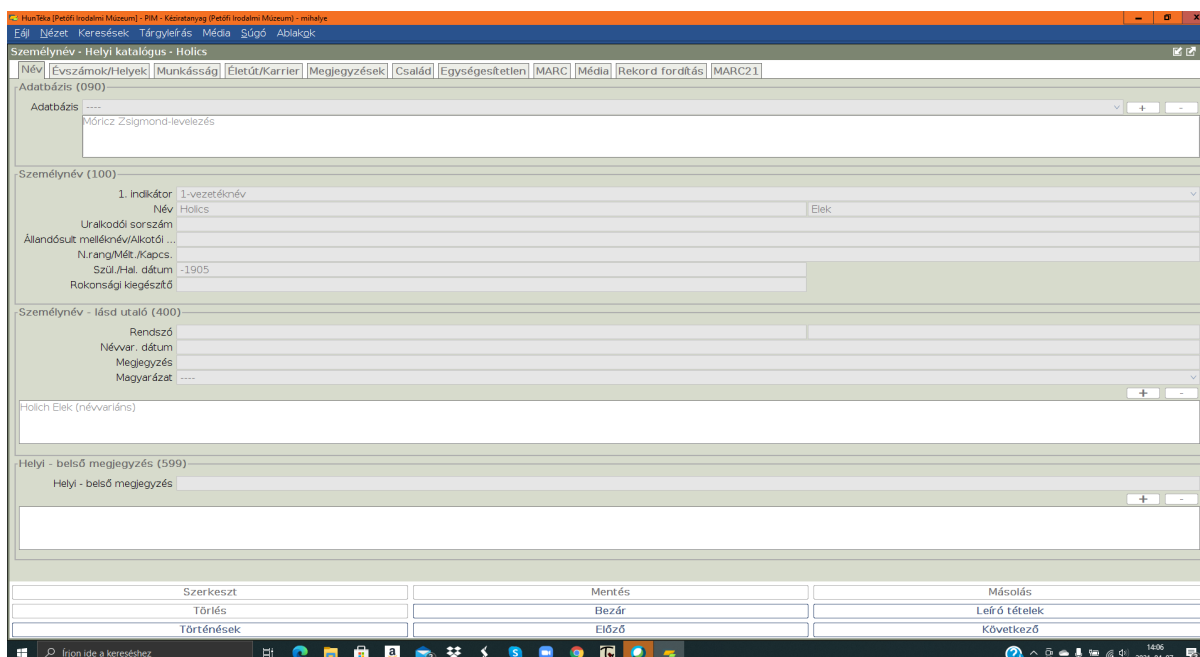
Ezen a területen is több ponton összekötöttük a közgyűjteményi infrastruktúra egyes elemeit a szövegtárolás szerkesztőjei környezetével.

Egyfelől kialakítottuk a PIM Névterével való együttműködést, amelyhez a Huntéka (gyűjteménykezelő-rendszer) bizonyos fejlesztéseire volt szükség, másfelől beépítettünk az Oxygen-frameworkbe egy funkciót, amelynek segítségével a szerkesztők közvetlenül a PIM Névterében kereshetnek, illetve azon keresztül azonosíthatják a különböző entitásokat.

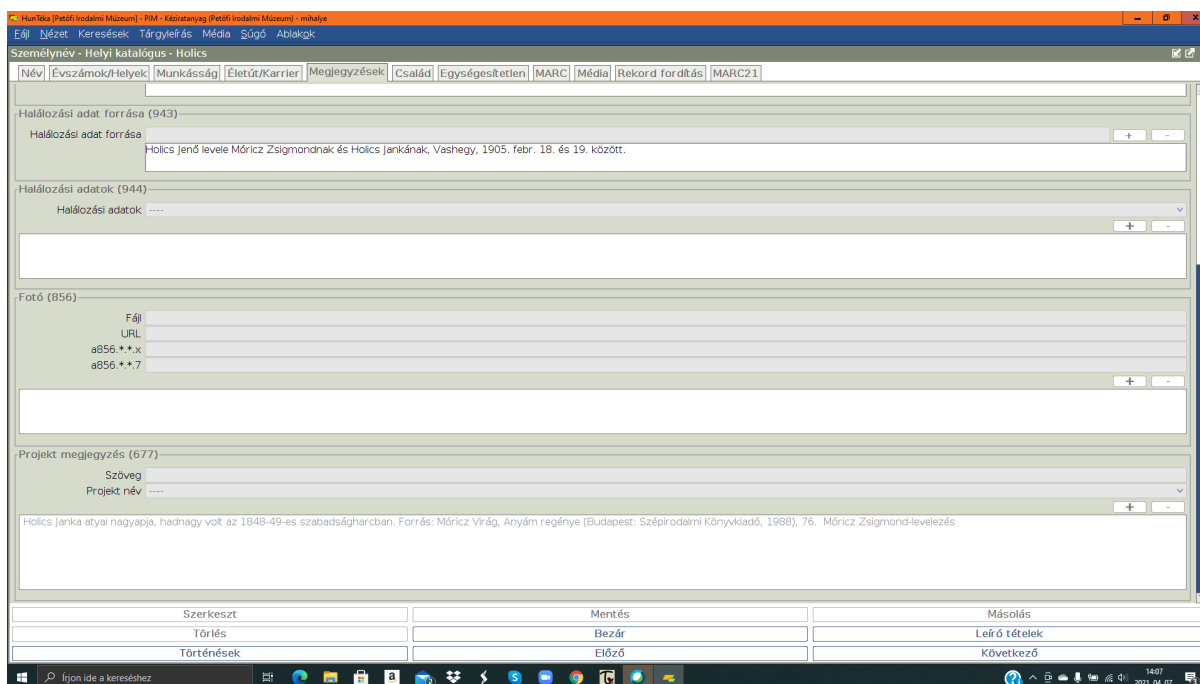
1. A PIM Névterével való együttműködés kialakítása első szakaszában a személynevek problematikájával foglalkoztunk, melynek során a következő kérdésekre kellett választ adnunk:
 - hogyan jussanak el a projekt során előkerülő új adatok a PIM névtér szerkesztőihez?
 - mi legyen azokkal a nevekkkel, amelyek nem kerülhetnek be az éles névtérbe, mert még nincs elég adat, de fontosak a projektben?
 - mi legyen azokkal a nevekkkel, amelyek csak az adott projektben bírnak jelentőséggel?
 - mi legyen azokkal a jegyzetekkel, amelyek csak az adott projekt szempontjából adekvátak?

Megoldásként kialakítottunk egy általános workflow-t a szövegtároló projektek során előállt adatok továbbításához, illetve a Huntékában a következő fejlesztéseket végeztük el:

- új 090: Adatbázis a digitális filológiai projekteknek (összes a projektben előforduló név)

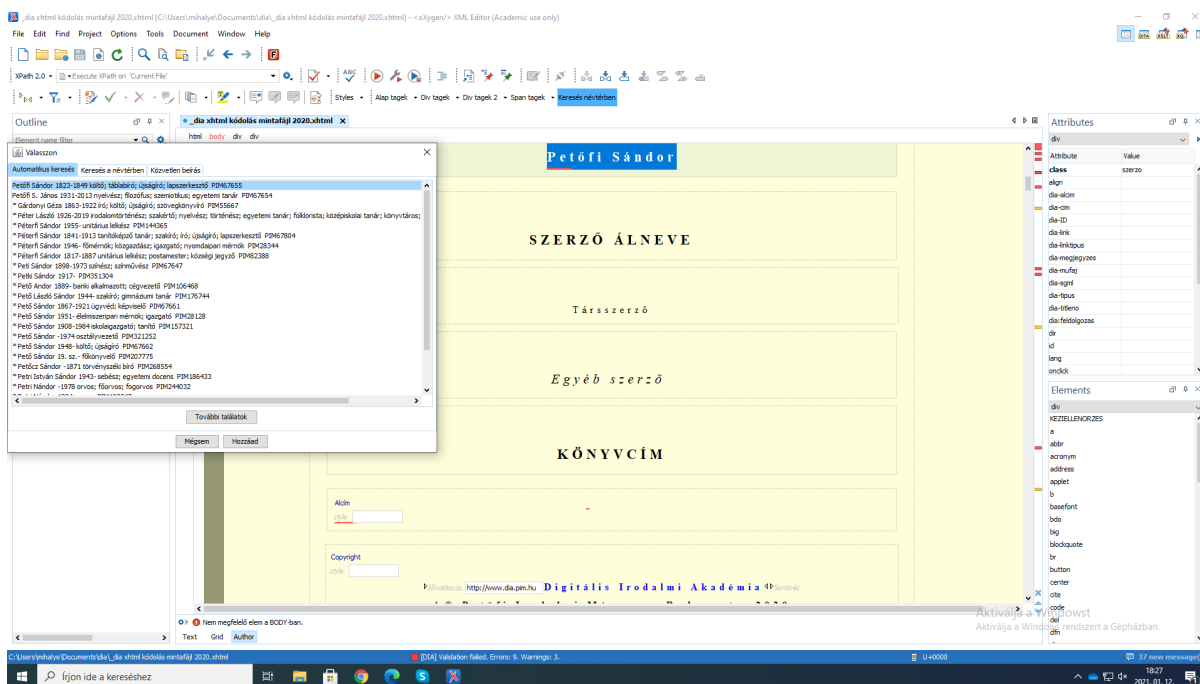


- új forrástípusok felvétele: analitikus forrás és lektorált internetes forrás
- új megjegyzés mező: projekt-megjegyzés összekapcsolva a projekttel



Ezek segítségével minden keletkezett adat helyet kapott a meglévő névtérstruktúrába. Folyamatban van a módszer kiterjesztése egyéb entitástípusokra (helynév, testületi név, terminus, műcím, bibliográfiai tétel), valamint az adatok Opac felületén való megjelenítése.

2. Az Oxygen-frameworkbe integrált névtér-funkció jelenleg a DIA (Digitális Irodalmi Akadémia¹⁹) frameworkjében valósult meg:



19 <https://pim.hu/hu/dia>

A szerkesztő a szövegben szereplő névalakot kijelölve egy kattintással beazonosíthatja a nevet a PIM Névterében, illetve egy újabb kattintással beemelheti a nevet azonosító adatokat annotációként. Folyamatban van a funkció implementálása a többi Oxygen-frameworkbe, illetve bővítése az egyéb entitástípusokra.

Publikációs, szűrési, keresési, rendezési lehetőségek


ATEI XML megjelenítési lehetőségei korlátlanok. A pilot projekt során arra koncentráltunk, hogy a felhasználók, kutatók számára leginkább hasznos vizualizációs módokat hozzunk létre. Ezért fejlesztettünk egy új eszközt a digitális faksimile és az átírt szöveg együttes megjelenítésére, a text-image linking szövegfeldolgozó módszert elsőként alkalmazva Magyarországon. Ennek lényege, hogy a kép zónái össze vannak kapcsolva az átírt szöveggel, tehát a felhasználó együttesen láthatja a kettőt.

Háromféle nézetet is kialakítottunk:

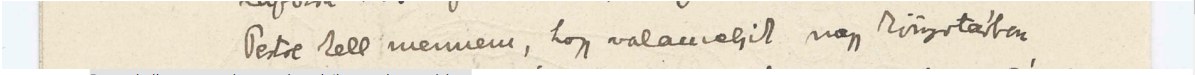
1. A kurzorral az átírt szöveg adott sora fölé menve megjelenik a faksimile megfelelő sora:

gyelmét előre is megköszöntem. Így várhatnék nyugodtan Nagyságod érkezésére, de közben még egy praktikus gondolatom támadt; ezt pedig az itteni nyomorúságos könyvtárnak köszönhetem. Tíz könyv közül nem kapok meg 3. at, a

1



legalapvetőbb munkák nagyrésze hiányzik vagy ki van adva. Nekem is már szép könyvtáram van, de mivel most januárban újra Németországba akarok



Pestre kell mennem, hogy valamelyik nagy könyvtárban próbáljak szerencsét. Gondolom Melich professor Úr segítségével csak bejutok valahová. Az itteni könyvtárban nem kaptam meg Nagyságodnak két művét, a Zsidó dalokat, továbbá a regényt. Nem kaptam meg a „Pester Lloyd” régi számait, melyekben Silberstein Adolfnak a kritikái jöttek, nem az akkori napilapok jelentékeny részét, hát így mit csináljak itt, hisz minden legkisebb anyagra szükségem van, mert az így össze gyűjtött anyagot i. történeti szempontból kell felhasználnom.

Most azért írok, hogy megtudjam Nagyságodtól: meddig marad Budapesten? mikor ió Kolozsvárra?



2. A kurzorral az átírt szöveg adott sora fölé menve láthatjuk a faksimile megfelelő részletét kisebb kontextusban:

legalapvetőbb munkák nagyrésze hiányzik vagy ki van adva. Nekem is már szép könyvtáram van, de mivel most januárban újra Németországba akarok utazni, minden krajcárkát félre kell tennem, könyvekre így most sokat ki nem adhatok. De az itt is keresett könyvekre, lapokra stb. föltétlen szükségem van; nincs más mód: Pestre kell mennem, hogy valamelyik nagy könyvtárban próbáljak szerencsét. Gondolom Melich professor Úr segítségével csak bejutok valahová. Az itteni könyvtárban nem kaptam meg Nagyságodnak két művét, a Zsidó dalokat, továbbá a regényt. Nem kaptam meg a „Pester Lloyd” régi számain, melyekben Silberstein Adolfnak a kritikái jöttek, nem az akkori napilapok jelentékeny részét, hát így mit csináljak itt, hisz minden legkisebb anyagra szükségem van, mert az így össze gyűjtött anyagot i. történeti szempontból kell felhasználnom.

Most azért írok, hogy megtudjam Nagyságodtól: meddig marad Budapesten? mikor jó Kolozsvárra?

3. Osztott képernyőn láthatjuk a faksimilét és az átírt szöveget (mindkettőn kijelölve az éppen olvasott szövegrész):

legalapvetőbb munkák nagyrésze hiányzik vagy ki van adva. Nekem is már szép könyvtáram van, de mivel most januárban újra Németországba akarok utazni, minden krajcárkát félre kell tennem, könyvekre így most sokat ki nem adhatok. De az itt is keresett könyvekre, lapokra stb. föltétlen szükségem van; nincs más mód: Pestre kell mennem, hogy valamelyik nagy könyvtárban próbáljak szerencsét. Gondolom Melich professor Úr segítségével csak bejutok valahová. Az itteni könyvtárban nem kaptam meg Nagyságodnak két művét, a Zsidó dalokat, továbbá a regényt. Nem kaptam meg a „Pester Lloyd” régi számain, melyekben Silberstein Adolfnak a kritikái jöttek, nem az akkori napilapok jelentékeny részét, hát így mit csináljak itt, hisz minden legkisebb anyagra szükségem van, mert az így össze gyűjtött anyagot i. történeti szempontból kell felhasználnom.

Most azért írok, hogy megtudjam Nagyságodtól: meddig marad Budapesten? mikor jó Kolozsvárra? meddig marad itt s innen hová szándékszik utazni? Mert ha június végén utazna Kolozsvárra, s ha

A szövegben szereplő annotációk kurzorral a (sárgával) kiemelt szövegrészek fölé menve jelennek meg.

A keresési, szűrési műveleteket itt is facetek használatával tesszük egyszerűvé, támogatóvá.

DBK Digitális Bölcsészeti Központ

barát

Keresés

barát

5 db találat (35 ms)

Vissza 1 **2** 3 Tovább

levél
[Heeger János – Mórincz Zsigmond \(1898-06-16\)](#)
S ha legalább egy héten egyszer írunk egymásnak – elég szép emléket birjuk nemcsak kéziratainknak, ej de kitűnő. ezért gratulálok hanem, ha úgy tetszik, egy sokkal maradandóbb köteteknek, ah melynek neve **barátság**, aha s mely non pecunia, rea fende et officio paritur.

levél
[Mórincz Bálint, Mórincz István, Pallagi Erzsébet – Mórincz Zsigmond, Mórincz Dezső \(1899 után\)](#)
Igen sokat aggódtam rajtatok hogy nem tudunk nektek pénzt küldeni, de csak azzal bíztattuk egymást hogy kérnétek ti zaklanátok ha másképp nem boldogulnátok. Hát Borosékat ott hagyta? meg szűnt a **barátság**? vagy az azért tart még?Hát az újságírást, hogy érte a 20 frtot forintot egész évre, vagy egy hónapra, azt hiszem egész évre, akkor pedig az talán kevés anyi fejtörésért.Mi élünk falusi emberek módjára, egy kicsiny faluba, még iskolája nincs, nem is volt soha, más faluba járnak a gyerekek.

levél
[Heeger János – Mórincz Zsigmond \(1898-06-19\)](#)

Szöveg típusa
[levél](#) (5)
Szerző
[Heeger János](#) (3)
[Mórincz Bálint](#) (1)
[Mórincz István](#) (1)
[Mórincz Zsigmond](#) (1)
[Pallagi Erzsébet](#) (1)

Automatikus névfelismerés

A Nyelvtudományi Kutatóközponttal való együttműködés keretében megkezdett újabb fejlesztés eredményeképpen a szerkesztőségi környezet része lesz egy másik funkció is, amely szintén a nevek beazonosítását szolgálja NLP (Natural Language Processing²⁰) eszközökkel. A Nyelvtudományi Kutatóközpont által fejlesztett e-magyar (<https://e-magyar.hu/hu/>) automatikus névfelismerő (Named Entity Recognition) moduljának²¹ beépítésével lehetséges lesz a szövegekben szereplő nevek automatikus megjelölése, méghozzá névtípusonként. Ezután következhet a fentiekben felvázolt beazonosítási procedúra.

Amennyiben nincs kapacitás a nevek manuális beazonosítására, a technológia mindenképp alkalmazható a publikációs felület keresőjébe beépítve, így a felhasználó a program által névként felismert szóalakok között tud böngészni.

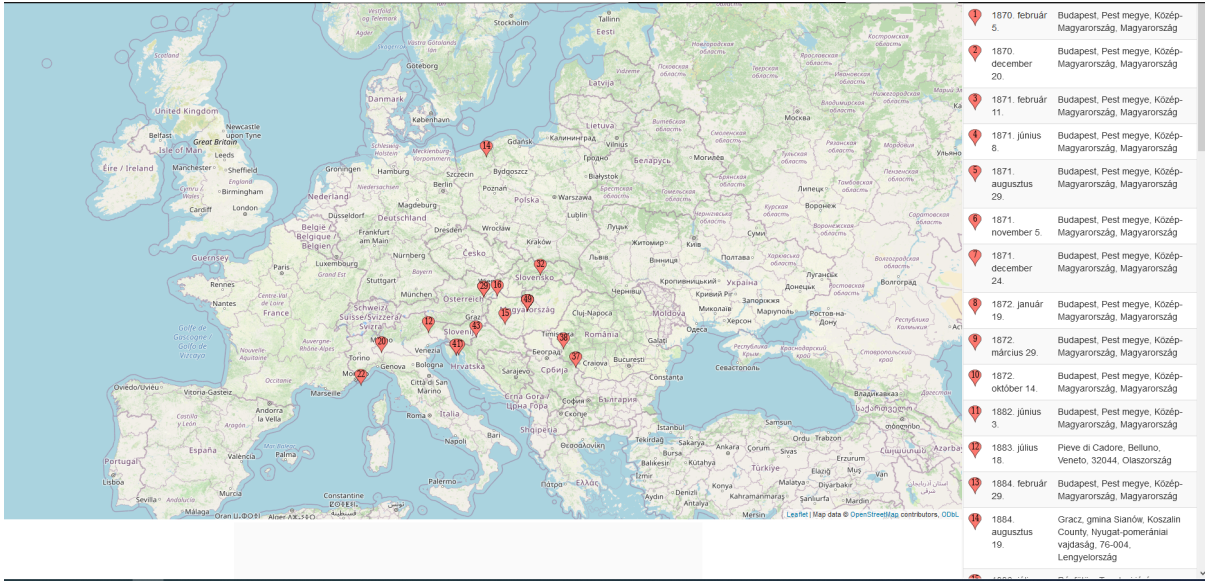
Adatvizualizációs lehetőségek

A metaadatok, illetve a szövegekben elhelyezett annotációk felhasználásával a legkülönbözőbb statisztikák, vizualizációk állíthatóak elő az egyes szövegtudásokról.

Levelezéskiadás esetén kézenfekvő a levelek feladási, illetve fogadási helyszíneinek térképen való nyomon követése:

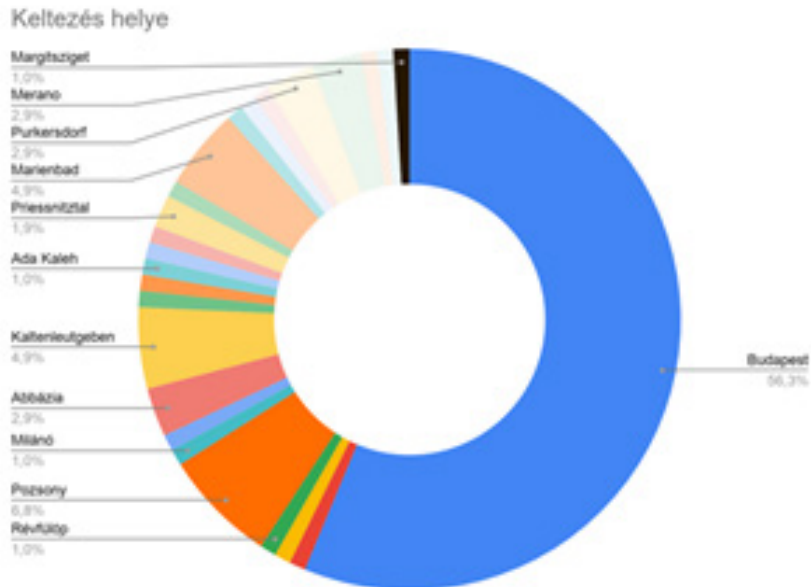
²⁰ https://en.wikipedia.org/wiki/Natural_language_processing

²¹ <https://e-magyar.hu/en/textmodules/emner>



Kiss József levélkeltezései. Ahol Kiss József élete során megfordult.

Érdekelhet minket a levélkeltezések helyszíneinek aránya:



Vagy kimutatást készíthetünk a levelezőpartnerek nemek szerinti eloszlásáról is:



Kiss József levelezőpartnerei:

Zöld: férfi

Sárga: nő

A lehetőségek száma végtelen, csak a képzeletünk szabhat határokat. Fontos azonban mindig szem előtt tartani, hogy a vizualizációk az adatok olyan aspektusait világítsák meg, amelyek anélkül nem, vagy nagyon nehezen értelmezhetőek. Hiszen a végső cél minden esetben a felhasználók számára hasznos szolgáltatások nyújtása.



A Mikes Kelemen program tapasztalatai és megújult működése

Görög Dániel
Országos Széchényi Könyvtár
gorog.daniel.aron@oszk.hu

Rényi Mátyás
Országos Széchényi Könyvtár
renyi.matyas@oszk.hu

The Mikes Kelemen Program: How it Works and Introduction of the Updated Service Interface

Author keywords: Mikes Kelemen Program, cultural heritage preservation, duplicate exchange, database, emigrant community

The presentation is an overview of the Mikes Kelemen Program - running since 2013 under the auspices of various public entities including the National Széchényi Library - in terms of its processes, results and future potential. Since its launch the Program has processed and offered into public use 250,000 documents collected from eight countries on four continents. The 90,000 documents so far placed offer us insight into the document needs of the domestic library system and those of Hungarian minorities abroad. After two years of development the initial HTML-based service interface listing the documents on offer was replaced by a new SQL database-based one in 2020.

The implementation was driven by knowledge gained in the first years of the Program, including document processing experience, utilization statistics for the documents on offer, and feedback from partner institutions that joined the Program.

The operation of the new database-based interface implements the management of duplicates exchange differently from the Hungarian practice. The Mikes Kelemen Program website is characterized by service-oriented operation and integrated processing and recommendation process management.

The operation enables accurate, reliable, automated and trackable document management, which may provide a blueprint for the overhaul of the national duplicates exchange system, replacing “digital paper-based” records such as Excel or Word.



A Mikes Kelemen Programot 2013-ban a Diaszpóra Tanács III. ülésének zárónyilatkozatával¹ indították útjára. Célja, hogy a Kárpát-medencén kívüli magyar diaszpórában felhalmozódott és a megsemmisülés szélére sodródott dokumentumörökséget összegyűjtse és rendezett módon hazaszállítva hasznosítsa.

Az adományok feldolgozására és a Kárpát-medencében történő elosztására a Nemzetpolitikai Államtitkárság az Országos Széchényi Könyvtárat kérte fel, annak hosszú időre visszanyúló, hungarikakutatásban² és fölőspéldány-kezelésben szerzett tapasztalati okán.

Az OSZK azzal, hogy elvállalta a Mikes Program lebonyolítását, visszanyúlt a fölőspéldány-elosztási hagyományhoz, módszereit azonban a 21. századi elvárásoknak megfelelően dolgozta ki. A fölőspéldány-csere olyan volumenét vállalta magára, amire már több évtizede nem volt példa, sem pedig kialakult gyakorlat. A tervezés nem annak alapján történt, mégis a korábbi fölőspéldány-kezelési gyakorlat és a Mikes Program folyamatai számos ponton átfedést mutatnak. Ugyanakkor az OSZK nemcsak saját állományának, hanem az egész Kárpát-medencei magyar könyvtári közösség gyarapítási lehetőségének tekinti a Mikes Program adományait.³

Mivel a nemzeti könyvtár saját kapacitása nem volt elégséges egy ilyen méretű adományegyüttes befogadásra, a program megvalósítása szükségessé tette az adományok számára egy önálló raktár létesítését. Ezért 2015-ben a Maglódi úton bérelt a Mikes Program számára egy 650 négyzetméteres raktárat, ezt követően pedig – egy meglehetősen nagyszabású költözés után – a Váci úton, egy 700 négyzetméteres bérleménybe került a folyamatosan bővülő állomány.

A hazaküldött adományok összetétele igen vegyes képet mutat. Az egészen apró, két-három könyvből álló könyvadományok mellett előfordul tízezres nagyságrendű magánkönyvtár is. Az állomány gerincét azoknak a nagy észak-amerikai szervezeteknek az adományai jelentik, amelyek több esetben a teljes könyvtári állományukat hazaküldték.

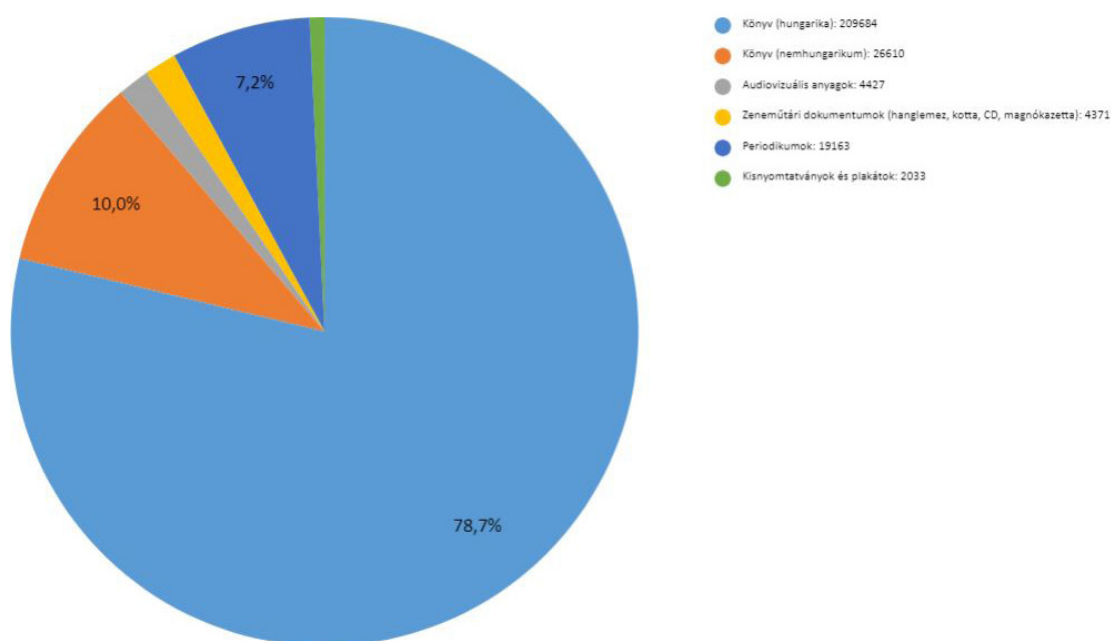
Az elmúlt 7 év során összesen 9 ország 29 városának több tucat magyar közössége ajánlotta fel a Mikes Programnak adományait, az adományok összmennyisége megközelíti a 300 000 dokumentumot. Az adományok több mint 90%-a az Amerikai Egyesült Államok, Kanada és Ausztrália magyar közösségeitől érkezett.

- 1 Magyar Diaszpóra Tanács III. ülés Zárónyilatkozat: <https://bgazrt.hu/wp-content/uploads/2019/03/06-A-Magyar-Diaszp%C3%B3ra-Tan%C3%A1cs-III.-%C3%BCI%C3%A9s%C3%A9nek-Z%C3%A1r%C3%B3nyilatkozata.pdf> (2021. 05. 22.)
- 2 Kovács Ilona: *Hungarika/diaszpóra-kutatás az Országos Széchényi Könyvtárban a Mikes Kelemen Program előtt* = Könyvtári Figyelő, 27. (63.) évf. 2017. 3. sz. 325-336. p. http://epa.oszk.hu/00100/00143/00348/pdf/EPA00143_konyvtari_figyelo_2017_03_325-336.pdf (2021. 05. 22.)
- 3 Mészáros Tamás: *Értékmérés a diaszpórában, avagy új gyarapítási lehetőség a Kárpát-medencei könyvtárak számára* = Könyv, Könyvtár, Könyvtáros, 25. évf. 2016. 1. sz. 3-7. p. http://epa.oszk.hu/01300/01367/00274/pdf/EPA01367_3K_2016_01_003-007.pdf (2021. 05. 22.)

A Nemzeti Könyvtár gyarapodása

A Törzsgyűjtemény mellett szinte az összes OSZK-s különgyűjtemény részesült a Mikes Program adományaiból, a Kézirattártól a Zeneműtárig. A kéziratanyagokra jellemző, hogy a személyi és szervezeti/intézményi archívumok nem válnak el egymástól.

Eddig⁴ négy nagyobb ciklusban érkeztek az adományok, melyek jelenleg a következő dokumentumtipológiát mutatják.



1. diagram

Mikes Kelemen Program feldolgozott dokumentumai

A feldolgozás lépései sok tekintetben megfeleltethetők a korábbi OSZK-s fölőpéldány elosztási gyakorlatnak. Első lépésként tipizálják a kiadványokat, majd a könyv jellegű dokumentumok esetében különválasztják a hungarikákat. Ezt követi a dokumentum azonosítása. Általánosan megállapítható, hogy az 1952 utáni dokumentumok – ellentétben a korábban megjelent anyagokkal – viszonylag gyorsan összevethetőek a törzsgyűjtemény állományával. Körülbelül 50-70 dokumentum azonosítására képes egy kolléga egy 8 órás munkanap folyamán, de a dokumentumok bizonyos köre csak a hagyományos cédulakatalógusok és a növedéki naplók, papíralapú nyilvántartások alapján azonosíthatók be egyértelműen. Az eddig feldolgozott 209 684 hungarika mintegy 10%-át az OSZK épületében dolgozó gyarapító kollégák azonosították. A sorsuk a további hasonlítás során dől el. Az OSZK a kiadványok többségét két példányban gyűjti, így hiány esetén az első és másodpéldányok kerülnek az állományba.

4 A program első három évének összefoglalója: Rényi Máttyás Attila: *A Mikes Kelemen Program 3 éve* = Könyvtárvilág, 5. évf. 2017. 6 sz. <https://mke.info.hu/konyvtarvilag/2018/01/a-mikes-kelemen-program-3-eve/4555/> (2021. 05. 22.)



A Széchényi Könyvtár Integrált könyvtári rendszerének gyarapító moduljából pontosan kimutatható a Mikes Program gyarapodáshoz való hozzájárulása. Eddig összesen 7569 könyvjellegű dokumentumból 8066 példányt vettünk állományba, ezek közül 2182 kiadványból a Mikes Program előtt nem volt példánya az intézménynek. Az OSZK mintegy tízezer részegységgel pótolta periodikagyűjteményének hiányait, és 114 eddig hiányzó időszaki kiadványt indított el az állományban. Az OSZK különgyűjteményei 3085 tétellel gyarapodtak a Mikes Program révén. A kéziratár jelentős gyarapodásai közé tartozik Nadas János (1903-1992) hagyatékának egy kisebb része, András Károly (1917-2004) újságíró, lapszerkesztő, politológus hagyatéka, Várdy Béla (1935-2018) amerikai magyar történész, egyetemi tanár hagyatéka, valamint Vászolyi Erik (1933-2013), a permi és az ausztráliai nyelvek tudós kutatójának hagyatéka.⁵ A különgyűjtemények gyarapodását részletesen bemutatja a nemzeti könyvtár 2020 júniusában indult blogsorozata⁶.

A Mikes-raktárban azonosított és jegyzékre vett dokumentumok mechanikus raktári rend szerint, egyszerű numerus kurrens azonosító alapján kerülnek a polcokra a raktár közel 2200 polcfolyóméterén. A véges raktári kapacitás és a folyamatosan beáramló anyag miatt az állományt rendszeresen mozgatni, forgatni kell. Nehézséget jelent, hogy az anyag folyamatos cirkulálása mellett statikus szempontokat is figyelembe kell venni, ami leginkább a folyóirat-gyűjteménynél okoz nehézséget. Statikus szempontot jelent az azonos kiadású művek, a többkötetes dokumentumok és az időszaki kiadványok összevárása, egy listára rendezése, ami a kijárlás megkezdését megelőző teljes feldolgozással lenne elérhető. A felhasználóbarát statikus szempontok teljes betartása időben jelentősen kitolja az amúgy is több éves kijárlási időszakot, így szükséges a különböző szempontok harmonizálása a feldolgozási folyamat során. Emellett az anyag eltérő volumenű beáramlása és elszállítása, továbbá a változó projektfeltételek adta ciklikusság olyan körülményeket teremt, ami a Mikes Program munkafolyamataitól nagymértékű adaptivitást követelt meg az elmúlt években.

A jelenleg (2021.05.22.) rendszerben szereplő dokumentumok 57%-a hazai, 43%-a külföldi kiadású. 61% kapott szakirodalom, 38% szépirodalom minősítést, 1% alatti az egyéb besorolás, amit a zenei és különgyűjteményi dokumentumokra használunk. A tipizáláshoz tartozik, hogy szakirodalom alá soroltuk a művészeti, vallási témájú, illetve az egyéb besorolás létrehozatala előtt a zenei dokumentumokat is, vagyis mindent, ami nem szépirodalom.

75 000-nél több már kijárlott tétel vár sorsára a raktárban, 22 000 vár első kijárlásra, és 3165 van éppen most kijárló listán.

A Kárpát-medence könyvtárainak gyarapodása

A dokumentumok kijárlása 2016. áprilisában indult meg, ekkor az OSZK honlapjának Mikes-aloldalán kerültek közlésre a listák. A felület hátránya volt, hogy a megrendelőket

5 Zichy Mihály: *Hazatérő hagyatékok - kép az emigrációról* https://nemzetikonyvtar.blog.hu/2020/09/04/hazatero_hagyatekok_kep_az_emigraciorol (2021. 05. 22.)

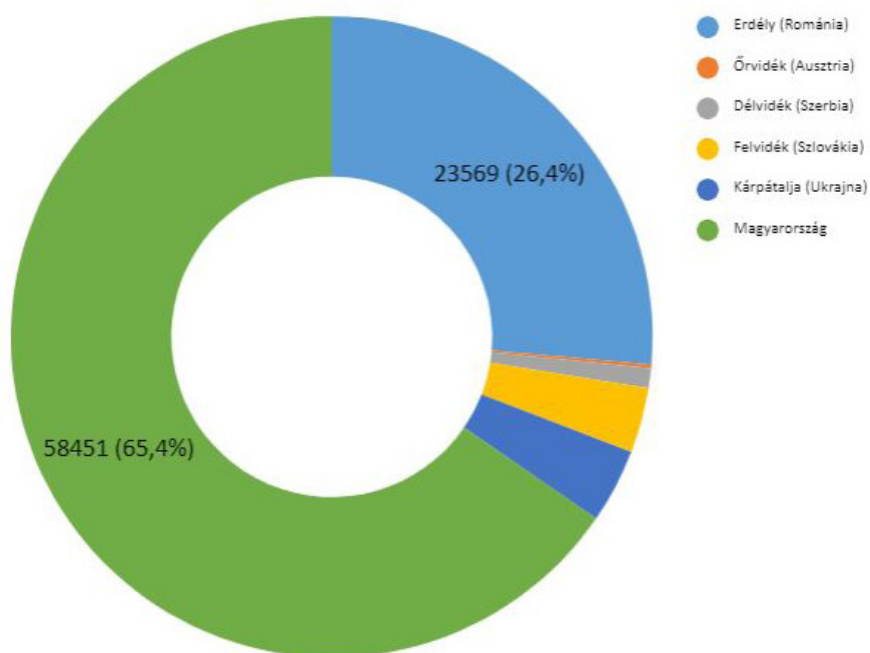
6 Kaposváriné Dányi Éva: *A Mikes Kelemen Program margójára - 1. rész* https://nemzetikonyvtar.blog.hu/2020/06/05/a_nemzeti_osszetartozasrol_a_mikes_kelemen_program_kapcsan (2021. 05. 22.)

nem lehetett egyértelműen azonosítani és a felajánlott dokumentumokról közölhető információ is limitált volt. Kezdetben ajánlólistáink akár 1000 dokumentumot is tartalmazhattak, de az első két év tapasztalati megmutatták, hogy ez a hétről hétre kikerülő mennyiség – kiegészülve a Könyvtári Intézet fölőspéldányjegyzékeivel – kezelhetetlen a gyarapító könyvtárosok számára, így 2017-től 500 címbe szabtuk meg listáink méretét.

A felajánlott dokumentumok hasznosulását sok tényező befolyásolja. Egyrészt a kérőoldali körülmények, úgymint a gyűjtőkör és az elszállítási kapacitás, másrészt a felajánló oldalról a könyvek állapota, a kiadványra beérkezett kérések száma, a listák nagysága, a felkínált dokumentumok keresettségé. A program indulásakor a listák hasznosulása – köszönhetően az újdonság erejének is – sok esetben elérte a 60-70%-os arányt. Négy év összesítésében a program keretében felajánlott dokumentumok 48%-a hasznosult.

Az 545 listán felajánlott 203 966 dokumentumra 147 669 kérés érkezett, amiből 84 880 kérést tudott teljesíteni a Program. Dokumentumtípus szerinti bontásban nagy különbségek mutatkoznak. A dokumentumokat a következő módon csoportosítjuk: könyv (hungarika), könyv (nemhungarika), zenei dokumentumok.

A kísérleti fázisban járó periodika (7%) kijánlás mellett könyvek közül a nemhungarika dokumentumok (35%) hasznosulási aránya az alacsonyabb, míg a hungarika könyv (49%) és zenei dokumentumok (51%) hasznosulása közel azonos. Ha a hasznosulási tendenciákat előre vetítjük, a Mikes Program kijánlott tételeinek közel 50%-os hasznosulásával tervezhetünk.



2. diagram: Teljesített kérések ország szerinti megoszlása



A Program fennállása óta teljesített kérések 2/3-a belföldi igénylés, míg a jelenlegi 184 partnerünk közül 30 határon túli. Tehát a 19% határon túli partner generálja a kérések 33%-át. Ebből az adatból világosan kiderül, hogy a határon túli magyar intézmények Mikes-dokumentumok iránti igénye jóval magasabb a hazai intézményekénél. Ezen igények kielégítésének pedig elsődleges feltétele a határon túlra való kiszállítás biztosítása, ami egybevág a Nemzetpolitikai Államtitkárság elképzeléseivel és a nemzeti könyvtárnak a 30/2014. (IV. 10.) EMMI rendeletben előírt feladatával, a határon túli könyvtárak hungarikumgyarapításának támogatásával.⁷

Az új szolgáltatási felület

2020 májusában, két év fejlesztés után elindult a Mikes Program új integrált rendszere a <https://mikesprogram.oszk.hu/> webcímen.⁸ A korábbi HTML-alapú kiajánlási nézet statikus megjelenési felületet biztosított, a tételek életútját nem tudta követni. Az új adatbázis képes a feldolgozási és kiajánlási folyamatokat egy rendszerben kezelni, így egyesítve a munkafolyamatok eddigi fragmentált egységeit és csökkentve a hibalehetőségeket. Az új adatbázis a Z39.50 protokoll segítségével, a kiadvány IKR-azonosítója alapján lekéri az OSZK adatbázisban kiválasztott katalógusrekord előre definiált mezőinek bibliográfiai adatait. A Mikes-rendszerbe importált és ott keletkező rekordok így a kiajánlást megelőző, majd az azt követő munkafolyamatokban is integráltan vehetnek részt. A rendszer biztonságos működése érdekében az adatbázisból rekordot törölni nem lehet, és az események naplózásának köszönhetően a változtatások visszakövethetők. A Mikes Program így az OSZK a saját állományába vett, vagy partnerhez eljuttatott dokumentumokkal is el tud számolni.

A dokumentumok útját hely és státusz jelölők jellemzik. Hely alatt értelemszerűen a pillanatnyi feldolgozási helyet, vagy raktározási egységet adjuk meg. A státusz jelölő arról hordoz információt, milyen munkafolyamatokon és állapotváltozásokon esett át a tétel. Státuszt változtatni, ahol indokolt, csoportosan, például polc vagy kiajánlási lista alapján is lehet. Némelyik státuszváltás automatikus helyváltással jár együtt. Azonban adott státuszról csak korlátozott számú, a feldolgozási láncban következő státuszra lehet változtatni, így elkerülendő a dokumentumok szabálytalan mozgása; például megelőzhető, hogy kiajánlási listára kerüljön a OSZK gyarapítási munkatársa által tartott szemlén át nem esett dokumentum. A lehetséges státuszváltásokat vizuálisan ábrázolva kirajzolódik a dokumentumok előtti három fő útvonal (OSZK beküldés, kiajánlás, selejtezés), illetve látható, hogy a partnerek által nem igényelt tételeket elvileg végtelenszer át lehet küldeni a kiajánlási folyamaton.

7 30/2014. (IV. 10.) EMMI rendelet: http://njt.hu/cgi_bin/njt_doc.cgi?docid=168674.348436 (2021. 05. 22.)

8 A felülethez felhasználói útmutató elérhető útmutató: Rényi Máttyás Attila: *A Mikes Kelemen Program és Megújult Szolgáltatási Felülete*. Könyv, Könyvtár, Könyvtáros 29 (7-8.), 3-12. <http://ojs.elte.hu/3k/article/view/1082>. (2021. 05. 25.)



periodika listákon közlünk. A kiajánlási folyamat után megmaradt dokumentumok *maradéklistákon* szerepelnek, ahol esélyt kapnak az újabb hasznosulásra a programhoz később csatlakozó felhasználók megrendelése által.

A bibliográfiai tételek az OSZK katalógusának rekordazonosítói szerint vannak rendezve, így az ismétlődő tételek és a többkötetes művek kötetei többnyire egymás mellé kerülnek. A tételek szűrése „szakirodalom/szépirodalom” és „hazai/külföldi” paraméterek, valamint ezek kombinációi alapján lehetséges. Listanézetben a kiadvány címét, szerzőjét, kötetszámát, megjelenési évét, állapotára vonatkozó megjegyzéseket, valamint a tételre beérkezett igénylések számát látja a felhasználó. Utóbbi két lehetőség új szolgáltatás: mivel a már leadott igénylések száma is látható, több azonos kiadvány példányainak kiajánlása esetén csökkenthető az ütköző kérések előfordulása. Illetve a felhasználók bizalmának fenntartása érdekében fontos, hogy dokumentumaink fizikai állapotáról előzetesen tájékoztatni tudjunk; különösen mert az eddig feldolgozott hungarika könyveink kiadási évének módusza és mediánja is az 1970-es évtized elejére esik.

Az adatbázis beindítása óta eltelt egy évben 14765 érvényes igénylésből 11965-öt teljesítettünk, az igénylések teljesülésének aránya így a korábbi 55%-ról 81%-ra emelkedett, és a teljesített igényléseknek majdnem harmada határon túli. A számok azt jelzik, hogy a tájékoztatás színvonalának emelésével és a kiszállítás megszervezésével a kiajánlás eredményessége fejleszthető.

A felhasználók egy adott tételre vonatkozó igénylését akár rögtön a listán, illetve később, a felhasználói adatoknál (*Saját tér*) is visszavonhatja.

A lista lezárulta után – az eddig megszokottnál gyorsabban, akár automatizálhatóan – megtörténik a kiértékelés. Az elnyert kiadványról automatikus levélben kapnak értesítést a könyvtárak. A *Saját tér / Igényléseim* menüpontban nyomon követhetik az új felület bevezetése óta leadott rendeléseik sorsát. Az egyes tételek állapotát különböző háttérszínek jelzik. Az el nem nyert dokumentumoknál lehetőségük van arra, hogy értesítést kérjenek a dokumentum újbóli felbukkanásáról.

A Mikes Program legnagyobb jövőbeni kihívása, egyben az adatbázis fejlesztésének fő iránya a 75000 dokumentumnál nagyobb, kiajánláson sikertelenül átesett, tehát kontraszelektált anyag kezelése. A megoldást egy olyan kiajánlási felület jelentheti, amelyen határidők és heti fel- és letöltések nélküli platformon érhető el sokezer tétel; a mérete miatt nem feltétlenül lista nézetben, de részletes keresőkkel ellátva. Egy ilyen felületen a partnernek nem kellene több száz vagy több ezer tételt végignéznie, viszont olyan dokumentumot is megtalálna, amely évekkorábban esett át a feldolgozáson és első kiajánláson. Ilyen platformon nem csak az átláthatatlan mennyiségű megmaradt dokumentum kapna még egy hasznosulási esélyt, hanem az igénylések teljesítésének aránya is javulna, hiszen a kínálat nagysága csökkenti a versenyt, és a megrendelt tétel rögtön eltűnne nyilvános kiajánlási felületről.

A Mikes Program adatbázisa együtt tartva a beérkezett dokumentumok bibliográfiai adatait, hozzájárulhat az emigráció olvasási és könyvgyűjtési szokásainak vizsgálatához,⁹ aminek köszönhetően lehetővé válik a terület hosszú távú kutathatósága és a jövőben – további fejlesztéseket követően – a magyar diaszpóra könyvadatbázisának is alapjául szolgálhat.

9 Az emigráns irodalom kritikai értékelésében a cikk társszerzője él a Mikes Program nyújtotta lehetőségekkel. Görög Dániel: Emigráns esszék 1-2.
<http://www.naputonline.hu/2021/02/14/gorog-daniel-emigrans-esszek-1-resz/>
<http://www.naputonline.hu/2021/05/07/gorog-daniel-emigrans-esszek-2/> (2021. 05. 22.)



A könyvtári együttműködés informatikai támogatása: adatmodell, workflow, rendszerfelépítés: Együttműködési lehetőségek a nyílt forráskódú platformokban: OKP, FOLIO, ReShare

Lendvay Miklós

Országos Széchényi Könyvtár, Országos Könyvtári Platform Projektvezető

lendvay.miklos@oszk.hu

ORCID: [0000-0001-8065-5282](https://orcid.org/0000-0001-8065-5282)

IT solutions supporting the library collaboration: data model, work-flow, system architecture: Collaboration Opportunities within Open Source Platforms: HNLP, FOLIO, ReShare

An essential goal of library informatics is to create open-source systems through community collaboration. Primary examples of open solution Integrated Library Systems, such as KOHA, Evergreen, or the Open Library Environment (Kuali OLE), have been born out of this notion.

Since 2016, the librarian and developer professional communities have been working together to take this framework to a higher level. Building on learnings from prior system developments, a new modular, micro-service based platform was created. The platform was named FOLIO, short for 'The Future of Libraries is Open', to reflect its open and flexible nature. Today, FOLIO platform and its relevant modules are widely used by a number of medium-sized and national libraries (e.g. the Italian National Library in Florence).

The objectives of the Hungarian National Library Platform (HNLP) development, launched in 2016, are very much in alignment with the above: to re-conceptualise services offered by national libraries, to explore new ways of collaboration, to revolutionise common catalogue and inter-library loan, and to make entity-based data connections available beyond the world outside libraries through integration to the Hungarian National Namespace. And first and foremost, to offer the most advanced services and state-of-the-art IT technology to library users.

The National Széchényi Library Hungary has been part of the FOLIO community since its inception, to have a stake in its strategic direction and to benefit from the developments taking place internationally. Our long-term vision is to enable seamless module compatibility between the two systems so that libraries can use a flexible configuration that best serves their needs. The main pillars of the development are identical for both FOLIO and HNLP: (i) an entity-based data model, (ii) the creation of a meaning-based integrated architecture through modularity for any number of institutions and any institutional hierarchy, and (iii) the free configurability of workflows across the system / flexible workflow design. Where are HNLP, FOLIO and ReShare, the major collaborative module for interlibrary loan of the latter on this path right now? What solutions are provided for the basic pillars, and what objectives are still to be achieved?

Keywords: new generation library platform, Hungarian National Library Platform (HNLP), the Future of Libraries is Open, FOLIO, namespace



A könyvtárinformatika fontos törekvése, hogy közösségi összefogással nyílt forráskódú rendszereket hozzon létre. Az Integrált Könyvtári Rendszerek terén így született meg pl. a KOHA, Evergreen vagy az Open Library Environment (Kuali OLE).

2016-tól ezeknek a rendszereknek a tapasztalataiból kiindulva az ezeket használó könyvtárosok és a fejlesztők közössége újabb szintet lépett: olyan moduláris, microservice alapú platformot hoztak létre, amelynek a nyitottságát és flexibilitását a nevében is kifejezték: a könyvtárak jövője nyitott - The Future of Libraries is Open (FOLIO). Ma már egy nemzeti könyvtár (a Firenzei Olasz Nemzeti Könyvtár) és számos középmezretű könyvtár vezette be a FOLIO platformot, annak az adott könyvtárak számára releváns moduljait.

A 2016-ban indult Országos Könyvtári Platform (OKP) fejlesztés alapcélja ugyanez volt: új alapokra helyezni az országos könyvtári szolgáltatásokat, új utakat nyitni az együttműködésben, forradalmasítani a közös katalógust és a könyvtárközi kölcsönzést, a Magyar Nemzeti Névtérhez integrálva a könyvtári világon túlra is kinyitni az entitás alapú adatkapcsolatokat. És legfőképpen: a könyvtárhasználók számára a technika által nyújtott legmodernebb szolgáltatásokat nyújtani.

Az Országos Széchényi Könyvtár a kezdetektől képviselteti magát a FOLIO közösségben, hogy a nemzetközi fejlesztésekből az OKP is profitáljon, és az OKP is inspirálhassa a FOLIO fejlesztési irányait. A hosszú távú cél az, hogy a két rendszer moduljait szabadon lehessen kombinálni, és a könyvtárak az igényeiknek legjobban megfelelő konfigurációt használhassák. A fejlesztések fő pillérjei mind a FOLIO, mind az OKP esetében azonosak: (i) az entitás alapú adatmodell; (ii) a modularitás, tetszőleges számú intézmény és intézményi hierarchia értelemszerű integrált kapcsolódásának a megteremtése; (iii) a munkafolyamatok rendszeren átívelő szabad konfigurálhatósága, azaz flexibilis workflow-kialakítás.

Hol tart ezen az úton az OKP, a FOLIO és annak egy jelentős kollaboratív modulja, a könyvtárközi kölcsönzés (ReShare)? Milyen megoldásokat nyújtanak az alappillérek tekintetében, és mik a még elérendő célkitűzések?

Ökoszisztéma, kapcsolódás, együttgondolkodás

Az ember alapvető igénye a kapcsolódás; csakis közösségben, kapcsolódásban bontakozhat ki emberhez méltó teljességében. A XXI. század technológiai értelemben teljesen új és újszerű kapukat nyitott ennek az alapvető emberi igénynek a kiszolgálására. Új modellek jöttek létre az információ létrehozásában, megosztásában, újrahasonosításában; de az online emberi találkozások is általánossá váltak, bővítve a személyes találkozások és interakciók lehetőségét. Ezek a minőségileg teljesen újat hozó fejlődési ugrások visszahatnak az alapvető igényekre, újabb képességek fejlődnek az embernél, és újabb, váratlan kihívásokkal találja szembe magát az új lehetőségekkel való bánásmód kapcsán. A technológia áldásai és káros hatásai egyaránt újabb dimenzióban, erőteljesebb intenzitással jelentkeznek.

A könyvtár az a központi hely, amely a hiteles információ széleskörűen elérhető közvetítésében és a csendes, alkotásra alkalmas nyugodt terek biztosításával kiemelkedő szerepet játszik a szellemi munka területén. Emellett, sok esetben ezzel párhuzamosan és ettől elszakadva jelent meg egy olyan információáradat a neten, amely a könyvtári állományokhoz képest nem a megszokott szerkesztőségi, hitelesítési folyamatokon keresztül jön létre, és strukturáltságát elsősorban az automatizált gépi felismerés szemantikája teremti meg, robusztus keresőmotorok alkalmazása révén. Az internetes tartalomszolgáltatás könnyen és sok helyről és eszközön elérhető, emiatt a felhasználók nagy része a tájékozódásban, kommunikációban és a tartalmak megosztása kapcsán is innen indul. A szakmailag igényes, hiteles és ritka anyagok gyűjteménye még mindig a könyvtárba vezeti a felhasználót, de a két világ találkozása érdekében a könyvtárak munkafolyamatait át kell alakítani, a tartalmak és a szereplők sokszínűségét figyelembe kell venni, ezeket egymással demokratikus módon, egyenrangú szereplőként össze kell kötni, hogy valóban releváns és hiteles információk legyenek könnyen elérhetők a könyvtárak konzorciumainak gyűjteményeiben is. A könyvtáron belüli, a könyvtárak közötti, külső szereplők (kiadók, tudósok, kutatók, fotográfusok, művészek, szerzők, nyomdák, folyóirat szerkesztőségek stb.) bevonásával szélesített közösség együttműködésében az online tér kommunikációs és megosztási lehetőségeit modern platformokban célszerű hasznosítani úgy, hogy ezek, az alkalmazott informatikai rendszerek tekintetében jelenleg még elkülönülő közösségek és szolgáltatott tartalmak szervesen összekapcsolódhassanak, egy értelemszerű, szemantikus hálóban.

Meghatározó az ökoszisztéma, amely az egyes embert, illetve a közösségeket körülveszi. Ebben az ökoszisztémában célszerű megtalálnunk és áramvonalasítanunk az egyes könyvtárak, könyvtárak konzorciumainak és gyűjteményeinek helyét, és a most már közvetlenül is bevonható és bevonni kész szereplők sokaságának együttműködését. Ma már a legtöbb résztvevő nemcsak fogyasztó, hanem tartalom-előállító is egyben. Ha a könyvtárat meg szeretnénk tartani a szellemi élet középpontjának, akkor a sokrétű kapcsolódások folyamatosan változó csatornáit kell megteremtünk és a meglévőket akadálymentesen átjárhatóvá tennünk az együttműködést támogató könyvtári platformon.

A COVID jelentősen felerősítette az online együttműködések informatikai támogatását, miközben a digitális készségek is minden szereplő tekintetében ugrásszerűen fejlődtek. Ez részben megkönnyíti, másrészt a helyzetből adódóan meg is követeli a digitális transzformációt.

A technológiai átállás nem technológiai kérdés!

A technológia átállás azonban nem technológia kérdés. Az adatok és a technológia valójában csak a jéghegy csúcsa! Amennyiben a könyvtárat, mint intézményt, illetve a könyvtári konzorciumokat, mint intézményi együttműködési egységet tekintjük, egy nyitott platformra való átállásnak az alapja, alfája és omegája az ember, a folyamat meghatározó eleme pedig az a kultúra, amelyben az emberi közösség létezik, és amelyet aktívan alakít. A sikeres átállási folyamat záloga a résztvevő emberek motiválása, fejlesztése, bevonása; valamint a kultúra ápolása és kiteljesítése.



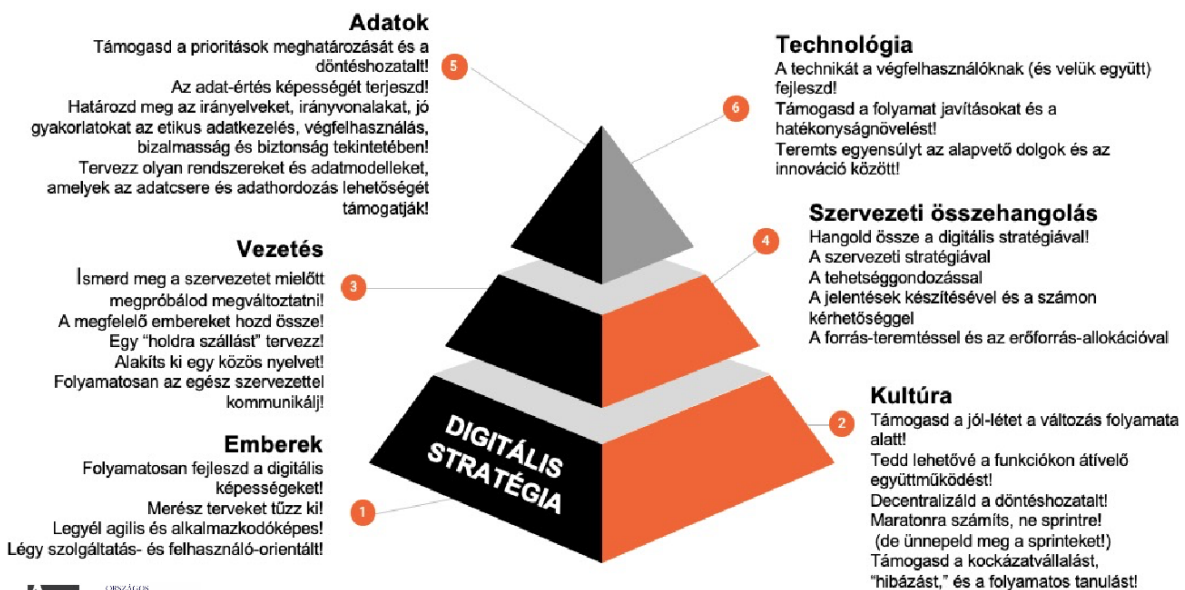
A technológiai átállás emberi és kulturális szempontjai, a projektmenedzsment lehetőségei és kihívásai

A projektmenedzsment hatékonyan támogathatja és katalizálhatja ezeket a folyamatokat. Az „Athenaeum 21” tanácsadó cég kutatta ezeknek a hatékonyságát, és számos alapvető jó tanáccsal tarisznyázta fel azokat, akik ezen az úton járnak. Bár ez a terület általában nem kutatás-fejlesztési terület, mégis: már a bevezetés folyamata is, az egyes területek összerakása, összefűzése, a kommunikációs láncok kialakítása intézményen belül és intézmények között egyaránt nem előre ismert, szervezésben és technológiai szempontból is valami korábban még nem létező új jön létre. Nem szabad félni a hibázástól, mert akkor eleve kizárjuk a folyamatból a még nem ismert innovatív elemeket, amelyek adott esetben a korábbi gyakorlatot képesek forradalmasítani, megújítani (1. ábra).

A technikai / technológiai megvalósítás (a jéghegy csúcsa) és az emberek és a kultúra mint az átállás alapjai (a fundamentum) között ott van a két nagy terület, amelyre érdemes a szervezet vezetésének és a projektvezetésnek egyaránt nagy figyelmet fordítani: a vezetés mibenléte és módszertana, illetve a szervezeti összehangolás.

Az Athenaeum 21 kutatása alapján a kudarcot vallott projektek négy legfontosabb jellemzői így fogalmazhatók meg:

- 1) Az emberek és a kultúra elsődleges fontosságának a figyelmen kívül hagyása
- 2) A vízió hiánya (vezetés)
- 3) Az infrastruktúra és az innováció közötti egyensúlyvesztés (technológia)
- 4) A maraton iránti elkötelezettség hiánya (vezetés)



Forrás / Copyright / Courtesy of: Athenaeum21 Consulting www.athenaeum21.com

1. ábra: A digitális stratégia megvalósításának alkotóelemei; szerző / forrás: Athenaeum 21

Az emberi lény többek között akkor egészséges, amikor önmagáról rendelkezhet, beleszólhat az ügyeibe, szabadnak érzi magát, szabadon szólhat és meghallják; lehetősége van megismerni és meghallani másokat. Ha szabadon kapcsolódhat, és nincs elszigetelve. Ha olyan viszonyok között tevékenykedhet, ahol ott születik a döntés, ahol a végrehajtás történik, így használhatja kompetenciáját a közösség javára és képes viselni a döntéseiért a felelősséget.

Az egyes ember és a közösség egészségének fontos része, hogy a közösségben, a társadalomban ezek a szempontok megvalósuljanak. Az első világháború folyamán a Monarchia, egy soknemzetű állam szülöttjeként Rudolf Steiner megfogalmazta és részletesen kidolgozta azt a modellt, ami alkalmas ennek tagolására, és az emberiség régi, a nagy francia forradalom által zászlóra tűzött szabadság - egyenlőség - testvériség hármasságát valósítja meg; ő ekkor ezt szociális hármastagozódásnak (soziale Dreigliederung) nevezte. Ennek a megközelítésnek a lényeges eleme az, hogy a forradalom három eszménye a társadalmi együttműködés egyes szektoraiban - **szellemi élet - jog - gazdasági élet** - valósulhat meg, és a társadalom minden szegmensében egyszerre nem lehetnek teljes mértékben jelen. A **jog** feladata biztosítani a résztvevők egyenlőségét. Szabadság uralkodhat a **szellemi területen**: bárki állítása szabad és érvényes lehet (figyelembe véve azt, hogy a közlés nem sértheti más méltóságát). A szabadság megvalósításának fontos lépése a szuverén ember önművelése, képességeinek folyamatos fejlesztése, a másik ember tiszteletben tartása és támogatása. Testvériség valósítható meg a **gazdaságban**, ahol mindenki a másik érdekében dolgozik és végzi a tevékenységét. Ezek a területek egymástól független végrehajtó szervekkel kell működjenek, hogy a saját területükön a másik szektor ne szűkítse be kibontakozásuk lehetőségét.

A FOLIO és az OKP alapelvek az együttműködés vonatkozásában

Mi következik mindebből egy könyvtári platform létrehozásakor? A technika az ember vágyainak és gondolkodásának tükrö, elhatározások, szándékok valósulnak meg benne. Ha a fent leírt szociális hármastagozódást meg kívánjuk valósítani, akkor olyan rendszereket kell létrehozni, amikben a szereplők szuverén módon vehetnek részt, információkhoz juthatnak és információt oszthatnak meg, szabadon, egymás tiszteletben tartásával.

A publikációk az ember kapcsolódási igényéből születnek, a könyvtár alapja a kommunikáció, a közlésvágy megvalósulása. Minél előrehaladottabb a demokratizálódási folyamat, annál több és többfajta publikáció jelenhet meg.

A megvalósítandó platformunkon a fentiek alapján kitűzött céljaink többek között a választás szabadsága; kapcsolódási lehetőségek sokasága; a hozzájárulás / szerepvállalás lehetősége; a szabad információhozzáférés; a munkafolyamatok együttműködést lehetővé tevő kialakítása.



A FOLIO közösség ezeket a vezérelveket fogalmazta meg a platform kifejlesztésének kezdetén:

- Nyitott, bizalmon és tiszteleten alapuló együttműködés
- Helyi döntéshozatalt lehetővé tevő mechanizmus
- Fenntartható közösség és ökoszisztéma létrehozása
- Nyitott és egyenlőséget biztosító platform minden közösségi partner számára

Mint ahogy az emberi szervezet szervei, olyan a FOLIO és az OKP platformja is – képes a közösségben szükséges együttműködési funkciókat intenzifikálni, a terhelés növekedésével a technológia képes tartalékokat mozgósítani, a futtatott folyamatokat párhuzamosítani. Ahhoz, hogy ez hatékonyan tudjon megtörténni, strukturált hozzáférest kell biztosítanunk sokféle résztvevőnek az online együttműködés lehetővé tételével. A korábbi szűk értelemben vett könyvtárosi feldolgozás és a felhasználók és olvasók különválasztása helyett sokféle szereplő számára nyílik ki a publikálás, adatgazdagítás lehetősége. Ilyen szereplők többek között: felhasználók, könyvtárak, konzorciumok, kulturális partnerek (levéltárak, múzeumok, kiadók, tudósok, stb.), adatok létrehozói általában, technológiát és szoftvert kialakító szakemberek, szoftver illetve hardver támogatást nyújtó szakemberek. Sokan és sokféleképpen építik ezt a közös halmazt, és minden területen az együttműködés hozza meg az eredményeket.

A FOLIO és OKP platform jellemzői

Régóta és ugyanezen az alapelven működik kiválóan a Wiki-birodalom: a WikiData / WikiMedia / Wikipedia / Wiktionary nyitott szoftver, amely szabadon újrahasznosítható és felhasználható adatok létrehozását teszi lehetővé egy nyitott struktúrában, ahol az adatok ellenőrzését is a közösség maga végzi, dinamikusan formált szerkesztői közösségekben, egy jól dokumentált struktúrában. A kultúra minden területe otthonra talál itt, sok nyelven, sokféle gondolkodásmóddal születnek tartalmak, amelyek életterülettől függetlenül szabadon átjárhatók. Az egyértelmű azonosítók alkalmazásával a Wiki-birodalom szegmensei pontosan és megbízhatóan kapcsolódnak össze, releváns tartalmak szinte végtelen hálóját alkotva. A legnagyobb kincs, hogy a szakterületről leghitelesebb információval rendelkező személyek szerkeszthetik az adott területről szóló cikkeket, oszthatnak meg képeket, térképeket stb.

A számítógépen levő digitális könyvtár is könyvtár, a világháló is kultúrkincs és a webarchiválás sem új keletű megőrzése a kulturális tartalmaknak. Hogy maga a könyvtár is elektronikus tartalmakból és részben webből van, lehetővé tette a megosztás forradalmát, és újradefiniálta a hagyományos könyvtár fogalmát. Így lettek a korábbi Integrált Könyvtári Rendszerekből könyvtári platformok, a nyílt forráskódú fejlesztések területén is. Nagymértékű nyitottsággal, tudva, hogy a platform valódi értékét azzal lehet mérni, hogy a platformon **kívüli** szereplőknek mit képes nyújtani. A szolgáltatás és a szereplők munkafolyamatainak és adatainak a hiteles összekapcsolása az elsődleges célkitűzés, és ez nem áll meg a platform határainál.

Ezek a követelmények meghatározzák az adatok és az alkalmazott technológia kapcsán támasztott elvárásokat is. Az adatok kapcsán általunk alkalmazott fő irányelvek: Támogasd a prioritások meghatározását és a döntéshozatal! Az adat-értés képességét terjeszd! Határozd meg az irányelveket, irányvonalakat, jó gyakorlatokat az etikus adatkezelés, végfelhasználás, bizalmasság és biztonság tekintetében! Tervezz olyan rendszereket és adatmodelleket, amelyek az adatcsere és adathordozás lehetőségét támogatják!

A technológia kapcsán a következő elvárások érvényesülnek: A technikát a végfelhasználóknak (és velük együtt) fejleszd! Támogasd a folyamat javításokat és a hatékonyságnövelést! Teremts egyensúlyt az alapvető funkciók és az innováció között!

A FOLIO és OKP egyaránt az ebben az értelemben lefektetett alapok megteremtését tartja elsődlegesnek. 5 év közös gondolkodás és fejlesztés eredményeképpen a közösség létrehozta az ezen alapelvek mentén működő platformot, amely 2020 nyarán érte el az akkor publikált Goldenrod (Aranyvessző) - a FOLIO release-ei virágok neveit viselik - release-zel a közepméretű könyvtárak által könyvtári platform terméként választható szintet. Az eddigi eredményhez elvezető projektek gyorsaságukkal és praktikusságukkal kiszolgálták a közvetlen könyvtárosi szükségleteket, így sok olyan hagyományos megoldás is megszületett, amely az új struktúrákban nem lett volna feltétlenül szükséges, de megkönnyítheti az átmeneti időszakot a korszerűbb megoldások irányába. A közösség a munka során azt is felismerte, hogy csak egy átfogó stratégiai vízió és erősebb irányítottság vezethet a minden elemében igazán újdonság erejű platform létrehozásához. Így 2021-ben a közösség kialakította a FOLIO új kormányzási modelljét, amely alapján választott közösségi tanács, terméktanács és technikai tanács, mindegyik tanács 15-15 választott taggal, irányítja a következő években a fejlesztést és bevezetéseket, a születő rendszerek összekapcsolását.

Alappillérek: adatmodell / modularitás / intézményi konfigurálhatóság / szabadon definiálható munkafolyamatok

A fejlesztések fő pillérjei mind a FOLIO, mind az OKP esetében azonosak: az entitás alapú adatmodell, a modularitás, tetszőleges számú intézmény és intézményi hierarchia értelemszerű integrált kapcsolódásának a megteremtése, a munkafolyamatok rendszeren átívelő szabad konfigurálhatósága, azaz a rugalmas workflow-kialakítás. A modulok és mikroszervizek szabadon kapcsolhatók össze, konfigurálhatók az igények függvényében. A könyvtárközi kölcsönzés / ReShare modul számos meglévő könyvtári rendszerhez kapcsolódik, összekapcsolva a konzorcium sokszínű résztvevőjének adatait és rendszereit.

Az elektronikus forrásmenedzsment (Electronic Resource Management – ERM) modul a FOLIO-ra épít. A különféle szempontok és konstrukciók miatt több változata is megszületett, amelyek választhatók, de párhuzamosan is üzemeltethetők. Létezik egy EBSCO változat, amely az EBSCO meglévő ERM-jét integrálja szervesen a FOLIO ökoszisztémába. A német FOLIO közösség kifejlesztette a FOLIO natív változatú ERM-jét, amely a német könyvtárszövetségek elmúlt 15 évben megfogalmazott igényeit valósította meg (az angol Knowledge Integration fejlesztő cég közreműködésével).



A német könyvtáros közösség egy összetett konzorciumban fűz össze számos meglévő szolgáltatást. Magyarországon a Qulto rendszerhez integrált ez a natív FOLIO ERM modul.

Az alapvető célkitűzés ebben az esetben is érvényesült: nyújts egy alapmegoldást a platformon belül, és tedd lehetővé a platformon kívüli meglévő megoldások kapcsolódását is! Ezáltal valósul meg a választás valódi szabadsága.

Flexibilis munkafolyamatok

Mindkét rendszer a Camunda workflow motort választotta. Ez részben a folyamatok lépéseit képes flexibilisen, választhatóan, feltételekhez kötötten összefűzni, másrészt képes feltételek teljesülése esetén folyamatokat automatizáltan végrehajtani. Ehhez természetesen az egyes szoftver alkotóelemeket olyan módon kell megtervezni és kialakítani, hogy a folyamatba ágyazhatók legyenek, azaz, hogy képesek legyenek a vezérlő paramétereket stb. feldolgozni, értelmezni, és azok függvényében eljárni. Minél elemibb funkciók végrehajtására tervezett az egyes szoftver-egység, azaz minél inkább érvényesül a mikroszerviz architektúra, annál áramvonalasabban fűzhetők össze egy workflow-ban a lépések.

Adatmodell, kapcsolt adatok

A szemléletünk központjában a végfelhasználó áll, aki nem egy konkrét tárgyat, nem is feltétlenül a tartalmat, nem meghatározott hordozón és nyelven keres, hanem értelmet, összefüggéseket – és ehhez szokott a Google felületeken, de a közösségi média azonnali reagálású platformjain is. Csakis szemantikus összefüggések érdekesek, minél megbízhatóbb kapcsolatokkal – a könyvtár az a hiteles hely, amely ma ezt a legstabilabban képes biztosítani, még a WIKI-nél is megbízhatóbban. A platform célja az, hogy ne kelljen tudni, hol és milyen módon kereshető valami; melyik könyvtárban, gyűjteményben, stb. – enélkül is célt érjen a keresés.

A FOLIO adatmodellje egy BIBFRAME2-höz közeli belső adatformátum, miközben a platform képes a jelenleg létező fontosabb adatcsere formátumok tárolására is. Jóval messzebb megy az OKP: az adatokat a legkisebb egységben tárolja, gráf-elemként, öt elemi egységből álló állításként; az OKP alapvető adatmodellje a Quintuplet.

Az általános tudásgráf dinamikusan bővíthető értékészlettel rendelkezik. Ebben a modellben a triplet képzésekor az állítmány nem a rekordok közötti kapcsolat minőségeként kerül letárolásra, helyette az állítmány, mint rekord épül be a kapcsolati láncba. A kapcsolatok közös pontja a kijelentés, amely képes egy adott alanyról egy darab elemi állítást megfogalmazni. A kijelentés tárgya lehet egy másik alany, literál érték, 'tétélesített' literál érték.

A „triplet” elemi kijelentések meghatározására szolgál. További specifikus adatok felvételéhez kijelentéseket kell tenni egy kijelentésről. Minden kijelentés egyformán igaz, amíg az adott kijelentésről nem teszünk egy „nem igaz” kijelentést. A kijelentés

„fa” a végtelenségig ágaztatható. Hátránya, hogy a keretrendszer nem ad iránymutatást versengő kijelentések kezelésére.

A triplet hagyományos három eleme:

- Subject: Az alany az a dokumentum, amelyre a kijelentés érvényes.
- Predicate: Az állítmány egy bővíthető értékkeszlettel rendelkező szótári elem, amely a kijelentést tipizálja.
- Object: A tárgy a kijelentés tartalmi része, amely tárolhat literál értéket, mutathat más a rendszerben elérhető entitásra.

A hagyományos gráf modell hátrányainak kiküszöbölésére és az új igények elérése érdekében került sor arra, hogy a fejlesztők a triplet három alapelemét két új elemmel is kibővítették, így született meg a Quintuplet. A két, triplethez képest újdonság a Quintupletben:

- Selector: A kijelentés elhelyezkedése az alany dokumentum típusa által definiált vászon dimenziói mentén.
- Lifecycle: A kijelentés életciklusa hordozza többek között a létrehozás idejét, a létrehozó ágenst, valamint a kijelentés érvényességi idejének kezdetét, végét, a kijelentés bizonyossági besorolását.

Ezzel a bővítéssel olyan régóta meglévő kihívásokra is választ talált a fejlesztő, amelyekre a hagyományosabb adatcsere-formátumok nem adnak kielégítő megoldást. Ilyenek:

- Képes kezelni a változatokat, versengő adatokat - érvényességi időszakra vonatkozhat a kijelentés.
- Adatformátum változás esetén kezeli a korábban érvényes, illetve a mindenkori formátumot is. Az adat módosításának esetén az aktuálisan érvényes formátumot tekinti mérvadónak.
- Flexibilisen definiálható megfeleltetéssel (mapping) bármilyen tetszőleges adatformátum kinyerhető.
- Nincsenek korlátai.



Kijelentések anatómiája: A Quintuplet felépítése



SUBJECT:
AZ ALANY
AZ A
DOKUMENTUM
AMELYRE A
KIJELENTÉS
ÉRVÉNYES



SELECTOR:
A KIJELENTÉS
ELHELYEZKEDÉSE
AZ ALANY
DOKUMENTUM
TÍPUSA ÁLTAL
DEFINIÁLT
VÁSZON
DIMENZIÓI
MENTÉN



PREDICATE:
AZ ÁLLÍTMÁNY EGY
BŐVÍTHETŐ
ÉRTÉKKÉSZLETTEL
RENDELKEZŐ
SZÓTÁRI ELEM
AMELY A
KIJELENTÉST
TIPIZÁLJA



OBJECT:
A TÁRGY A
KIJELENTÉS
TARTALMI RÉSZE,
AMELY TÁROLHAT
LITERÁL ÉRTÉKET,
MUTATHAT MÁS A
RENDSZERBEN
ELÉRHETŐ
ENTITÁSRA



LIFECYCLE:
A KIJELENTÉS
ÉLETCIKLUSA HORDOZZA
TÖBBEK KÖZÖTT A
LÉTREHOZÁS IDEJÉT,
A LÉTREHOZÓ ÁGENST,
VALAMINT A KIJELENTÉS
ÉRVÉNYESSEGI IDEJÉNEK
KEZDETÉT, VÉGÉT,
A KIJELENTÉS
BIZONYOSSÁGI
BESOROLÁSÁT.

| www.folio.org

Forrás / Copyright: HerMészSoft



2. ábra: Kijelentések anatómiája: a Quintuplet felépítése; szerző / forrás: HerMészSoft Kft.

A metaadatok jelentőségét jelzi, hogy számos konferencia foglalkozik ezzel a kérdéssel összetett módon. A 2021-es évből kettőt ajánlok a figyelmükbe, amelyek a mi projektünk számára relevánsak:

Bibliographic Control (besorolási adatok) a digitális ökoszisztémában – A digitális ökoszisztémában történő metaadat-kezelés

<https://www.bc2021.unifi.it>

OCLC Next Generation of Metadata - A metaadatok következő generációja:

<https://www.oclc.org/go/en/events/next-generation-of-metadata.html>

Könyvtárközi kölcsönzés ReShare: egy vízió / egy közösség / egy projekt

A ReShare egy közösség, egy szoftver és fenntartó grémium egyben. Az alkalmazott megoldások terén tükrözi a forrásmegosztás területén történt jelentős átalakulásokat. Képes kommunikálni a legkülönbébb meglévő rendszerekkel és formátumokkal. A szoftver és az adatok egyaránt közösségi tulajdonban vannak. Mindenki számára elérhető, mindenki által használható. A használat és a hozzájárulás mértéke nem mechanisztikusan kapcsolódik össze, a használók képességeik szerint járulnak hozzá a fenntartáshoz.

A középpontban a végfelhasználó áll. A nagy dobás ebben a megoldásban az, hogy a Discovery és kiszolgálás független – nem kell tudni, hol van az objektum, és én helyileg hol vagyok. A felhasználó központú workflow-knak köszönhetően a felhasználónak csak a kérést kell megfogalmaznia, és a rendszer utánajár, azonosítja a kívánt objektum hollétét és a megfelelő fizikai vagy éppen virtuális helyen átadja, majd (adott esetben kérésre egy másik helyszínen) visszaveszi a kölcsönzött anyagot. A megoldás flexibilis együttműködésre kész, rendszereken átívelő. Hálózatokat kapcsol össze, konzorciumok

Lendvay Miklós: A könyvtári együttműködés informatikai támogatása: adatmodell, workflow, rendszerfelépítés: Együttműködési lehetőségek a nyílt forráskódú platformokban: OKP, FOLIO, ReShare

és rendszerek együttműködését valósítja meg, korlátok nélkül kezeli az átfedéseket. Benne létrehozhatók közös gyűjtemények is. Sokfajta formátumot kezel. A technológia protokollok standardok alapján valósulnak meg. A UX design nemcsak a kinézetben, hanem az ergonomikus rendszerfelépítésben fejeződik ki általa. A mögöttes, megvalósuló funkcióból adódik a felület, a látható képernyőfelépítés.

Átalakulás a forrásmegosztás területén

ORSZÁGOS
SZÉCHÉNYI
KÖNYVTÁR



ReShare

Rendszer-központú workflow-k	Felhasználó-központú workflow-k
Monolitikus, vertikális stack-ek	Együttműködésre kész, rendszereken átívelő
Egyedi gyűjtemények megosztása	Közös gyűjteményeket tesz lehetővé
Nyomtatott kiadvány fókuszú	Sokfajta formátum
Egyetlen hálózat	Sok egymást átfedő hálózat
API-k	Standardokra alapozott protokollok
Szolgáltató / szállító tulajdonában levő	Közösség tulajdonában levő

Forrás / Copyright / Courtesy of: Index Data



| www.folio.org

3. ábra: Átalakulás a forrásmegosztás területén – és a ReShare erre adott válaszai; szerző / forrás: Index Data

A FOLIO közösség

Az, hogy maga a szoftvermegoldás közösségi tulajdonban van, és a közösség hosszú távon fenntartható módon fejleszti azt, megvalósítja a fejlesztők és a könyvtárosok szoros munkakapcsolatát. Végre nem szerződéseken, követelménykatalógusokon vagy tenderkiírásokon keresztül kommunikálnak a felek, hanem egy asztalnál ülve egyeztetnek folyamatosan, közös céllal dolgoznak nap mint nap. A könyvtárak számára nagy a választási szabadság: akár saját maguk is üzemeltethetik a rendszert, de több hozzáértő szállító közül is választhatnak, adott esetben részfeladatokat is adhatnak, és több szállító együttműködése is elképzelhető, sőt kívánatos. A szállítók többféle spektrumon nyújtanak szolgáltatást: az implementáció, a hosting és a support területén egyaránt.

A folyamatosan csatlakozó könyvtárak újabb lendületet adnak a fejlesztéseknek és a meglévő rendszerek összekapcsolásának. 2021-ben két olyan nagy hírű könyvtár is a FOLIO bevezetés mellett tette le a voksát, ahol jelentős fejlesztői kapacitás is van, és már eddig is sok olyan jelentős nyílt forráskódú szoftvert állítottak elő, amelyet a világ számos országában ismernek és használnak jelentős könyvtárak, köztük számos nemzeti könyvtár is. A Stanford University az International Image Interoperability Framework



(IIF) közösség egyik alapítója, emellett pl. a Blacklight, a Samvera fejlesztője. A Villanova University a világ szinte minden országában használt VuFind kereső fejlesztője.

Az OKP és a FOLIO fejlesztések egymás által inspirálódva zajlottak az elmúlt években. Olyan modulok jöttek létre ennek során, amelyek egymással összekapcsolhatók, vagy akár egy implementáción belül is párhuzamosan futtathatók.

Olyan platformok jöttek létre, amelyek képesek a leghatékabban módon összekapcsolni, rendelkezésre bocsájtani, fejlődni. XXI. századi megoldás, amely a WIKI platform mintájára kinyithatja a csatlakozó könyvtárak számára az együttműködés régóta vágyott lehetőségeit. A fejlődés minden irányban nyitott, a kreativitás és a növekvő igények inspirálják az újabb és újabb lépéseket. A könyvtárak jövője már nemcsak a FOLIO közösség alapítói által megfogalmazott ideában, hanem a valóságban is nyitott! A felhasználók kiszolgálása és bevonása érdekében egyre több résztvevő ismeri fel az ezáltal megnyíló lehetőségeket, és él a platform adta szinte korlátlan szabadsággal.

Bibliográfia

Az Athenaeum 21 Consulting honlapja: <https://www.athenaeum21.com>

Az Athenaeum 21 Consulting előadása: „A technológia nem a válasz; a „digitális” miért nem a legfontosabb aspektusa a digitális stratégiának”

<https://www.slideshare.net/MHzUX/technology-is-not-the-answer-why-digital-is-not-the-most-important-aspect-of-your-digital-strategy>

Rudolf Steiner hármas tagozódás:

<http://www.camphill.hu/wp-content/uploads/Chris-Shaefer-A-hármas-tagozódásról.pdf>

https://de.wikipedia.org/wiki/Soziale_Dreigliederung

<https://www.dreigliederung.de/gliederung>

A FOLIO közösség honlapja: <https://www.folio.org>

A FOLIO vezetés:

<https://wiki.folio.org/display/COMMUNITY/FOLIO+Governance+Model>

A FOLIO víziója:

<https://wiki.folio.org/display/COMMUNITY/>

[FOLIO+Vision%2C+Strategic+Objectives+and+Initiatives](https://wiki.folio.org/display/COMMUNITY/FOLIO+Vision%2C+Strategic+Objectives+and+Initiatives)

A FOLIO release-ek és fejlesztési tervek: <https://wiki.folio.org/display/PC/FOLIO+Roadmap+Process>

A FOLIO roadmap: <https://wiki.folio.org/display/PC/FOLIO+Roadmap>

Az Országos Könyvtári Platform honlapja: <http://okp.oszk.hu>

A ReShare információs honlapja: <https://projectreshare.org>

A projekt technikai vezetőjének, az Index Data-nak az információi a ReShare-ről:

<https://www.indexdata.com/reshare/>

A BTK Zenetudományi Intézet digitális archívum koncepciója az oktatás és a tudomány szolgálatában

Dr. Bolya Mátyás
ELKH BTK Zenetudományi Intézet (Budapest)
tudományos munkatárs
bolya.matyas@abtk.hu
ORCID: [0000-0002-6145-663X](https://orcid.org/0000-0002-6145-663X)

The Digital Archive Concept of RCH Institute for Musicology (Budapest)
at the Service of Education and Science

Mátyás Bolya, DLA
research fellow (RCH Institute for Musicology)
associate professor (Liszt Ferenc Academy of Music)

The article introduces a unique Hungarian folklore database. Its design and operation were implemented within the framework of the new digital archive conception of the Institute for Musicology. This concept seeks to keep pace with the development of digital technology. I will briefly present aspects along which our strategy was developed, assessing how the database can be utilized in the fields of education and research.

Hungarian folk music research, marked by the names of Bartók and Kodály, has a very strong tradition. Huge amounts of valuable material have been accumulated. Effects of this tradition are strongly felt even today. The connection between the archival background, folk music education and the performing arts is very strong. We are currently converting the heritage of our greatest researchers into digital knowledge. Results of this approach effect many areas of culture, creating an extremely complex system embedded in the communication space. Thorough knowledge of the context is essential to be able to design effective online systems.

Keywords: digital archives, folklore database, professional fields, folk music, education, ethnomusicology, crowdsourcing

1. Bevezetés

A Zenetudományi Intézet a zenetudomány különböző területein folyó kutatásokat fogja össze.¹ Egy ilyen összetett szerkezetű és sokféle kutatói hagyománnyal rendelkező kutatóintézet² átfogó online publikációs koncepciójának kidolgozása és megvalósítása csak lépésről-lépésre, nagyobb időtávban lehetséges. Írásomban a Népzene- és

1 Ezt az intézet osztályszerkezete is jól mutatja: *A Zenetudományi Intézet honlapja*, 2021, hozzáférés: 2021.05.31, www.zti.hu.

2 A Zenetudományi Intézet a kutatóhelyek mellett archívumokat is fenntart, amelyek általában az osztályokhoz kapcsolódóan működnek (20–21. Századi Magyar Zenei Archívum, Bartók Archívum, Népzene- és Néptánckutató Osztály és Archívum).



Néptánckutató Osztály és Archívum Népzenei Archívumában végzett digitális átállítás, majd az ennek eredményeképpen létrejött online hangarchívum³ felhasználási lehetőségeit és hatásait vizsgálom a kultúra különböző szegmenseiben – kiemelt figyelemmel az oktatás és a tudomány területére.

A koncepció igyekszik lépést tartani a technika fejlődésével és kihasználni a legmodernebb technikai lehetőségeket úgy, hogy közben tiszteletben tartsa a múlt tudományos eredményeit és publikációs szokásait is. Ez nem könnyű feladat. Miközben a technika fejlődése hihetetlen tempót diktál, a szűkös anyagi lehetőségek, illetve a hagyományos publikációs gyakorlat berögzülései lassítják a felzárkózási folyamatot. Míg akár néhány évvel ezelőtt a papíralapú publikáció volt az egyetlen elfogadott publikációs platform, ma már a tudományos kommunikáció egyre inkább az online térbe helyeződik át. Teret hódít a digitális bölcsészet is, amely nem csupán azt jelenti, hogy virtuózan alkalmazzuk a digitális technológia adta lehetőségeket a bölcsészettudomány területén, hanem felismerjük, hogy e lehetőségek visszahatnak a kutatás módszertanára, alapvetően megváltoztatva a kérdésfeltevéseket és a felhasznált források körét.

2. A digitális archívum koncepció elemei

Az analóg és a digitális világ határán egyensúlyozó archívumok szakembereinek számos, eddig ismeretlen kihívással kell szembenézniük. Az analóg gyűjteményi részek digitalizációját követően felépített adatbázisok szerepe mára teljesen átalakult és kiszélesedett; ezek az archívumok a digitális térben multifunkcionális eszközökké válnak.

Röviden – kezdve az 1990 előtti analóg feldolgozás folyamatától – bemutatom azokat a szempontrendszerket, amelyek segítségével igyekeztünk kialakítani stratégiánkat. A Népzenei Archívum működésén keresztül világossá válnak a digitális archívum koncepció legfontosabb elemei: az archiválás és a szolgáltatás szétválasztása; a tervezett adatfeldolgozás; a gyűjteményi egységek szuperadatbázisokban definiált részhalmazként való értelmezése; végül a megfelelő publikációs platform megválasztása és a célcsoportok felmérése.

2.1. Az analóg feldolgozási lánc⁴

A folklóresemény – esetünkben zömmel zenefolklórról van szó – rögzítése a gyűjtés során történt. Az így előállt adatokat a szakemberek az előrendezést, a szakmai dokumentációt és az archívumi adminisztrációt követően gyűjteményi egységekbe rendezték. Ezek jelentik ma a népzene kutatás számára az elsődleges és legértékesebb forrásanyagot. 1990 körül vált elérhetővé a digitalizálási technika, amely a kor színvonalán biztosította a gyűjtemény digitális kópiáit. E két forrás összegzésével, tárolás és kereshetőség biztosításával lezárult a klasszikus archiválási folyamat, és létrejött az analóg és digitális egységekből álló archívum. Ez a feldolgozási sor egészült ki egy újabb kimenettel a

3 Dinamikusán fejlődő adatbázisaink zászlóshajója az Arcanum Adatbázis Kft. partneri együttműködésével készült ZTI Hangarchívum: Bolya Mátyás, szerk., *Zenetudományi Intézet Hangarchívum*, 2021

4 A folyamat az 1. ábrán követhető.

2000-es évek elején; ezek voltak az adatbázisépítés első szárnypróbálgatásai.⁵ Statikus felületek, amelyek a mai értelemben nem tekinthetők adatbázisnak, azonban fontos lépésnek számítottak az analóg-digitális tudáskonverzió, illetve a szolgáltatási funkció problémáinak megértése terén.

2.2. Archiválás és szolgáltatás

A valódi adatbázismotorok megjelenése – immár az analóg feldolgozási láncsal egyenértékű – korszerű digitális feldolgozási lánc kialakítását tették szükségessé. Megjelent a digitális archívum koncepció legfontosabb eleme, az *archiválás* és a *szolgáltatás* koncepcionális szétválasztása. Az előbbi a megőrzést helyezi a fókuszba, utóbbi pedig az adatelérést. A szolgáltatáshoz optimalizált fájlformátumok és fájlformátumok kellenek, valamint megfelelő – vagyis fenntartható, fejleszthető, korlátlan elérésű – szoftverkörnyezet.

2.3. Adatfeldolgozás

A második elem az analóg adatok feldolgozásának megszervezése. Meghatároztuk a különböző feldolgozási fázisok mélységét és valós szakemberigényét. Esetünkben a folklórkutatók mellé adatbázisokkal foglalkozó szakembereket vontunk be a munkába már a tervezés fázisában. Azokat a munkafolyamatokat, amelyeket egy általános adatfeldolgozó is el tud végezni, elválasztottuk azoktól, amelyek egyértelműen kutatói feladatnak minősülnek.⁶

A gyűjteményi egységek digitalizálását és rendezését digitális utómunka követi, amely – feljavítva a jel-zaj viszonyt – jelentősen megnöveli az alapanyagból nyerhető információmennyiséget. A leíró adatstruktúra kialakítása után, a rendelkezésre álló dokumentáció alapján az adatok pontosításával és kereszthivatkozások létrehozásával gyűjtésrekonstrukció készül. Az így létrejött metaadat-struktúra a javított hang- és vizuális fájlokkal együtt feltöltve az intézet által fejlesztett szoftverkörnyezetbe olyan online publikációs felületen jelenik meg, amely összehangolt kutatói feladatok kiszolgálására is alkalmas.

2.4. Szuperadatbázis

Az informatikai rendszer fejlesztésénél fontos szempont volt, hogy az a vizsgált kulturális jelenséget minél több szempontból képes legyen rögzíteni, egyúttal lehetőséget adjon a tudományos rendszerezésre is. Itt kerül előtérbe a koncepció harmadik – már a tervezés fázisában kiemelt fontosságú – eleme, a publikálandó gyűjteményi egységek belső összefüggéseinek felismerése és megjelenítése. A mai tudomány szemlélet ugyanis nem adatokat közöl, hanem kulturális kontextusokat dekódol, értelmez és publikál. Ezt a nagyon fontos alapelvet csak superadatbázisok építésével lehet megvalósítani.

5 Két kiragadott, ma is elérhető példa: Fügedi János, szerk., *TáncFA: Tánc Film Adatbázis*, 2001; Németh István, szerk., *Publikált népzenei felvételek internetes adatbázisa*, 2003

6 Az általános adatfeldolgozó és a kutatói erőforrásigény aránya 90/10%-ra becsülhető.



Tehát a korábban különálló, egymással nem kommunikáló, zárt adatbázisok helyébe szuperadatbázisokban definiált részhalmazok lépnek.⁷

2.5. Célcsoport meghatározása

Egyre nagyobb teret nyer az *open access* szemlélet, a társadalom számára pedig fontossá vált, hogy a közpénzből finanszírozott tudományos munka transzparens legyen. Az érdeklődők rétege jelentősen kiszélesedett és elvesztette homogenitását.⁸ Ezért lett fontos elem a célcsoport meghatározása, keresési preferenciáik felmérése, majd ezek ismeretében az adatbázis optimalizálása, a publikált anyagban várható hibaszázalék meghatározása és a fejlesztési fázisok megtervezése. Ez a szempont hangsúlyosan jelenik meg a keresőmotorok algoritmusának és felhasználói felületének tervezésében.

Az informatikai eszközök fejlődésével új adatgyűjtési modellek jelentek meg, például a *crowdsourcing*. Ez lehet passzív⁹ és aktív, amely a felhasználói magatartástól vagy a szoftveres környezet kialakításától függ. A kulturális adatbázisok között számos példát találunk erre az adatgyűjtő technikára bel- és külföldön egyaránt: ilyen az Hungaricana Közgyűjteményi Portál Képeslapgyűjteménye,¹⁰ a Fortepan online fotóarchívuma,¹¹ vagy éppen a New York Public Library étlapgyűjteménye.¹² A zene területén sem ismeretlen.¹³ A ZTI Hangarchívum vágatlan helyszíni felvételeinek időködolása ebben a modellben zajlik.

3. A digitális platform szakterületi alkalmazása¹⁴

A digitális archívum tervezésének és elkészítésének megismerése után vizsgáljuk meg, hogyan hasznosulnak ezek az eredmények! Mire képes ma egy digitális platform? Milyen szakterületeken jelenik meg és ott milyen hatást gyakorol? Nyilvánvaló, hogy pontos felméréshez és a kulturális beágyazódás mértékének meghatározásához évekre lesz szükség, de a visszajelzések alapján biztosak lehetünk abban, hogy hasznos folyamatok indultak el.

A forrásanyag alapvetően a kultúra három nagy területén jelent meg, ezek a tudomány, az oktatás és az előadó-művészet. A 2. ábra középső része jelképezi a digitális platformot,

7 Jó példa erre a Népzenei Archívum etiópai gyűjteménye: Bolya Mátyás, et al., *Ethiofolk: Online etióp népzene- és néptáncarchívum*, 2019.

8 Már nem csak néhány szakértő kíváncsi egymás véleményére, komoly igényeket támasztanak például a felsőoktatásban tanuló diákok, a múzeumpedagógusok, az interdiszciplináris beállítottságú gondolkodók stb.

9 Ilyen a közismert navigációs alkalmazás is, a Waze. A felhasználó a telepítéssel beleegyezik, hogy az általa generált adatokhoz (tartózkodási hely, útvonalak, forgalmi adatok stb.) a szoftvert üzemeltető cég hozzáfér és feldolgozás után a közösség számára használható formában elérhetővé teszi.

10 *Képcsarnok: Képeslap*, 2011–2021.

11 *Fortepan*, 2010–2021.

12 Rebecca Federman et al, 2011.

13 Bozena Kostek, *Journal of the Acoustical Society of America*, 2019.

14 Az alkalmazási területeket a 2. ábra szemlélteti.

amely két fő részből áll: a mindenki számára elérhető *frontend*-ből – itt kapcsolódnak az oktatás és az előadó-művészet szereplői – és a csak kutatók számára publikus *backend*-ből.

3.1. Tudomány

Ma a forrásanyag gyors elérése, tízezres nagyságrendű adatelemek rendezése, valamint a teljes forráslánc láthatósága, ellenőrizhetősége alapvető kutatói elvárás. Az elmúlt évtizedek digitális forradalma hatalmas mennyiségű anyagot tett elérhetővé, így a harmadik évezred kutatói számára a források elérése nem csupán jogosultság kérdése, hanem egyben versenyfutás az idővel. Felértékelődtek a gyorsan kereshető, jól strukturált, permanens linkekkel hivatkozható, hiteles online tartalmak.

A platform a tudomány területén sokfunkciós eszközzé válik. Archivális szempontból a leltárt, a feldolgozást, az adminisztrációt és a kutatószolgálatot segítő felület, kutatói szempontból pedig publikációs lehetőség és egyúttal a publikálást megelőző kutatást támogató digitális környezet.

Nagy gyűjtemények belső összefüggéseinek feltárása digitális gyűjteménykezelés nélkül lehetetlen. Az adatok áttekintése után következik a klasszikus kutatói munka, vagyis a kulturális kontextus dekódolása, mintázatok keresése, rendszeralkotás. Ebben a környezetben olyan belső összefüggések tárulnak fel, amelyek a kétdimenziós táblázatokban láthatatlanok maradnak és rendkívül inspirálóak a kutatók számára.

3.2. Oktatás

A magyar népzene kutatás nagy hagyományokkal rendelkezik. A Bartók és Kodály nevével fémjelzett időszakban megszületett modern tudományág eredményei máig éreztetik hatásukat. A népzeneoktatás és az archivális háttér közötti kapcsolat az online platformok megjelenésével erősebbé vált. A korábban csak kutatók számára hozzáférhető gyűjtések ma már széles körben ismertek. Számos olyan funkció található – vagy fejleszthető a jövőben – amely segíti a tanítást és az e-tananyagok fejlesztését.

Jó példa erre a Folk_ME projekt¹⁵ – digitális tananyagfejlesztés archívumi támogatással –, amely számos egyedi fejlesztéssel, nemzetközi pályázat keretében, finn és ukrán partnerekkel valósul meg.¹⁶ Bár a hivatalos bemutató csak 2021 októberében várható, a kezdeményezésre a magyar kultúra közszereplői is figyeltek, így a szoftver keretrendszerének és fejlesztési tapasztalatainak felhasználásával magyar népzenei felvételsorozat indult a Zenetudományi Intézet, az LFZE Népzene Tanszék és a Hagyományok Háza közös kezdeményezéseként.

¹⁵ Bolya Mátyás, Both Miklós et al, *Folk_ME. Folk Music Education for Future Generation*, 2021.

¹⁶ Kottákat is kezelő soksávú lejátszó, lejegyzések komplex audiovizuális támogatással, lassítható videók, professzionális kép és hangminőség stb.



3.3. Előadó-művészet

Egyre erősebb társadalmi elvárás, hogy az értékmentés és értékközvetítés ne csak a kultúra egy szűk szegmensében jelenjen meg. A művészképzésen keresztül az online publikációk közvetlenül hatnak az előadóművészetre is. A művészek inspirációs forrásként tekintenek ezekre az anyagokra, miközben látens módon formálódik az értékrend és művészi ízlés is. Mindez hatással van a zeneipar további szereplőire is. Ma már producerek, mérvadó kulturális intézmények vezetői látják úgy, hogy sajátosan európai értékek jelenhetnek meg ezáltal a színpadokon, új és friss ötletekkel inspirálva a gyakran idegen mintákra épülő produkciónkat.

3.4. Kommunikációs tér

Az eddig tárgyalt szakterületek – kiegészülve a civil szféra erősödő jelenlétével – egy közös kommunikációs térbe ágyazódnak be. A 21. század marketingzajában a sikeres értékközvetítés elengedhetetlen feltétele a láthatóság, amely e tér csatornáinak megfelelő paraméterezésével valósítható meg.

4. Zárszó

A bemutatott stratégia elemeiben letisztult, de várhatóan a gyakorlati megvalósítás során részleteiben változhat. Az újabb és újabb technológiák új utakat nyithatnak meg, és a jogi háttér formálódása is hatással lehet a fejlődés irányára. Mivel úttörő kezdeményezésről van szó, természetesen szakmai viták kísérik a folyamatot, de az elért eredmények és a pozitív visszajelzések – a hasonló nemzetközi trendek mellett – egyértelműen jelzik az irány helyességét.

4.1. Eredmények

A Hangarchívum jelenleg 200 000 rekordot tartalmaz, körülbelül 14 000 óra hanganyaggal.¹⁷ A vágatlanul közzétett hangfelvételekhez metaadatokkal ellátott dallamszegmensek kapcsolódnak. A metaadatok tartalmazzák a dallamok kezdőpontjának idő kódját, de emellett lehetőség van – a youtube-hoz hasonlóan – a felvételek tetszőleges pontjára hivatkozni.¹⁸ A metaadatok másik fontos eleme a helyek geokoordinátája,¹⁹ amelynek segítségével térképen is ábrázolható a gyűjtemény vagy a keresési eredmény. Az adatbázis keresztkapcsolatokkal köti össze a teljes gyűjtési dokumentációt, így elérhetők a dallamok lejegyzései,²⁰ a kutatói használatra készült egyedi mikrobarázdás

17 Ez zömmel a magnószalag-gyűjtemény népzenei vonatkozású anyagát jelenti. 2019 elején ez kiegészült a Kodály-rend teljes anyagával, beleértve a támlapokat és a hangfelvételeket is.

18 Ilyenkor a hangfájl kezdete a referenciapont, ehhez képest határozzuk meg az adott hivatkozási pont távolságát másodpercben.

19 Megkülönböztetjük a gyűjtés helyét az adott folklóradat érvényességi területétől (vagyis a használat helyétől).

20 *Népzenei hangfelvételek lejegyzései: ZTIAP 501-18740, 2019-2021*). Korábban a magnetofonszalagok hanglemez-átjátszásairól készült lejegyzéseket tartalmazta a lemezátjátszások sorrendjében tárolva, később ez kiegészült az újabb hordozókról készült lejegyzésekkel (magnókazetta, DAT-kazetta, MiniDisc stb.). 180 000 oldal.

lemezek adatlapjai,²¹ illetve a gyűjtési jegyzőkönyvek is.²² Az adatbázishoz különböző jogosultságokkal lehet hozzáférni; a fejlesztés negyedik ütemében alakítottuk ki a kutatói felületet.²³ A Hangarchívum bármely elemére permanens és egyedi linkkel lehet hivatkozni.

4.2. Tervek

A közeljövő tervei között a platform fejlesztése mellett további gyűjteményi egységek vagy kutatási eredmények digitális elérhetősége szerepel. 2021-ben megújul a Néptánc Tudástár²⁴ és publikálni szeretnénk az intézet népzenei és néptánc témájú fényképgyűjteményét is. A platform legfontosabb fejlesztési irányai a következők:

- digitális azonosítási lehetőség az adathierarchia bármely pontján (DOI, Handle, ARK stb.),²⁵
- tetszőleges részadatbázis definiálása és külön felületen való publikálása, API-kommunikáció kiépítése,
- további gyűjteményi egységek integrálása (fotók, filmek stb.),
- az adminisztrációs felületről elérhető menedzsmenteszközök és projektkövetés,
- MI-alkalmazás a gyűjteményrendezés és a kutatástámogatás területén,
- integráció nagyobb – hazai és nemzetközi – adatbázis-rendszerekbe.

21 *Kutatói hangfelvétel-másolatok (AP-lemezek) borítói, 2019–2021.* Az AP az MTA BTK ZTI Népzenei Archívumában található egyedi gyártású hanglemezek azonosítója, az Akadémiai Pyral rövidítése. A kutatók e hanglemezeket használták a felvételek lejegyzéséhez, publikációikban pedig rendszerint az AP-számot adták meg forrásként. 18 000 borító.

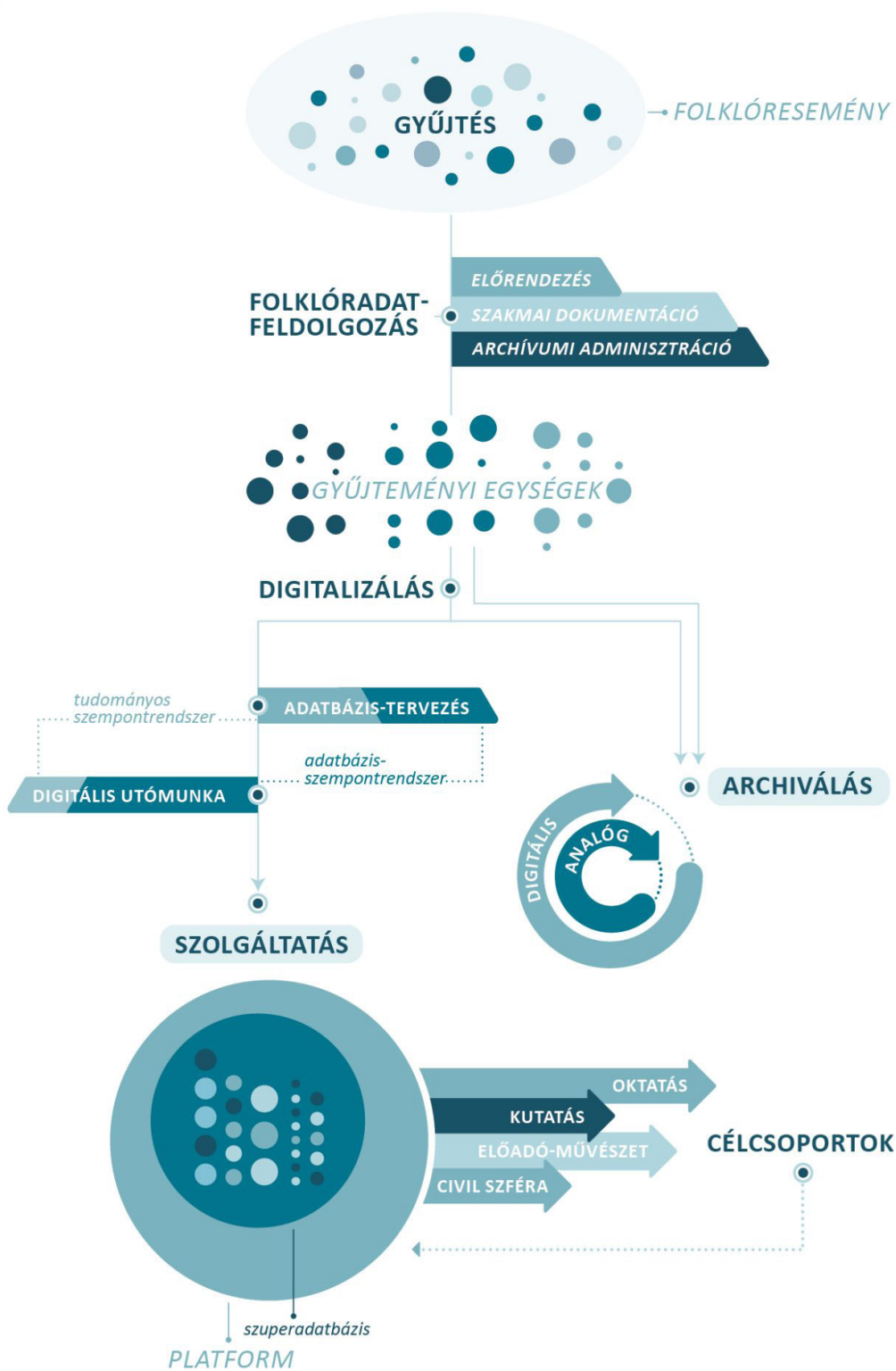
22 Csak kutatói jogosultsággal elérhető tartalom.

23 Ez a jogosultság hozzáférést biztosít a jegyzőkönyvekhez, a valamilyen okból nem publikus felvételekhez stb.

24 Fügedi János, *Néptánc Tudástár, 2007–2021.*

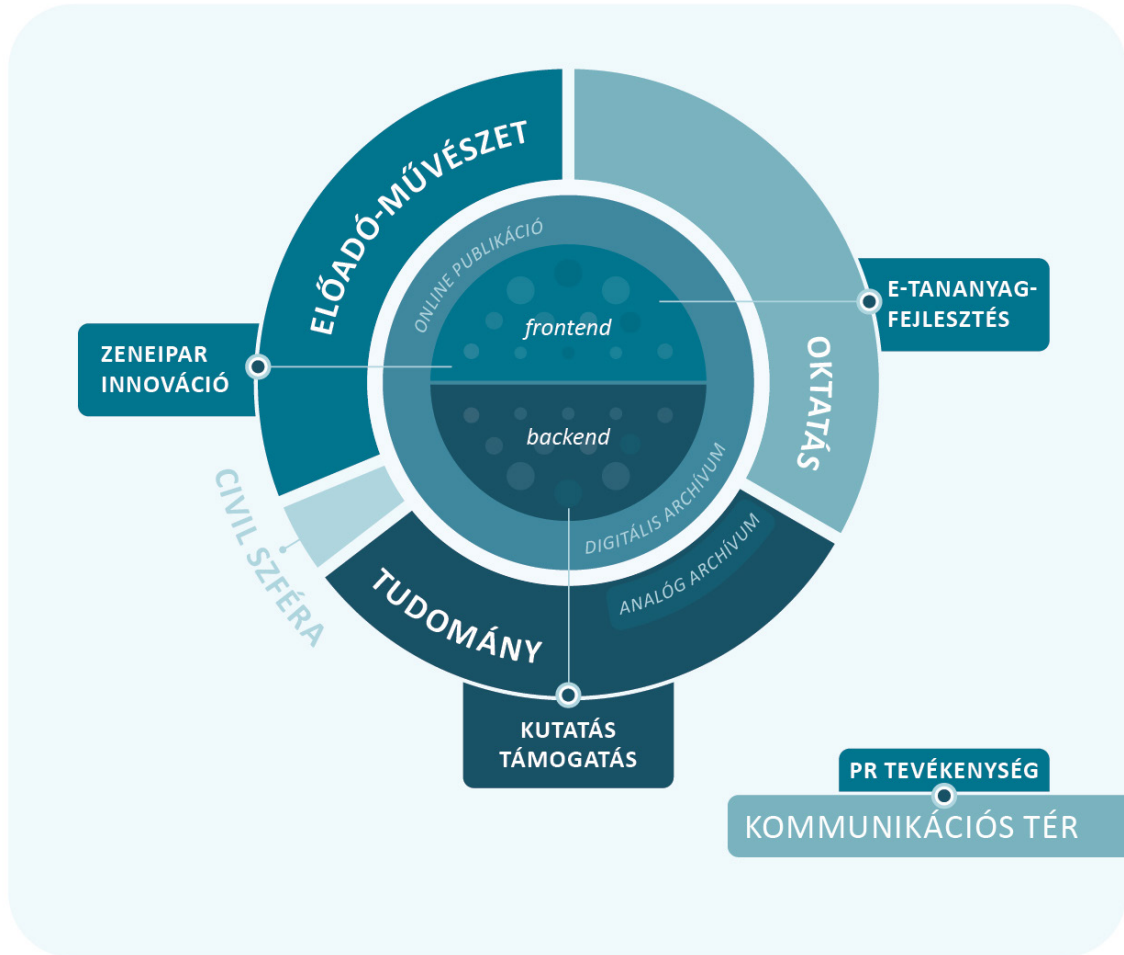
25 A közgyűjteményekben használható perzisztens azonosítók összefoglalását lásd: Lukas Coster, *Code4Lib Journal* Issue 47, 2020.

DIGITÁLIS ARCHÍVUM KONCEPCIÓ



1. ábra. A digitális archívum koncepció és kontextusa

SZAKTERÜLETEK



2. ábra. A digitális platform szakterületi megjelenése



Bibliográfia

- A Zenetudományi Intézet honlapja. 2021. Hozzáférés: 2021.05.31, www.zti.hu.
- Bolya Mátyás, Both Miklós és Kukár Manó. *Ethiofolk: Online etióp népzene- és néptáncarchívum*. MTA BTK Zenetudományi Intézet / Polyphony Project, 2019. Hozzáférés: 2021.05.31. www.ethiofolk.com.
- Bolya Mátyás, Both Miklós et al. *Folk_ME: Folk Music Education for Future Generation*. Polyphony Project, 2021. Hozzáférés: 2021.05.31. www.folk-me.com.
- Bolya Mátyás, szerk. *Zenetudományi Intézet Hangarchívum*. ELKH BTK Zenetudományi Intézet, 2021. Hozzáférés: 2021.05.31. <https://zti.hungaricana.hu/>.
- Coster, Lukas. „Persistent Identifiers for Heritage Objects.” *Code4Lib Journal Issue 47*, 2020. Hozzáférés: 2021.05.31. <https://journal.code4lib.org/articles/14978>.
- Federman, Rebecca et al. *What's on the menu?* New York Public Library, 2011. Hozzáférés: 2021.05.31. <http://menus.nypl.org/about>.
- Fortepan. 2010–2021. Hozzáférés: 2021.05.31. <https://fortepan.hu/>.
- Fügedi János, szerk. *TáncFA: Tánc Film Adatbázis*. MTA Zenetudományi Intézet, 2001. Hozzáférés: 2021.05.31. <http://www.zti.hu/tanc/tancfa.htm>.
- Fügedi János. *Néptánc Tudástár*. ELKH BTK Zenetudományi Intézet, 2007–2021. Hozzáférés: 2021.05.31. DOI: <https://doi.org/10.23714/nzntk.ntt.ind.hu>. http://db.zti.hu/neptanc_tudastar/.
- Képcsarnok: *Képeslap*. Hungaricana Közgyűjteményi Portál, 2011–2020. Hozzáférés: 2021.05.31. <https://gallery.hungaricana.hu/hu/search/results/?list=eyJxdWVyeSI6ICJUSVBUz0oS1x1MDBIOXBlc2xhcCkifQ>.
- Kostek, Bozena. „Music information retrieval: The impact of technology, crowdsourcing, big data, and the cloud in art.” *Journal of the Acoustical Society of America* 146 (2019): 2946. Hozzáférés: 2021.05.31, <https://asa.scitation.org/doi/10.1121/1.5137234>.
- Kutatói hangfelvétel-másolatok (AP-lemezek) borítói*. BTK Zenetudományi Intézet, 2019–2021. Hozzáférés: 2021.05.31. https://library.hungaricana.hu/hu/collection/mta_btk_ZTI_APboritok/.
- Németh István, szerk. *Publikált népzenei felvételek internetes adatbázisa*. MTA Zenetudományi Intézet, 2003. Hozzáférés: 2021.05.31. <http://db.zti.hu/24ora/dalok.asp>.
- Népzenei hangfelvételek lejegyzései: ZTI AP 501–18740*. BTK Zenetudományi Intézet, 2019–2021. Hozzáférés: 2021.05.31. https://library.hungaricana.hu/hu/collection/mta_btk_zti_AP/.

Tömeges könyvtári felhasználóképzés távoktatási formában
az SZTE Klebelsberg Könyvtár e-learning platformján¹

Várnai-Vígh Adrienn E.
Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára (Szeged)
adrienn.varnai-vigh@ek.szte.hu

Kokas Károly
Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára (Szeged)
karoly.kokas@ek.szte.hu

MOOCs (Massive Open Online Courses) are one of today's popular and effective educational solutions. This presentation connects directly to another, held by our colleagues. Last spring, we have managed to create a complete e-learning package with our team at the SZTE Klebelsberg Library in order to meet the unprecedented challenges during the pandemic. We have more than 10 years of experience from organizing the Introduction to library and information science compulsory course for the students of the Faculty of Humanities and Social Science, which means about 5-600 students per semester. In our presentation, we will share insights about our experiences managing the e-curriculum and the system of tools behind it: starting with the platform and the database of the exam questions, followed by the communication with the students and ending with the grading method. We will also address the initial extreme difficulties, emerging requests and evaluate the first year's experiences as well.

Keywords: MOOC, user experience, library user training, e-learning package

¹ Az előadás prezentációja itt megtekinthető: <https://prezi.com/view/VISEiUc9RI24UCGxVcpC/>

Központ kongresszusi termében (olykor 750 fős létszámmal!). Szakértelem alapján váltakozó tanári gárda oktatott, a hallgatók tananyagként a honlapra feltöltött részletes prezentációkat kaptak. A kezdő évektől eltekintve hamar átálltunk arra, hogy saját fejlesztésű elektronikus vizsgáztató rendszerünkkel (EVR) kérjük számon az anyagot: ez egy kérdésadatbank felépítését jelentette, amiből a gép generálta az időzített kérdőívet és végezte el a pontozást is.

1.2. A „tudományos írás” és a kurzus

A képzés kezdeteitől olyan szándék vezetett minket, hogy a szükséges műveltségi elemeket (pl. könyvtárak típusai, történeti áttekintés stb.) keverjük a könyvtári informatikai elmélet alapjaival (fogalmi tisztázás) és ehhez tegyünk gyakorlati tudnivalókat. Filozófiánk az volt és az ma is, hogy egy „szerszámosládát” adjunk át a kezdő bölcsészeknek, amit már tanulmányaik elejétől használhatnak, akár a beadandó munkákat, a referátumokat, vagy a BA, majd MA szakdolgozatokat is figyelembe véve.

Az utóbbi időben szerencsésen találkozott követett irányvonalunk a téma egyes aspektusait kézikönyvben (tankönyvben?)



2. ábra: Gyurgyák János könyvének borítója

feldolgozó Gyurgyák János történész törekvéseivel, akinek születő könyvének munkálataiban közre is működünk, ahhoz számtalan könyvtárosi javaslatot tettünk. *A tudományos írás alapjai*³ címmel megszületett munka és maga a születés folyamata is igen jótékony hatással volt a kurzusra, mert az oktatást is feldúsítottuk – a nálunk is ajánlott könyvként szereplő mű – főbb mondanivalóival, praktikus tanácsaival. (A kitűnő és korszerű könyv pedig a szegedi gyakorlatból vett át sok elemet, főként adatbáziskezelési, katalógushasználati és általában könyvtári informatikai megfontolásokat, ill. a szakirodalmi keresés módszertanát illető megoldásokat.) A gyakorlatban – úgy gondoljuk – hogy e könyv tanulmányozása és a kurzusunk látogatása, kiegészítve a könyvtár és szolgáltatásai valós kipróbálásával, együtt ad jó és megbízható kulcsot a hallgatói lét elején lévő fiataloknak.

2.1. Az oktatás keretei

2006 óta oktatjuk tehát az SZTE BTK hallgatóit modern könyvtárhasználati ismeretekre. Ezalatt változott az oktatás helyszíne, a hallgatói létszám, a tananyag, az előadók és előadások. Az oktatás jelenlegi kereteinek értelmezésekor ezért összehasonlításokat is kell tenni.

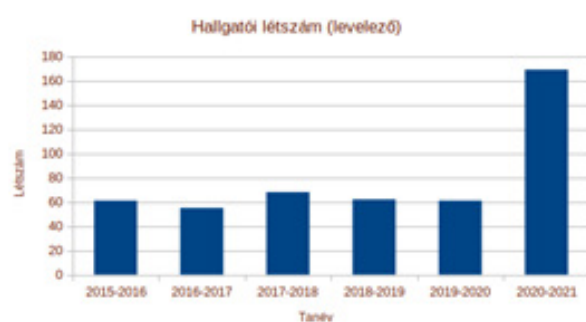
³ Gyurgyák János. *A tudományos írás alapjai*. Budapest: Osiris, 2019



A Könyvtár- és informatikai alapismeretek nevű tárgy számos képzés mintatantervének része; nappali tagozaton az őszi, levelezőn a tavaszi félévben vehető föl. A nappali tagozatos oktatás helyszíne kezdetben a Tanulmányi és Információs Központ Kongresszusi terme volt, mely 700 fő befogadására alkalmas. A félév során az előadáson résztvevők létszáma azonban folyamatosan csökkent, így kihasználatlan volt egy ekkora tér. Újabb létszámcsökkenést eredményezett, amikor e tárgy több szakon kikerült a mintatantervből. Így ekkortól egy 200 fő befogadására alkalmas teremben tartottuk az órákat, 350-400 hallgató részére. Ez különösen a félév elején okozott problémákat, amikor a nagy jelenléti aktivitás miatt nem fértek be ennyien. Ezt a problémát megoldja a kibertérbe helyezett, önálló tanulással feldolgozható tananyag, mert létszámkorlát nélkül lehetséges jelentkezni a kurzusra.



3. ábra: Hallgatói létszám alakulása nappali tagozaton (2015–2021)



4. ábra: Hallgatói létszám alakulása levelező tagozaton (2015–2021)

A távoktatási jellegű kurzus vonzóbbnak is látszik a tantermi megoldásnál, mert így elkerülhető pl. az óraütközés, valamint mindenki saját tempójában haladhat a tananyaggal. Az utóbbi hat tanév létszámdinamikáján ez látszik is: 2020/2021 I. félévében több, mint 500-an vették föl a kurzust. A nappali tagozattól némileg eltér a levelező, ám a létszámok itt is hasonló tendenciát mutatnak.

Oktatási helyzet	Tananyag	Oktatási forma	Vizsgamódszer
Normál	ETA-ban archivált + előadás	Tantermi	Offline számítógépes teszt (Elektronikus Vizsgáztató Rendszer)
Pandémias 1.	ETA-ban archivált + kiegészítő tananyagcsomagok	Önálló tanulás, online támogatással	Beadandó feladatok
Pandémias 2. / távoktatás (jövő)	E-tananyag + ETA-ban archivált forma	Önálló tanulás, online támogatással	Online teszt (CooSpace)

5. ábra: Tananyag, oktatási forma és vizsgamódszer összehasonlítása

A következőkben összevetjük az oktatás, tananyag és vizsga formáit a különböző oktatási helyzetekben. Normál oktatási helyzetben tantermi oktatás zajlott: az Elektronikus Tananyag Archívumban (ETA) elérhetővé tett tananyagot kiegészítették az előadások,

s a kettő együtt volt szükséges a sikeres vizsgálóhoz, melyet egy elektronikus vizsgáztató rendszerben bonyolítottunk le a könyvtár számítógépein.

Az első pandémiás helyzet a levelezősöket érintette. Az előző évek gyakorlata szerint, egy konzultációt tartottunk volna, ám ez most elmaradt, s teljes egészében önálló tanulásra álltak át. Az elérhető tananyagokhoz szükséges tantermi magyarázat pótlására pedig kiegészítő tananyagcsomagokat hoztunk létre. Jelenléti vizsgáztatásra nem volt mód, s mivel a létszám alacsony volt, ezért beadandó feladatokkal teljesíthették a kurzust.

2020 őszére jelentős része elkészült a bevezetni kívánt e-learning tananyagunknak, így ez lehetővé tette az újabb pandémiás helyzetben a távoktatásra történő teljes átállást. A tananyag saját e-learning keretrendszerünkben érhető el, emellett pedig továbbra is archiváljuk azokat.⁴ A nappalis hallgatók nagy létszáma nem teszi lehetővé a beadandó feladatokkal való teljesítést, ezért a korábbi offline teszt alapján online vizsgatesztet vezettünk be.



Oktatási helyzet	Előadók száma	Előadásforma	Előadásminőség	Tananyag-formátum
Normál	5<	Szolgáltatás-bemutató, prezentáció	Változó, nem korrigálható	PDF, PPT, prezi
Pandémia 1.	-	-	-	PDF, PPT, prezi
Pandémia 2. / távoktatás (jövő)	5-10	Rögzített, e-learning	Egységesíthető, korrigálható	Olvásó- és videóleckék

6. ábra: Az előadás- és tananyag-formátumok különbségei

A tantermi oktatás során általában ötnél több fő vett részt az oktatási tevékenységben. Az előadások nagy része valós idejű szolgáltatásbemutató volt, kisebb részben prezentáció; változó minőségben hangzottak el, utólagos korrigálásukra nem volt mód. Az archivált tananyagok többsége ppt, pdf vagy prezi. Az első pandémiás helyzetben egyáltalán nem volt oktatás, a hallgatók a tananyagra és a kiegészítő segédletekre voltak utalva.

Az e-learning tananyagokat 5-10 fő részvételével készítettük el. Egy-egy tananyag elkészítésére több idő fordítható, többször ellenőrizhető, korrigálható. Így sokkal egységesebb minőségben kerülhetnek a hallgatók elé. Két formát használtunk: a szöveges, elméletibb olvasóleckét és a szolgáltatást bemutató, gyakorlati videóleckét.

A tantermi oktatás során az előadó bemutat egy szolgáltatást, melyet a hallgatók offline módon követnek, azaz egy kivetítőn nézik. Az elsajátított ismereteket offline számítógépes teszttel kértük számon, melyet kezdetben csak a tananyag elméleti

4 Az e-learning keretrendszer és a tananyag-archívum elkészültét az EFOP-3.4.3-16-2016-00014 pályázat segítette elő.



Oktatási helyzet	Létszám	Tananyag vs. vizsgamódszer	Vizsgaeredmény
Normál	300+	gyakorlati + offline hallgatók → offline teszt	Rossz: 3-as alatti
Pandémias 1.	60	gyakorlati + online hallgatók → beadandó feladat	Jó: 3,98
Pandémias 2. / távoktatás (jövő)	500+	gyakorlati + online hallgatók online teszt	Jó: 3,5 körüli

7. ábra: A tananyag és a számonkérés egymáshoz való viszonya

részére alapoztunk. Később megpróbáltunk gyakorlati tudnivalókat is beépíteni, de az eredmények még így is igen rosszak voltak: hármast alatti volt az átlageredmény. Ez a módszer mind az oktatásban, mind a számonkérésben nagy energiabefektetést igényelt részünkről. Az első pandémias félévben a tananyag továbbra is gyakorlati jellegű maradt, ugyanakkor egy-egy anyagrész megismerésekor a hallgatók online lehettek, így rögtön ki is próbálhatták a szolgáltatást. A beadandó feladat szintén gyakorlati jellegű volt: a tananyag alkalmazásával kellett feladatot megoldani. Az értékelés szubjektívebb lehetett, s így sokkal jobb eredmények születtek: az átlag majdnem elérte a négyest. Ez a megoldás a feladatok ellenőrzésében roppant energiaigényes, ezért csak kis létszám mellett alkalmazható.

Az e-learning egyik nagy előnye a nagy létszám kezelhetősége. Egy másik előny, hogy a korábbiakkal ellentétben önmagában használható tananyagot kapnak: mind a szövegek, mind a videók rögzített módon tartalmazznak egy elméleti kérdéskört vagy egy szolgáltatás bemutatását, s így bármikor újra elővehetők. Az online tananyaghoz online jelenlét szükséges, így egy szolgáltatás ekkor is azonnal kipróbálható. Az online teszt fölvetette a visszaélések lehetőségét, ám az eredmények ezt nem igazolták: kiegyensúlyozott osztályzatok születtek, s az átlag kimozdulva a korábbi nagyon rossz tendenciából növekedést hozott – megközelítette a 3,5-öt. A tananyag és a vizsgához szükséges kérdésbank elkészítése energiaigényes, ám ez megtérülő energiabefektetésnek látszik.

A kurzussal kapcsolatban az alábbi kulcsszavakat tartjuk fontosnak: **gyakorlati ismeret** szeretnénk átadni **minőségi tananyagokkal**, melyek mind részünkre, mind a hallgatók számára biztosítják a **ráfordított energia megtérülését**, s **hatékonyan** kialakítják a szükséges kompetenciákat.

2.2. Tananyag

A tananyag összeállításakor a előző évek tematikájára támaszkodtunk, annak e-learning formájú földolgozását vizsgáltuk meg. A tananyagokat többen készítettük, lektorálásuk nagyrészt megtörtént.⁵ Mivel a tananyagok változását a keretrendszerben nem

⁵ A tananyag az SZTE Klebelsberg Könyvtár e-learning keretrendszerében: https://edu.ek.szte.hu/courses/konyvtar_es_informatikai_alapismeretek/?tab=tab-curriculum

követjük, ezért erre az ETA-t használjuk; abban minden egyes változat megtalálható.⁶ Az e-tananyaghoz kérdések is tartoznak, melyek segítségével a hallgatók ellenőrizhetik a tananyag elsajátítását és készülhetnek a vizsgára. Ezeket részben a tananyag készítői állították össze, részben kiegészítettük.



Tudás	Képesség	Attitűd	Autonómia / Felelősség
Felismeri a leggyakrabban előforduló dokumentumtípusok bibliográfiai leírását.	Irodalomjegyzékből beazonosítja a dokumentumokat, azokat megtalálja katalógusban vagy adatbázisban, s hozzáféréstől tájékozódik.	Törekszik a felhasznált szakirodalom korrekt hivatkozására.	Hivatkozásainak korrektségéért hivatkozásképző rendszert használ.
Alapvető ismeretei vannak a hivatkozásképzésről.	Elsajátította egy hivatkozásképző rendszer használatának alapjait.		
Ismeri a bölcsészeti- és társadalomtudományi területeken legjelentősebb magyar és külföldi adatbázisokat, azok tartalmát és keresési technikáit.	Kiszűri az adatbázisokban lefuttatott keresések releváns találatait, szűrők segítségével optimalizálja a találatok számát.	Tanulása eredményességéért rendszeresen és tudatosan használja az online adatbázisokat.	Önállóan eredményesen használja az online adatbázisokat, de tudja, hogy szükség esetén hol kérhet segítséget.

8. ábra: Tanulási eredmények

Fontosnak tartjuk a tanulási eredmény alapú megközelítést, amelyben meghatározzuk a tananyag elsajátításával elérhető kompetenciákat. Az ábrán néhány kiemelt szempont látható.

2.3. Kérdésbank

A véletlenszerű feladat kiválasztáson alapuló tesztvizsgálathoz egy nagyszámú kérdést tartalmazó kérdésbank szükséges. Kérdésbank létrehozása nem volt ismeretlen feladat, hiszen a vizsga kezdetek óta így működött. A vizsgálathoz jó kiindulást jelentettek a tananyag ellenőrző kérdései, azonban ennél sokkal többre volt szükségünk, ezért egy négy fős munkacsoport dolgozott a kérdések összeállításán. Több, mint 250 kérdést sikerült megfogalmaznunk, melyek áttestek egy lektoráláson és egységesítésen is. Ezután következett az online kérdésbank létrehozása.

A Szegedi Tudományegyetemen online kérdésbank és vizsgák létrehozására a Coospace (<https://www.coosp.etr.u-szeged.hu/>) nevű rendszert használjuk. A rendszerben több kérdéstípus található, melyekből a teszt céljának megfelelően választottunk. Törekedtünk arra is, hogy egy-egy témakörhöz könnyebb és nehezebb kérdések is tartozzanak, legyen köztük elmélet, de gyakorlat is. A feladatok elérhető pontszámát egytől háromig határoztuk meg, s az összetettebbeknél lehetőség volt részpontszámok elérésére is. Össze kellett még hangolni a tesztfeladatlapon megjelenő kérdések számát

⁶ A tananyag az SZTE Elektronikus Tananyag Archívumban: <https://bit.ly/3mBMyBz>



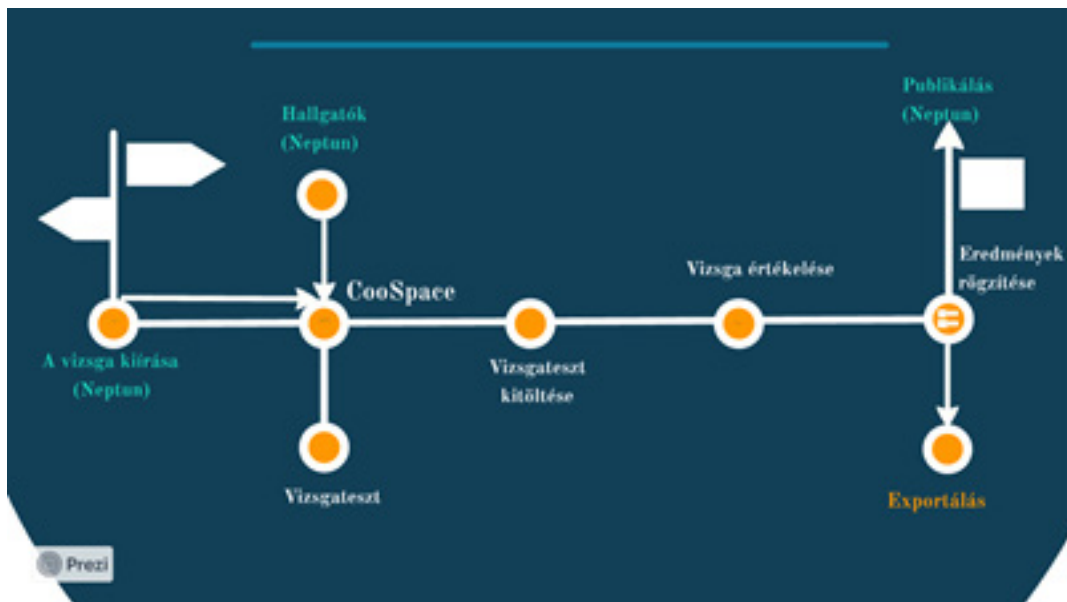
és témakörönkénti eloszlását, az elérhető összpontszámot és ponthatárokat, valamint a kitöltésre rendelkezésre álló időt.

A kérdésbankban mappák és almappák hozhatók létre, ezek strukturálása segíti a teszt létrehozását. Egy kérdéssor létrehozásánál a véletlenszerű feladatkiválasztáshoz megadhatók a mappák és az abból kiválasztandó feladatok száma. Ezért fontos a kérdések megfelelő strukturálása, mert így a véletlenszerű kiválasztás mellett is megoldható, hogy egy feladatlap minden témakörből tartalmazzon kérdéseket, s minden résztvevő egyenlő arányban kaphasson könnyebb és nehezebb feladatokat. A létrehozott kérdéssorból az előnézet segítségével mintafeladatsort is kérhetünk, így ellenőrizhető a megfelelő működés.

Az összes szempont figyelembevételével egy véletlenszerűen kiválasztott 20 feladatból álló kérdéssorra 40 percet kaptak a hallgatók, így egy-egy kérdés megválaszolására átlagban 2 perc jut. Ez enged némi rugalmasságot a teszt kitöltésében. Feladatonként 2 perc ugyanis arra nem elegendő, hogy valaki készületlenül, online kereséssel mindet helyesen megoldja. Arra viszont elég lehet, hogy ha valaki foglalkozott a tananyaggal és utána akar nézni egy feladatnak, akkor megtalálja azt, s ezzel jobb eredményt érheszen el.

2.4. Értékelés

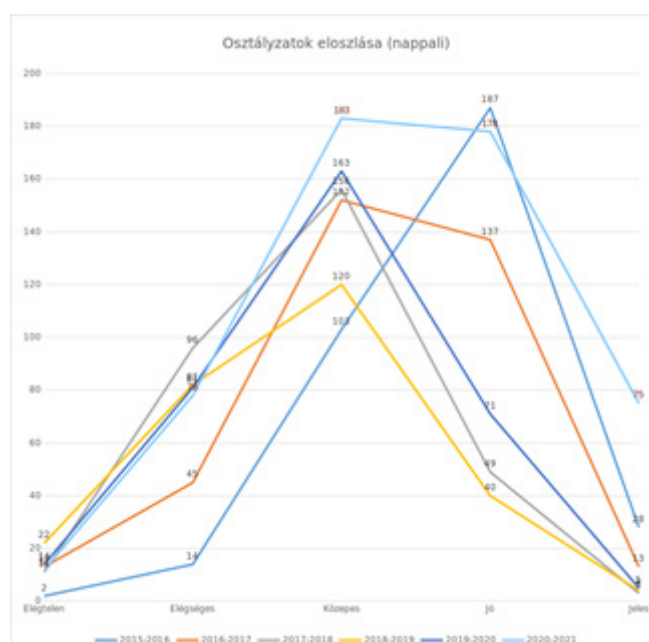
Tekintsük át a vizsga és értékelés folyamatát.



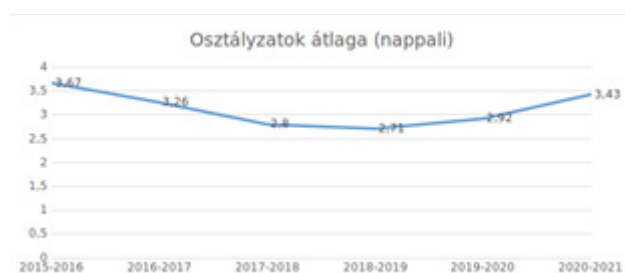
9. ábra: A vizsga és értékelés folyamata

(1) A tanulmányi rendszerben (Neptun) kiírásra kerül egy Coospace-es tesztvizsga. Itt állítható be a teszt elindíthatóságának ideje, a jelentkezők száma. (2) A Neptunban kiírt vizsgaalalom létrejön a CooSpace színtérben. Itt állítható be a kitöltés maximális ideje, az értékelés típusa és szabálya, részpontok számítása, a beadott feladatlap megtekinthetősége és módja. (3) A létrejött vizsgaszíntérben két irányból történik

hozzárendelés. (a) Neptun-CooSpace szinkronizációval a vizsgára a Neptunban jelentkezett hallgatók automatikusan hozzárendelődnek a CooSpace-vizsgaszíntérhez; indokolt esetben személyre szóló kitöltési idő is beállítható. (b) Hozzárendeljük a CooSpace-kérdésbankból elkészített vizsgatesztet; ekkor lehetőség van újabb ellenőrzésre egy próbakitöltés indításával. Ha a kérdéssoron változtatni szeretnénk, azt a kérdéssor-szerkesztőben tehetjük: a változások automatikusan érvényesülnek mindenhol. (4) A megadott időintervallumon belül elindítható a vizsga. Bármikor beadható vagy az idő lejártakor automatikusan beadásra kerül. (5) A vizsga értékelésekor fontos szempont volt, hogy a kérdések *automatikusan* értékelhetőek legyenek, s a hallgatók *azonnal* lássák az elért eredményt. Ugyanakkor lehetőség van manuális felülbírálatra, *újrapontozásra*. A feladatlap értékelése is megnézhető, s akár a helyes válaszok is. Ez utóbbit azonban nem szerettük volna, így csak az egyes feladatokra kapott *pontszámot* látják. A kitöltött feladatlapokról *statisztika* is kérhető, ahol az adott vizsgán egy kérdés összes előfordulását és a rá adott válaszok arányát látjuk. (6) A jegybeírás rendkívül egyszerű: két kattintással elindítható a publikálás a Neptunba. (7) Az eredmények *exportálhatók* és *archiválhatók*: ezekből monitorozzuk a teljesítményeket és az átlagok alakulását.



10. ábra: Osztályzatok eloszlása nappali tagozaton (2015–2021)



11. ábra: Osztályzatok átlaga nappali tagozaton (2015–2021)



Úgy gondoljuk, sikerült jó eredményt elérnünk az e-tananyag és a jól előkészített online tesztvizsga párossal. Ez látszik az átlag javulásán, de a jeles osztályzatok számának jelentős növekedésén is.

3. Összegzés

Az e-tananyag és archivált változata akár *különleges hallgatói igények kielégítésére* is alkalmas: 2020/2021 I. félévében magyar nyelvű képzési programban részt vevő kínai hallgatók oktatását kellett megoldanunk. A tananyagban nem változtathattunk, de számukra egy nyelvi egyszerűbb tesztet állítottunk össze, kevesebb feladattal és hosszabb kitöltési idővel. Az eredmények itt is igazoltak minket: a nyolc hallgatóból heten sikeres vizsgát tettek, az osztályzatok átlaga hármassal lett.

Az elektronikus tananyag, az erre épülő online teszt és a távoktatási forma tehát jól vizsgázott – ezen az úton szeretnénk tovább haladni. A jövő feladatai lesznek: 1. a tananyag és a kérdésbank folyamatos karbantartása, *fejlesztése*; 2. a tananyag *kiegészítése*, amely már régóta problémaként jelentkezett. Vannak ugyanis hallgatói csoportok, akik bár a BTK-n tanulnak, képzésük mégis közel áll pl. az egészségügyi szakokéhoz (pszichológus hallgatók). Ebbe az irányba mutat az is, hogy 2020/2021 II. félévétől nyitottunk az egyetem összes hallgatója felé: különféle képzési területekről érkezők vették föl – vannak a bölcsész- és társadalomtudományokba sorolhatóak is, de egészségügyi, valamint természettudományi területről is.

Az e-tananyaggal elért eredményben benne van sokéves megelőző munkánk: törekvésünk arra, hogy naprakész, a tanulmányokat hatékonyan segítő ismereteket nyújtsunk, próbálkozásaink a gyakorlati ismeretek elsajátításának fölmérésére, problémamegoldó gondolkodás a sajátos helyzetű vagy igényű hallgatók eseteinek kezelésében, valamint a nagy létszám ellenére a mindenkori rendelkezésre állás a kommunikációban.

Az online oktatás könyvtári támogatása a Szegedi Tudományegyetemen

Fülöp Tiffany
SZTE Klebelsberg Könyvtár
tiffany.fulop@ek.szte.hu
ORCID: [0000-0002-2219-8455](https://orcid.org/0000-0002-2219-8455)

Nagy Gyula
SZTE Klebelsberg Könyvtár
gyula.nagy@ek.szte.hu
ORCID: [0000-0002-8391-2851](https://orcid.org/0000-0002-8391-2851)

Fortunately, the forced change to online education required by the pandemic situation did not catch the University of Szeged completely unprepared. Our institution has a long tradition of using electronic teaching materials in various projects and situations. One of the most important tool is the SZTE Repository of Educational Resources, operated by the Klebelsberg Library, which stores almost 4,500 electronic learning material and it has more than 1 million downloads. In the first part of our study, we show our experiences gained during the development and running of the database. In the second part, we would like to talk about the development of an online education platform (<https://edu.ek.szte.hu>). It has launched last year and we presented here our first e-learning course, 'Introduction to library and information science'. This course was previously held-in-person by the library staff to hundreds of students per semester at the university as a general education course. Redesigning the course took several months. During this time, we had to face several technical and education-related issues and questions which gave us the experience and the necessary knowledge about the methods of switching from classroom teaching to online courses.

Keywords: online education, e-learning, e-learning platform, MOOC

Bevezető

A Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára már régóta zászlajára tűzte az online oktatás támogatása iránti erős elköteleződést, legutóbb minderről a 2018-as Networkshopon számoltunk be¹. Mindezt figyelembe véve a 2020-ban megjelenő COVID világjárvány szerencsére nem érte teljesen felkészületlenül intézményünket ezen a téren. A korábban elindult fejlesztések egy stabil alapot adtak az ilyen formán megnövekvő igények kiszolgálásához is. Elmondható, hogy a normális üzemmenetet kiszolgáló működési környezet mellett a járvány egy sürgető helyzetet teremtett, amely igen termékenyen hatott a SZTE Klebelsberg Könyvtár egyetemi oktatást kiszolgáló online szolgáltatásainak továbbfejlesztésére.

1 Nagy, Gyula. A Könyvtárak Szerepe a Digitális És Elektronikus Tananyagok Archiválásában. NETWORKSHOP 2018 Konferenciakiadvány (2018): 91–97.
<https://doi.org/10.31915/nws.2018.13>



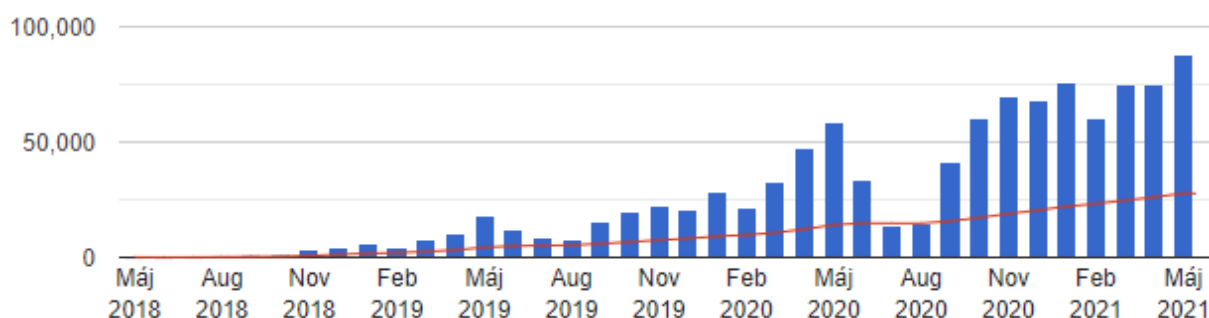
Mostanra mindenek egyik zászlóshajójává vált a Szegedi Tudományegyetemen a könyvtár által üzemeltetett SZTE Elektronikus Tananyag Archívum², amely 4500-at megközelítő elektronikus tananyagot tárol, a mindezekre vonatkozó letöltésszám pedig meghaladja az 1 milliós értéket. Az impresszív eredményeken túl tanulmányunkban a rendszer kialakítása és üzemeltetése kapcsán szerzett tapasztalatainkról számolunk be.

Írásunk második részében a könyvtárunkban a tavalyi évben elindult online oktatási platform (<https://edu.ek.szte.hu>) létrehozásáról és az itt megszülető első e-learning tananyagról (*Könyvtár- és informatikai alapismeretek*) szeretnénk beszámolni. Jelen konferenciakötetben munkatársaink a már elkészült digitális kurzus első tanéves oktatási tapasztalatairól is beszámolnak. Saját tanulmányunkban elsősorban a fejlesztési folyamatra fókuszálunk.

A kurzust korábban a könyvtár munkatársai tantermi keretek között oktatták az egyetemen, szemeszterenként több száz hallgatóval, általánosan művelő kurzusként. A kurzus átalakítása több hónapos folyamatot jelentett, melynek során számos technikai és oktatással kapcsolatos kérdést végigjártunk, így az online formába történő átültetés mikéntje és tapasztalatai komoly tanulságokkal szolgáltak. Egy korábban elnyert, EFOP 3.4.3-as pályázat egyik projektvállalása miatt már a tervezés fázisában ért bennünket a kitörő pandémia, mivel ekkor már dolgoztunk a digitális kurzus létrehozásán. Mindez az elképzelés igen sürgetővé vált akkor, amikor a személyes óratartás ellehetetlenült, továbbá a könyvtári szolgáltatások egy jelentős része – így a könyvtár jelenléti használatának megtapasztalása is – elérhetetlenné vált a hallgatók számára.

A könyvtár szerepe online oktatás támogatásában

Az SZTE Elektronikus Tananyag Archívum 2018 őszén indult el, ahogyan az 1. ábrán látszik, egyre növekvő népszerűsége tett szert a közönség körében, a tavaly őszi időszaktól kezdődően a tananyagok havi letöltési száma folyamatosan és stabilan a 60-90 ezres nagyságrendben mozgott.



1. ábra Az SZTE Elektronikus Tananyag Archívum látogatottsági adatai

Az archívum folyamatosan gyarapodott ebben az időszakban is, ahogyan a tartalmi fejlődés várhatóan a jövőben sem fog megállni. Gyűjtőkörbe tartozik minden oktatással kapcsolatos anyag és ezekhez kapcsolódó segédanyag: online oktatási csomagok (bármely e-learning és MOOC keretrendszerekből archivált oktatási csomag); jegyzetek, tankönyvek; prezentációk; oktatási témájú képi tartalmak; oktatóvideók

² SZTE Elektronikus Tananyag Archívum. Hozzáférés: 2021.05.30. <http://eta.bibl.u-szeged.hu>

és előadások felvételei; oktatási témájú audiófájlok és előadások hangfelvételei; útmutatók (oktatáshoz kapcsolódó tájékoztató anyagok, kurzusleírások, tanmenetek); önálló tanítási erőforrások; tanulási objektumok; tanulást segítő dokumentumok; feljegyzések előadásokról; handoutok; kötelező olvasmányok; gyakorlófeladatok; tesztek; vizsgakérdések.

Az archívum fő előnyei közé tartozik, hogy nincs feltöltési méretkorlát, nincs fájl típusbeli megkötés, bármennyi anyag feltölthető, a tananyagok szakszerű metaadatolást kapnak, és a könyvtárosok elvégzik a részletes szakterületi besorolást is. Ilyen módon biztosítva van az elektronikus tananyagok hosszú távú archiválása a teljes egyetemre nézve, amely egyúttal több fajta hozzáférési szinten valósulhat meg. Lehetőség van bizonyos anyagok esetében a tananyagok EduID-val való hozzáférési szintre való korlátozására. Az adatbázis felhasználóbarát kezelését részletesen kidolgozott útmutatókkal, súgóanyagokkal, oktatóvideókkal és professzionális támogató ügyfélszolgálat biztosításával segítjük, mind a hallgatói, mind az oktatói, tananyagfeltöltői irányokba. Az elektronikus tananyag repozitórium egy meglévő, jól bejártatott ökoszisztémába illeszkedik, hiszen a Contenta repozitóriumrendszer³ része, tételei a repozitóriumok közös keresőjében⁴ is megtalálhatóak.

Az adatbázis népszerűségének különösen jót tett, hogy az egyetemen futó, elektronikus tananyagfejlesztést célzó EFOP pályázatok keretében elkészült tananyagokat az ETA-ba kell feltöltenie (validálás) a készítőknél. Egy célzott projekt keretében a könyvtár munkatársai elvégezték a teljes egyetemi e-tananyagportfólió felderítését és begyűjtését az egyes tanszéki, intézeti, kari oldalokról, hogy ezek egységes felületen, egy helyen váljanak elérhetővé, egyúttal jelentős tartalomnövekedést biztosítva az adatbázis számára. Az adatbázis használata természetesen beépült a könyvtár különböző felhasználóképzéseibe is, továbbá Youtube információs-népszerűsítő videó is készült az archívumról⁵.

A pandémia alatt folyamatos híradásokban számoltunk be az archívum lehetőségeiről és az összes többi hiánypótló online könyvtári szolgáltatásról a központi egyetemi honlapon, a Coospace tanulmányi rendszer megfelelő szinterein, a könyvtár saját honlapján és Facebook oldalán. Az adatbázis jelenleg körülbelül 1 terabájtnyi tananyagot tárol és szolgáltat, típusok szerint a 2. ábrán látható megoszlás alapján.

3 Contenta repozitóriumok. Hozzáférés: 2021.05.30. <http://contenta.bibl.u-szeged.hu>

4 SZTE Klebelsberg Könyvtár Repozitóriumok közös keresője. Hozzáférés: 2021.05.30. <https://contentas.bibl.u-szeged.hu>

5 Az SZTE Elektronikus Tananyag Archívum (ETA) használata. Hozzáférés: 2021.05.30. <https://youtu.be/dVEuqoPYeSw>



2. ábra Elektronikus tananyagok megoszlása tananyagtípus és fájltypus szerint az SZTE ETA-ban

A digitális kurzus kivitelezésének előkészítése

Az első lépés a keretrendszer kiválasztása volt. Lehetőségként felmerült az eXeLearning⁶ szoftvereszköz használata, de viszonylag gyorsan elvetésre került. Ez egy kompakt szerzői környezetet biztosít a készítőknél, ahol nincs szükség jelentős fejlesztői ismeretekre, azonban megjelenését tekintve viszonylag azonos sémára épül fel a segítségével elkészített tananyag, melynek online elérhetővé tételét maga az eszköz nem biztosítja.

Ezután még két további opciót vizsgáltunk meg: az egyik a Moodle⁷ e-learning keretrendszer volt, amely egy olyan – a tanulás szervezését segítő – tartalomkezelő rendszer (CMS – Content Management System), mely jellegéből fakadóan nem csak a tananyag elkészítéséhez, hanem az oktatásszervezés különféle aspektusaihoz is alkalmas eszközt biztosít. A rendszer inkább az online szerveződő tanítás interaktív, folyamatos kapcsolódását támogatja, melynek működtetése sok igényt támaszt, és aminek egy jelentős része kihasználatlan maradt volna a részünkről.

A következő opció egy MOOC⁸ (Massive Open Online Course) jellegű, vagyis egy olyan online kurzus létrehozása volt, amely alapvetően nyílt hozzáférésű, ingyenes, ezáltal tulajdonképpen bárki szabadon elvégezheti. Az ilyenfajta szabadság mellett kényelmes és rugalmas hozzáférést biztosít a felhasználók számára. Mindenki a saját tempójában végezheti a kurzust, az elsajátítást pedig vizuális elemek és önellenőrző tesztek segítik. A MOOC-megközelítés jelenleg nemzetközileg is kedvelt irányvonalát jelenti az online tanulásnak. Az ilyenformán szervezett tananyag emiatt nagyobb hatékonysággal kecsegtetett, hiszen sokan használják és ismerik az általában bevett és alkalmazott módszereit. További előnye volt számunkra a nagy létszám vonatkozásában, hogy kevesebb egyéni revíziót igényel, a tananyag elsajátítása önállóságot ad a hallgatóknak a tanulásra, így a mi céljainknak is jobban megfelelt.

Az e-tananyag csomag kivitelezése a technikai oldaltól tekintve egy komplett weboldal elkészítését jelentette, amely a WordPress tartalomkezelő rendszer segítségével lett létrehozva. Az így kialakított felület jelenleg az SZTE Klebelsberg Könyvtár e-tananyag

6 eXe eXeLearning. Hozzáférés: 2021.05.30. www.exelearning.org

7 Moodle - Open-source learning platform. Hozzáférés: 2021.05.30. www.moodle.org

8 Kokas, Károly. 2013. Könyvtárak a Rubiconnál. EDUCATIO 22 (3): 363–376. Hozzáférés: 2021.05.30. <http://www.edu-online.eu/hu/letoltes.php?fid=tartalomSOR/2287>

portáljaként (<https://edu.ek.szte.hu>) funkcionál, ahol azóta már nem csak az írásunkban most bemutatott digitális kurzus, hanem néhány további, szintén a könyvtár által, a KDS-K pályázat⁹ keretében készített komplex online oktatóanyag (Klebsberg Kuno élete és munkássága, Egyetem Szegeden, Szegedi tudósok hálózatban) is megtalálható¹⁰.

A weboldal környezetének létrehozása után egy olyan bővítményt telepítettünk, amely kialakítja és támogatja a MOOC típusú anyagok szolgáltatását: a kurzushasználati beállításokat, a leckék megfelelő felépítését (olvasóleckék és oktatóvideók, ráfordítási idő és haladási irány) és az önellenőrző kérdéseket. A telepített bővítmény a LearnPress – WordPress LMS Plugin¹¹.

A digitális kurzus tartalmának létrehozása

A tartalmat tekintve fontos volt az említett struktúra megléte, miszerint óraanyagokra és ellenőrzőkérdésekre van szükség. A jelenlegi kurzus tartalmi elemeit figyelembe véve egyrészt olvasóleckék készültek, amiknél cél volt a jól strukturált, médiaelemekkel ellátott modern oldalak kialakítása. Nagy hangsúlyt fektettünk az aktív (lenyíló, kattintható) eszközök és a grafikai elemek felhasználására, amik vezetnek a diákok figyelmét és segítik a tanulást. Emellett prioritást jelentettek még az audiovizuális tartalmak, oktatóvideók is, ahol elsődlegesen gyakorlati ismeretek kerültek bemutatásra, narrációval ellátott képernyőfelvételek formájában. Ezeket utólag a videóban megjelenő kiemelésekkel tettük színesebbé, könnyebben használhatóvá.

A tananyag létrehozása egy több hónapon keresztül tartó projekt volt sok résztvevővel, aminek a koordinálása is nagy munkát igényelt. Az első megbeszélések a kurzus tartalmi újragondolásáról és a MOOC típusú anyagok jellegzetességeivel történő ismerkedésről szóltak. Nagyrészt ugyanazok a kollégák készítették el az online anyagokat, akik a korábbi, hagyományos előadásokat is tartották, azonban a digitális kurzus sokkal több szabadságot és lehetőséget biztosított számukra is. Néhány esetben magunkban is újragondoltuk és újraértékeltek, hogy valójában milyen ismeretekre van szüksége egy egyetemistának a modern könyvtárhasználat során a 21. század második évtizedében ahhoz, hogy minél hatékonyabban tudjon megfelelni a saját szakja által támasztott követelményeknek.

Az új tematika anyagainak elkészítése során bevontunk még néhány, az adott témában jártas munkatársat is. A tematika kialakítása, sorrendiségének megállapítása és a témák felosztása a kollégák között többszöri átgondolás és egyeztetés után nyerte el végleges formáját. Ahhoz, hogy jó struktúrájú tananyagszövegek és forгатókönyvek készüljenek, a kollégáknak meg kellett ismerkedniük az e-learning anyagok speciális struktúrájával: hosszúság, szövegen túli elemek összegyűjtése, plusz tartalmak, kiemelések meghatározása, elkülönítése.

9 KDS-K. Hozzáférés: 2021.05.30. <http://www.oszk.hu/kds-k>

10 Panek, Sándor. Digitális oktatóanyagok és 25 ezer lapnyi régi könyv a Klebsberg Könyvtártól. (2021) Hozzáférés: 2021.06.04. <https://www.delmagyar.hu/kozelet/helyi-kozelet/digitalis-oktatoanyagok-es-25-ezer-lapnyi-regi-konyv-a-klebsberg-konyvtartol-7587817>

11 LearnPress - WordPress LMS Plugin. Hozzáférés: 2021.05.30. <https://hu.wordpress.org/plugins/learnpress>



Először az olvasóleckéket készítettük el. Az elkészült anyagok nyers szövegeit egyik kollégánk lektorálta, majd azok folyamatosan kerültek át a technikai elkészítés fázisába. Amíg az ezért felelős munkatársak a tartalmat állították elő, párhuzamosan létrehoztuk a technikai környezetet: telepítettük a WordPress keretrendszert, kialakítottuk a honlap megjelenését, majd felkerült és konfigurálva lett a LearnPress bővítmény. Egy-egy olvasólecke kinézetének kialakításához a bővítmény az ingyenes szolgáltatásai között egy egyszerű szövegszerkesztőt kínál csak fel, ami nem alkalmas arra, hogy a tartalmat aktív elemekkel lássuk el. Ezért az első beérkező anyagokon kialakítottuk az olvasóleckék sablonját: létrehoztunk egy dizájnt és leprogramoztuk a grafikai, aktív elemeket W3.CSS nyelven. Ezután elkezdődött az elkészült nyers szövegek adaptálása a LearnPress bővítmény megfelelő aloldalaira.

Amíg az olvasóleckék technikai véglegesítése zajlott, a kollégák elkezdtek dolgozni az oktatóvideókkal. Akik vállalták, azok saját maguk készítették el a videóleckék nyers felvételeit, amely munkához sokszor saját felszerelést használtak, hiszen többen a járvány miatt otthonról dolgoztak. Ehhez a feladathoz tulajdonképpen egy képernyőfelvevő programra és egy mikrofonra volt szükség, hiszen az oktatóvideókat azzal tudtuk igazán felhasználóbaráttá tenni, hogy folyamatos hangos kommentár és plusz szöveges-grafikai információk egészítették ki a rögzített képernyőképet, ezzel is segítve a diákokat a megértésében. Az ilyen módon elkészített nyers felvételek ezután még további utómunkára szorultak.

A munkatársak másik felével közösen dolgoztunk a felvételek elkészítésén, ők megírták a forgatókönyvet, ami alapján a könyvtár által beszerzett Camtasia¹² képernyőfelvevő és videóvágó programmal elkészítettük a videóanyagokat. A program komoly effekt és animációs lehetőségeket biztosít, így a videók a vágás után kaptak egy-egy felvezető és lezáró animációt, részegységekre bontottuk őket (3-8 perc) és fejezetcímekkel, magyarázó szövegekkel, figyelmet irányító effektekkel (nagyítások, kiemelés) lettek ellátva.

12 Camtasia: Screen Recorder & Video Editor. Hozzáférés: 2021.05.30.
www.techsmith.com/video-editor.html

Search courses content

Könyvtár- és informatikai alapismeretek

Mi az a repozitórium?

Ebben a témakörben a repozitóriumokkal fogunk részletekbe menően megismerkedni.

Mindezek során két nagyobb témát fogunk körbejárni.

A repozitóriumokról
Mik ezek, mit tudnak és egyáltalán hogyan jöttek létre?

Konkrét példák
Melyik magyar és hazai változatait érdemes használni?

Fontos
A szövegben csillaggal jelölt fogalmakat felugró ablakban olvashatja, illetve ebből az anyagból is visszakéresheti a kurzushoz kapcsolódó fogalomtárban, s ott olvashatja az adott fogalom, szakszó részletesebb magyarázatát!

A repozitóriumokról

A hagyományos értelemben vett könyvtári információszerzés, mely elsősorban a helybenhasználatra és a kölcsönzésre alapult.

Előző Következő

3. ábra Képernyőkép a 'Könyvtár- és informatikai alapismeretek' c. digitális kurzusról

Az e-tananyag véglegesítése és az ellenőrző kérdések

Az elkészült olvasóleckék és oktatóvideók első körös utóellenőrzését az anyagok készítőire bíztuk. A kész órák végleges formáját látva így az is előfordult, hogy további ötletek merültek fel a bővítést, javítást illetően. Ezután, utolsó lépésként egy, addig a projektben részt nem vevő kolléga külsős szemmel is megnézte az anyagokat, hogy azok biztosan logikusak, jól érthetőek legyenek. Az ellenőrző kérdések például többször korrigálásra kerültek, mert fontos célnak tekintettük, hogy az olvasott és látott anyag minden fontosabb aspektusára rákérdezzenek. Ezek az igaz-hamis valamint az egy- és többválasztós kérdéstípusokba tartoznak, ahogy egy példa alapján látható is a 4. ábrán.



9) Ellenőrző kérdések

Kérdés: 1/14

TESZT BEFEJZÉSE

3. Mi az az OCR?



egy e-könyv formátum



egy e-könyv szerkesztésére alkalmas szoftver



egy eljárás, amely során a beszkenelt gépelt vagy kézzel írt dokumentumot átalakítjuk digitális, elektronikus formátummá, szöveggé

4. ábra Példa egy ellenőrző kérdésre

Összegzés

Az SZTE Elektronikus Tananyag Archívum hiánypótló szerepet töltött be egyetemünkön a járványidőszak alatt, a kidolgozott modell jól vizsgázott és nagy népszerűsége tett szert. Segítségével a könyvtár valóban be tudta tölteni oktatástámogató szerepét az online oktatás során is. Az adatbázis várhatóan a járványidőszak után is szerves részét fogja képezni a Szegedi Tudományegyetem online oktatási ökoszisztémájának, hiszen a könyvtár által ez a forrás vált az egyik legnagyobb, központi elektronikus tananyag lelőhellyé az egyetemen.

Az írás második felében ismertetett elektronikus tananyagfejlesztés intézményünk életében egy igazán komplex projektet jelentett, amely folyamat tapasztalataival sokat gyarapodott könyvtárunk a kurzus elkészítése során, amely azóta további online tananyagok elkészítésére motivált minket. Az ismertetett digitális kurzus kivitelezési munkálatai a tavaly nyári időszakban zajlottak, így az ősszel induló, rendhagyó, online oktatásra alapuló szemeszterben már az e-kurzus keretein belül sajátíthatták el a könyvtárhasználathoz szükséges tudást a hallgatók. A felületet több száz fő használta nagy megelégedéssel, melynek tapasztalatairól munkatársaink a korábban említett, kapcsolódó Networkshop előadásban számoltak be.

A digitáliskompetencia-fejlesztés új eszköze: a DigKomp Rendszer

Sörény Edina

Digitális Jólét Nonprofit Kft. (Budapest)

soreny.edina@djnkft.hu

Digital competence is a complex system of knowledge, practical skills, and behavioral elements that enables confident, critical, and creative use of digital technology-based tools and applications to make work, learning, leisure, and civic engagement more effective. Based on the 2019 data of the DESI index, the Government decided to develop a model for a digital competence development system. The Hungarian Digital Competence Framework (DigKomp System) develops and organizes tools into the system: DigKomp Learning Support Platform, DigKomp Training Register, DigKomp Certification Points, reference materials, which can serve the digital competence development of citizens more effectively than before.

Keywords: digital competence, digital competence framework, competence profile, DigKomp System

1. Bevezető

A 2006-ban megjelent kulcskompetenciákra vonatkozó Európai Ajánlás keretében az Európai Unió a digitális kompetenciát elismerte, mint az egész életen át tartó tanulásához szükséges nyolc kulcskompetencia egyikét¹. A digitális kompetencia ismeretek, gyakorlati képességek és viselkedési elemek komplex rendszere, amely lehetővé teszi a digitális technológián alapuló eszközök és alkalmazások magabiztos, kritikus és kreatív használatát a munkavégzés, a tanulás, a szabadidő szervezése, illetve az állampolgári aktivitás hatékonyabbá tételé céljából. A megfelelő szintű digitális jártasság ma már elengedhetetlen a világban való eligazodáshoz és az aktív léthez, továbbá a versenyképes digitális gazdaság működéséhez is.

2. A változás gyorsuló üteme

Nemzetközi jövőkutatások azt vizionálják, hogy a 2030-ra olyan munkakörök jönnek létre, amelyek 85%-a ma még nem létezik,² és a ma általános iskolába járó gyermekeknek a 65%-a teljesen új, még nem létező munkakörben fog dolgozni³. A digitális kompetencia kulcsfontosságúvá vált a foglalkoztathatóság szempontjából.

-
- 1 Ferrari, Anusca, DIGCOMP: A digitális kompetencia értelmezésének és fejlesztésének európai keretrendszere, hozzáférés: 2021.05.28., https://www.deaweb.hu/images/bongeszde/digcomp_teljes_hun_151231.pdf
 - 2 The next era of human|machine partnerships:emerging technologies' impact on society & work in 2030, hozzáférés: 2021.05.27., https://www.delltechnologies.com/content/dam/delltechnologies/assets/perspectives/2030/pdf/SR1940_IFFTforDellTechnologies_Human-Machine_070517_readerhigh-res.pdf
 - 3 Centeno Mediavilla, Isabel Clara;Vvuorikari, Riina; Punie, Yves et al., Developing digital competence for employability: Engaging and supporting stakeholders with the use of DigComp, hozzáférés:2021.05.29., <https://publications.jrc.ec.europa.eu/repository/handle/JRC118711>



Nemcsak a szerepét tekintve transzverzális készség a foglalkoztathatóság fejlesztése érdekében, hanem azért is, mert az Európai Unió munkahelyeinek körülbelül 85%-ában legalább alapvető digitális készségszintre van szükség⁴. Az előrejelzések alapján gyakoribb lesz az életciklusok során a munkahelyváltogatás, ez akár a 15–20 munkahelyet is jelenthet. Emellett a munkakörök „tartalmának” gyors változása miatt akár már 2022-re a jelenlegi munkaerő 54%-ának átképzésre lesz szüksége. Az Európai Bizottság Közös Kutatóközpontja (Joint Research Centre – JRC) által 2019 nyarán rendezett bilbaoi workshopon elhangzott, hogy a munkavégzés formája is jelentősen át fog alakulni. Míg a 2018. évi elemzések alapján a gépi/algorithmizált és a humán munkaerő munkavégzésének aránya 29%–71%, addig a 2022. évi előrejelzések alapján a gépi/algorithmizált munkavégzés 42%-ra nő, míg ezzel párhuzamosan a humán munkaerő alkalmazása 58%-ra csökken. (Developing digital competence for employability: Engaging and supporting stakeholders with the use of DigComp⁵, JRC, 2019.).

3. Digitális környezeti kihívások

Az Európai Unió kutatások alapján a polgárok 45%-ának nincs, vagy nagyon alacsony szintű a digitális tudása. A munkaerő 37%-ának pedig nincs, vagy csak nagyon alacsony szintű a digitális kompetenciája. A vállalkozások 40%-a IT-szakemberhiánnyal küzd. Mára a munkakörök 90%-ánál elvárás bizonyos szintű digitális készség megléte, ami területenként eltérő. Zajlik a munkakörök átalakulása, új munkakörök létrehozása (pl.: adatgazda, azaz data steward).

A technológiai környezet gyors változása miatt az állampolgároknak hatékony, a digitális készségek fejlesztéséhez segítséget adó eszközök kifejlesztése vált szükségessé.⁶ Emellett az EU-ban a tagállamok digitális környezeti kihívásokra való felkészülése, felkészítése az elmúlt évek egyik kiemelt és prioritizált területe. Az európai digitális menetrendhez kapcsolódóan, a digitális versenyképesség alakulásának monitorozása érdekében 2015-ben bevezetésre került a DESI index⁷ (Digital Economy and Society Index - Digitális Gazdaság és Társadalom Index), ami 5 területen évente méri a tagállamok digitális fejlettségét.

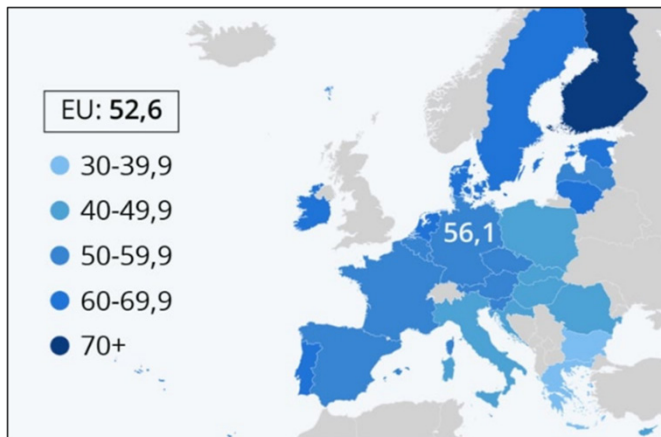
4 Insights into skills shortages and skill mismatch, Cedefop, 2018.-7. p.-

<http://www.cedefop.europa.eu/en/publications-and-resources/publications/3075>

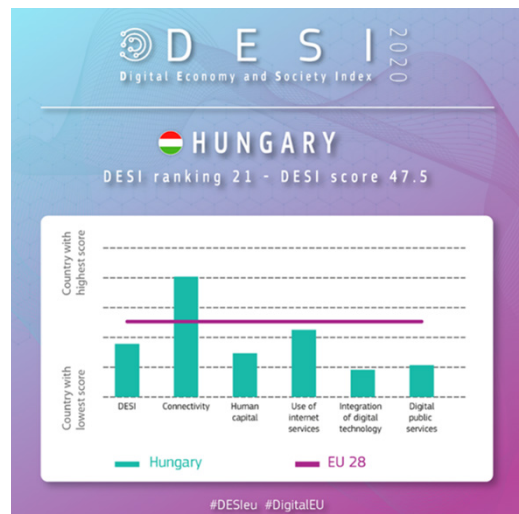
5 Centeno Mediavilla, Isabel Clara; Vuorikari, Riina; Punie, Yves et al., Developing digital competence for employability: Engaging and supporting stakeholders with the use of DigComp, hozzáférés:2021.05.29., <https://publications.jrc.ec.europa.eu/repository/handle/JRC118711>

6 Hodson Deirdre, European Distance Learning Week: Tackling the digital skills gap in the EU – policies and programmes, hozzáférés: 2021.05.29., https://pt.slideshare.net/eden_online/european-distance-learning-week-tackling-the-digital-skills-gap-in-the-eu-policies-and-programmes/4

7 The Digital Economy and Society Index (DESI), 2020., hozzáférés: 2021.05.29., <https://digital-strategy.ec.europa.eu/en/policies/desi>



1. ábra Az EU tagállamok DESI indexe (2020)

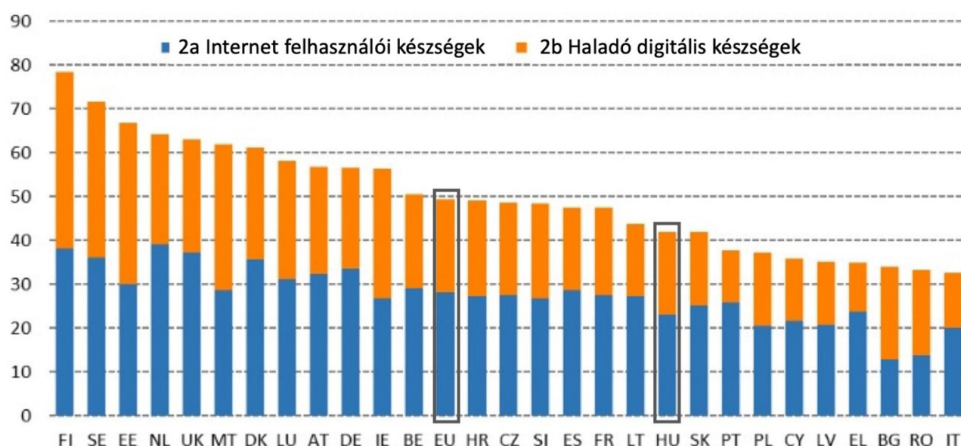


2. ábra Magyarország DESI-mérőszáma és digitális felkészültsége (2020)

4. Miért van szükség a DigKomp Rendszerre?

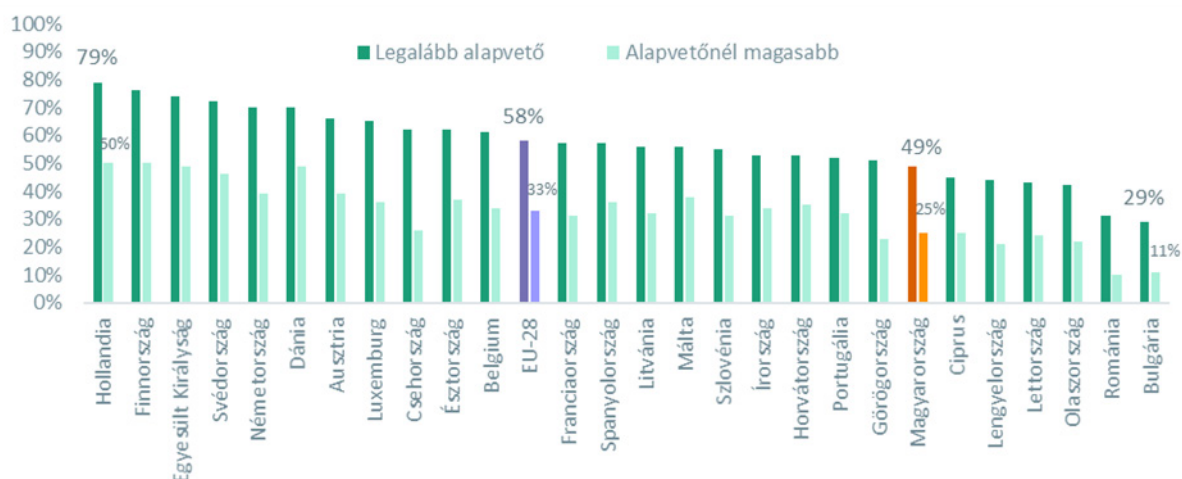
A DESI 4., Humántőke-területe méri a polgárok digitális kompetenciájának fejlődését. Ma már aligha kérdőjelezhető meg, hogy az állampolgárok digitális kompetenciájának fejlesztése elsőrendűen fontos cél, a fejlesztés sürgősségét a DESI index Magyarországra vonatkozó 2020. évi adatai is jelzik, és számos más elemzés is alátámasztja. A koronavírus-járvány következtében kialakult helyzet is rávilágított, hogy az állampolgároknak milyen sok területen van szükségük a digitális technológia használatában való magasabb szintű jártasságra. Bár a kényszer hatására sokféle hasznos és előremutató megoldás született, eközben a nyilvánvaló lemaradás is megmutatkozott.

Digitális Gazdasági és Társadalmi Index – DESI 2020 Humán tőke érték (0-100), 2019



Forrás: <https://ec.europa.eu/digital-single-market/en/human-capital>

3. ábra: Humántőke-érték (0–100), Internetfelhasználói készségek



4. ábra DESI index - Digitáliskészség-szintek, EU 2020

Az állampolgárok digitális kompetenciájának fejlesztése és a versenyképes digitális gazdaság működéséhez, a megfelelő tudással rendelkező munkaerő biztosításához, az online közszolgáltatások igénybevételének növelésére és így az életminőség javításához szükséges digitális készségek magasabb szintre emelése deklaráltan kiemelt kormányzati feladattá vált, aminek eredményeként megszületett a 1341/2019 (VI.11) Korm.határozat.⁸ A dokumentum rendelkezett a Digitális Kompetencia Keretrendszer fejlesztéséről és bevezetésének lépéseiről, amelyben kijelölésre került a fejlesztéshez keretet biztosító GINOP 6.1.2-15-2015-00021 „Digitális szakadék csökkentése”⁹ című projekt. A Digitális Jólét Nonprofit Kft. előkészítette egy új digitáliskompetencia-fejlesztő eszközrendszer, a Digitális Kompetencia Keretrendszer (a továbbiakban: DigKomp Rendszer) fejlesztésére vonatkozó hosszú távú koncepció javaslatát, a működési modelljét, illetve a digitális kompetencia értékelési és igazolási rendszerének lehetőségeit, továbbá a digitális készségeket fejlesztő képzések regiszterét és az ösztönzőrendszer kialakítására vonatkozó javaslatokat.

5. A DigKomp Rendszer fő célja, elemei és működési modellje

A tervezett DigKomp Rendszer elsődleges célja, hogy az állampolgárok digitális kompetenciájának fejlesztését rendszerszerűvé és mérhetővé tegye, mely tanúsítvánnyal igazolható. Ennek része a fejlesztési területek és célok világos, a nemzetközi referenciákkal is összhangban álló szabványos megfogalmazása, naprakészen tartása, az egyéni tanulást támogató megoldások kidolgozása, az értékelés és tanúsítás mechanizmusának kiépítése, a digitális kompetencia fejlesztésére történő ösztönzés.

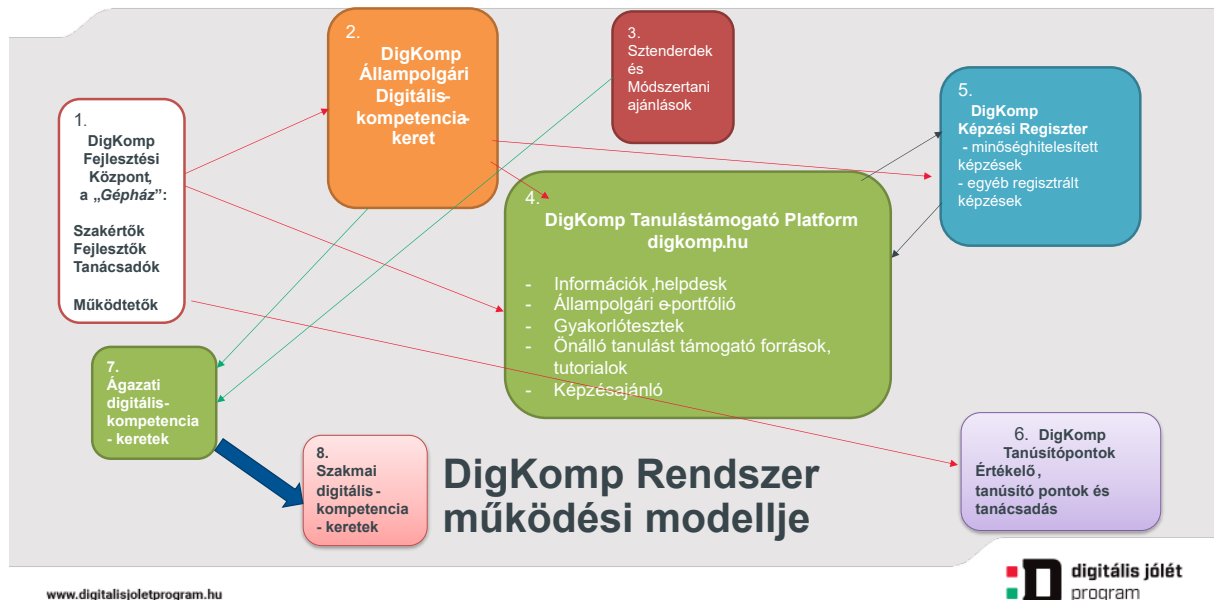
8 1341/2019. (VI. 11.) Korm. határozat a Digitális Kompetencia Keretrendszer fejlesztéséről és bevezetésének lépéseiről, hozzáférés: 2021.05.29., <https://net.jogtar.hu/jogszabaly?docid=A19H1341.KOR×hift=ffffff4&txtreferer=00000001.TXT>

9 GINOP-6.1.2-15 – Digitális szakadék csökkentése, hozzáférés: 2021.05.29., <https://www.palyazat.gov.hu/doc/4456>

A DigKomp Rendszer olyan eszközöket fejleszt és szervez rendszerbe, amelyek az eddiginél hatékonyabban szolgálhatják az állampolgárok (például: diákok, munkavállalók, álláskeresők, fogyatékosokkal élők, stb.) digitális kompetenciájának fejlesztését.

A DigKomp Rendszer tervezett fő elemei:

- a DigKomp Központ, ami a „Gépház” szerepét tölti be;
- a DigKomp Tanulástámogató Platform, ami segítséget nyújt a digitális kompetencia egyéni fejlesztéséhez, de akár csoportos tanulás keretében is használható a jövőben kifejlesztésre kerülő feladattára, illetve az azokhoz kapcsolódó tutorialtár révén;
- a DigKomp Képzési Regiszter, ami összegyűjti és strukturált módon teszi elérhetővé az információkat a digitális kompetencia fejlesztésére irányuló képzésekről és egyéb digitáliskompetencia-fejlesztő programokról;
- a DigKomp Tanúsítópontok, ahol a kompetenciaprofilokhoz kapcsolódó DigKomp-tanúsítvány megszerzése történik.



5. ábra A DigKomp Rendszer modellje

6. DigKomp Állampolgári Digitáliskompetencia-keret – Adaptáció és Innováció

A fejlesztés alatt álló DigKomp Rendszer részét képező állampolgári digitáliskompetencia-keret olyan referenciaanyag, amely strukturált formában írja le a ma korszerűnek tartott állampolgári digitális kompetencia összetevőit. A hazai fejlesztés keretében az állampolgári keret az Európai Bizottság által 2017-ben kidolgozott és jelenleg érvényes referenciának tekintett DigComp 2.1¹⁰ állampolgári digitáliskompetencia-keret alapján került kialakításra olyan módon, hogy a hazai sajátosságok megjelenítése mellett lehetővé teszi a nemzetközi összehasonlítást is.

¹⁰ DigComp 2.1: Állampolgári digitáliskompetencia-keret, /magyarul/, 2019. hozzáférés: 2021.05.29., <https://dpmk.hu/2019/07/25/a-digitalis-kompetencia-unios-referenciakerete-magyarul/>



Az Európai Unió állampolgári digitáliskompetencia-keret fejlődése és a hazai DigKomp Állampolgári Digitáliskompetencia-keret			
DigComp 1.0	DigComp 2.0	DigComp 2.1	DigKomp Állampolgári Digitáliskompetencia-keret hazai kompetencia-keret
EU kompetencia-keret	EU kompetencia-keret	EU kompetencia-keret	
2013	2016	2017	2020
5 kompetenciaterület	5 kompetenciaterület	5 kompetenciaterület	5 kompetenciaterület
1. Információ /3 kompetenciaelem/ 2. Kommunikáció /6/ 3. Tartalomkészítés /4/ 4. Biztonság /4/ 5. Problémamegoldás /4/	1. Információ- és adatközlés 2. Kommunikáció és közös munka 3. Digitális tartalomfejlesztés 4. Biztonság 5. Problémamegoldás	1. Információ- és adatmenedzsment 2. Kommunikáció és együttműködés 3. Digitális tartalmak 4. Biztonság 5. Problémamegoldás	1. Információk és adatok kezelése, használata /3/ 2. Kommunikáció és együttműködés /6/ 3. Digitális tartalmak létrehozása /4/ 4. Biztonság /4/ 5. Különböző problémák kezelése /4/
21 kompetenciaelem	21 kompetenciaelem	21 kompetenciaelem	21 kompetenciaelem
3 jártassági szint	8 jártassági szint	8 jártassági szint	8 jártassági szint
Alapszint /A/ Középszint /B/ Felsőszint /C/	Alapszint /A/ Középfelhaladó/B/ Haladó/C/ Magasan specializált v2.0	Alapszint /1-2/ Középszint/3-4/ Haladó szint /5-6/ Mesterszint /7-8/	Alapszint /1-2/ Középszint/3-4/ Haladó szint /5-6/ Mesterszint /7-8/
https://ec.europa.eu/irc/en/publication/digcomp-framework-developing-and-understanding-digital-competence-europe https://www.deaweb.hu/images/bongesz/de/digcomp_teljes_hun_151231.pdf	https://ec.europa.eu/irc/en/publication/eur-scientific-and-technical-research-reports/digcomp-20-digital-competence-framework-citizens-update-phase-1-conceptual-reference-model	https://ec.europa.eu/irc/en/publication/eur-scientific-and-technical-research-reports/digcomp-21-digital-competence-framework-citizens-eight-proficiency-levels-and-examples-use	munkaanyag

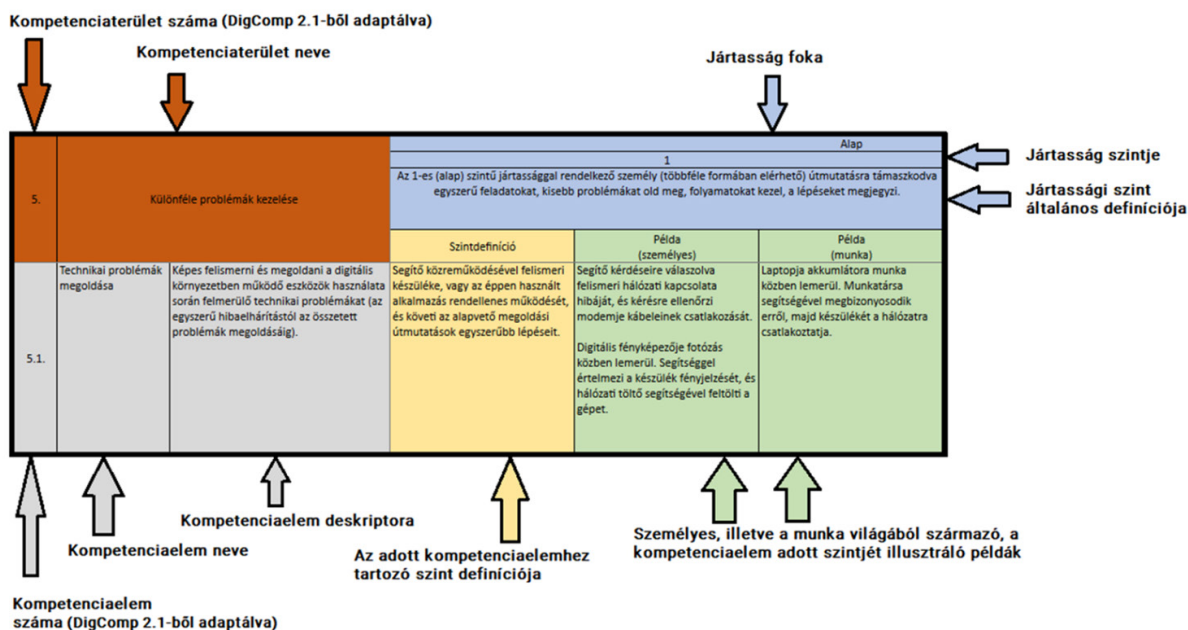
6. ábra A DigComp-tól a DigKomp-ig

A hazai állampolgári keret (DigKomp ÁDKK) döntő mértékben az uniós referencia, a DigComp 2.1 keret szerkezetét és szintezési alapelveit követi, ami 5 kompetenciaterületet, 21 kompetenciaelemet és ahhoz kapcsolódó 8 jártassági szintet tartalmaz.

A hazai fejlesztés hozzáadott elemei:

- Törekvés az online szolgáltatásokkal kapcsolatos kompetenciaelemek megjelenítésére a gyakorlati példákon keresztül.
- A kompetenciaterületek megnevezésének és a kompetenciaelemek deskriptorainak pontosítása.
- Hazai környezetben értelmezhető példák készítése – a privát életterben és a munka világához kapcsolódóan. (Teljes példarendszer létrehozása: minimum 500 példa).
- Szintdefiníció készítése (168) – minden kompetenciaelemhez, minden jártassági szinten készült szintdefiníció a tananyagfejlesztők és a képzésszervezők támogatása érdekében.
- Új dimenziók létrehozása: belső területközi kapcsolódások jelzése.
- Profilban való gondolkodás – cél az aktuális digitális tudás profilalapú értékelése, valamint előre definiált digitáliskompetencia-profiljavaslatok megalkotása. Az alapszintű digitális készségszintre előre definiált profil a DigKomp Állampolgári Bázis és a magasabb, haladó szintű digitális készségszintre az Állampolgári Plusz létrehozása.
- A DigKomp ÁDKK módszertani előkép az ágazati, szakmai keretekhez. A tervezett DigKomp Rendszer nemzetközileg is egyedi vonása, hogy az állampolgárok

digitális kompetenciájának fejlesztését szolgáló eszközök mellett a gazdaság számára szükséges ágazati és szakmai szintű digitáliskompetencia-fejlesztéshez is eszközöket kíván nyújtani. Erre szolgál az ágazati digitáliskompetencia-keretek kidolgozása, az azok alapján kidolgozandó szakmai digitáliskompetencia-elvárások megfogalmazása, illetve az azt segítő módszertani útmutató.



7. ábra A DigKomp ÁDKK – állampolgári digitáliskompetencia-keret, részlet

7. Összefoglaló

Kijelenthetjük, hogy a népesség digitális kultúrájának és gyakorlati készségeinek fejlesztése hosszú távú feladat. A digitális technológia gyors fejlődése és annak elterjedése ma már potenciálisan az egész lakosságot érinti. A DigKomp Rendszer egyidejűleg képes a digitálisan legkevésbé jártas népesség kompetenciafejlesztésére, de a magas felkészültségi szinttel rendelkezőket is további tanulásra és fejlődésre ösztönzi. A hazai digitális tudás és jártasság fejlődése Magyarország DESI értékeinek fokozatos növekedésére, valamint a munkavállalók munkaerőpiaci elhelyezkedésének esélyeire is hatással van, továbbá támogatja a Nemzeti Digitalizációs Stratégia által 2030-ra megfogalmazott célértékeket is.



Bibliográfia

- Carretero Gomez Stephanie; Vuorikari Riina; Punie Yves. DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use. 2017. <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digcomp-21-digital-competence-framework-citizens-eight-proficiency-levels-and-examples-use>
- Centeno Mediavilla, I., Vuorikari, R., Punie, Y., Okeeffe, W., Kluzer, S., Vitorica, A., Lejarzegi, R., Martnez De Soria, I. and Bartolom, J., Developing digital competence for employability: Engaging and supporting stakeholders with the use of DigComp, Publications Office of the European Union, Luxembourg, 2019. <https://publications.jrc.ec.europa.eu/repository/handle/JRC118711>
- Ferrari, Anusca. DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe., 2013. <https://ec.europa.eu/jrc/en/publication/digcomp-framework-developing-and-understanding-digital-competence-europe>
- Ferrari, Anusca. DIGCOMP: A digitlis kompetencia rtelmezsnek s fejlesztsnek eurpai keretrendszer, 2013. https://www.deaweb.hu/images/bongeszde/digcomp_teljes_hun_151231.pdf
- The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution, World Economic Forum, 2016. <http://reports.weforum.org/future-of-jobs-2016/chapter-1-the-future-of-jobs-and-skills/#view/fn-1>
- Hodson, Deirdre. European Distance Learning Week: Tackling the digital skills gap in the EU – policies and programmes, 2016. https://pt.slideshare.net/eden_online/european-distance-learning-week-tackling-the-digital-skills-gap-in-the-eu-policies-and-programmes/4
- Insights into skills shortages and skill mismatch: Learning from Cedefop’s European skills and jobs survey, Cedefop. 2018. <http://www.cedefop.europa.eu/en/publications-and-resources/publications/3075>
- The next era of human|machine partnerships: emerging technologies’ impact on society & work in 2030. Institute for the Future for Dell Technologies, 2017. https://www.delltechnologies.com/content/dam/delltechnologies/assets/perspectives/2030/pdf/SR1940_IFTFforDellTechnologies_Human-Machine_070517_readerhigh-res.pdf
- Vuorikari Riina; PUNIE Yves; Carretero Gomez Stephanie; Van den Brande Godelieve. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model, 2016. <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digcomp-20-digital-competence-framework-citizens-update-phase-1-conceptual-reference-model>

A könyvtári szolgáltatások átalakulása és fejlesztése a pandémia idején

Békésiné Bognár Noémi Erika

Országos Széchényi Könyvtár, Könyvtári Intézet, Kutatási és Elemző Osztály
bognar.noemi@oszk.hu

Nagy Andor

Országos Széchényi Könyvtár, Könyvtári Intézet, Kutatási és Elemző Osztály
nagy.andor@oszk.hu

Abstract

The Covid19 pandemic has not only posed new difficulties to libraries but has accelerated the process of digitizing library services. Although different services, events and information sources online were only an option for libraries before, in 2020 it became essential for them to be able to fulfill their purpose even during obligatory closures: preserving human culture, making it accessible, providing information ensuring equal opportunities. Librarians all over the world re-evaluated their duties and tasks, the forms of communication and tools used.

In January 2021, we conducted a survey of Hungarian libraries at the Research and Analysis Department of the Hungarian Library Institute, during which we tried to explore how certain library services and background activities changed during the pandemic. Based on the obtained results, it can be said that the operation of libraries not only did not stop during the obligatory closure of library buildings, but many libraries began to develop electronic and remote services on a large scale and tried to ensure that they could serve their users during this period. Not only by making some information sources available online and re-prioritizing tasks, but also by digitizing events, trainings, exhibitions and mainly locally available services.

Keywords: library service, digitization, digital transformation, remote access, pandemic, covid19

1. Bevezetés

Az 1997. évi 140. törvény¹ megfogalmazása szerint, az információs társadalom és a demokratikus jogállam működésének alapfeltétele a könyvtári rendszer, amelynek alapvető feladata az információk szabad hozzáférhetőségének biztosítása. A 2020 elején kirobbanó világjárvány és az azt követő korlátozó intézkedések megnehezítették ezen feladat ellátását, ám a felméréseink alapján a szolgáltatás a kényszerű zárvatartás idején sem szünetelt. A könyvtárak olyan intézkedéseket hoztak, amelyek nem csupán újszerűségükkel tűntek ki, de sok közülük a könyvtári rendszer hosszútávú fennmaradását is támogathatja.

1 1997. évi CXL. törvény a múzeális intézményekről, a nyilvános könyvtári ellátásról és a közművelődésről, 2021. Hozzáférés: 2021. 05. 31. <https://net.jogtar.hu/jogszabaly?docid=99700140.tv>



A Könyvtári Intézet Kutatási és Elemző Osztályán 2021 januárjában egy felmérést készítettünk a magyarországi könyvtárak és KSZR szolgáltatóhelyek körében², amelynek során azt kíséreltük meg feltárni, hogy miképp alakultak át az egyes könyvtári szolgáltatások és háttértevékenységek a pandémia idején, így helyzetképet adva a magyarországi könyvtárak jelenlegi működéséről. Mostani tanulmányunk ennek a kutatásnak az előzményeit, a céljait, az alkalmazott módszereit és az eredményeit foglalja össze röviden.

A koronavírus-járvány eddig nem tapasztalt emberi és szakmai kihívások elé állított mindannyiunkat. Megváltoztak a munkakörülmények: mind a könyvtárosoké, mind az olvasóké. A korlátozások idején főleg az otthon tanuló diákokat, hallgatókat, az otthonról tanító pedagógusokat, tanárokat és a home officeban dolgozó felnőtteket kellett kiszolgálni. Mivel Magyarországon a könyvtári szolgáltatásokra a veszélyhelyzet idején is megmaradt az igény, így a hazai könyvtárak igyekeztek minden elérhető eszközt megragadni a könyvtárhasználók kiszolgálására. A 2020 márciusa óta tartó időszak legnagyobb részében fizikai tereiknek a bezárására kényszerültek a könyvtárak, azonban ez nem jelentett szolgáltatáskiesést. Az olvasókkal való kommunikáció hagyományos módjai és a korábban megszokott szolgáltatási rend nem működhetett tovább. Míg korábban az egyes szolgáltatások, a rendezvények és az információforrások online elérhetővé tétele csupán lehetőség volt a könyvtárak számára, addig a 2020-2021-es években már elengedhetetlen feltételévé vált annak, hogy a zárvatartás idején is megfelelhessenek küldetésüknek, ami az emberi kultúra megőrzése, hozzáférhetővé tétele, az információs esélyegyenlőség biztosítása.

A rendkívüli körülmények miatt, a könyvtárosok ahogy a világon mindenhol, úgy hazánkban is, újraértékelték szerepüket és a feladataikat, a szolgáltatási folyamatokat, a szolgáltatások fizikai és virtuális tereit, valamint a dokumentumok, információk eljuttatására alkalmazott kommunikációs formákat, eszközöket. Az ország könyvtárosainak felkészültsége, kreativitása és elhivatottsága révén új, a pandémiához alkalmazkodó megoldások születtek. A könyvtárak egy jelentős részének sikerült rugalmasan reagálni erre a nem várt helyzetre, és régi szolgáltatásaik átszervezésével, illetve újak kialakításával megtalálták azokat a lehetőségeket, amelyekkel a zárvatartás ellenére használóikat támogatni tudják, mind az analóg és mind a digitális tartalmak elérésében. Ebben az időszakban felértékelődött a távolról elérhető tartalom és a biztonságos könyvtárhasználat.

1.1 A Könyvtári Intézet szerepe a koronavírus könyvtárakra gyakorolt hatásának feltérképezésében

Az Országos Széchényi Könyvtár és a Könyvtári Intézet munkatársai, köztük mi is, arra törekszünk, hogy különböző módokon segítsük mind a könyvtáros szakma, mind az olvasók tájékozódását a pandémia idején, ezért 2020 márciusától nyomon követjük a koronavírus-járvány könyvtárakra gyakorolt hatását.

2 Békésiné Bognár Noémi Erika [et al.]: *A könyvtárak és a világiárvány hullámai: gyorsjelentés a magyarországi és a határon túli könyvtárak szolgáltatásairól*. Hozzáférés: 2021. 05. 31.
https://ki.oszk.hu/sites/default/files/hirfajlok/a_konyvtarak_es_a_vilagjarvany_hullamai.pdf

A járvány újabb hullámainak megjelenésével a teljes nyitás és a korábban megszokott működési, illetve szolgáltatási rendhez való visszatérés egyelőre még várat magára, így a járvány hatására bevezetett változtatások mélyebb vizsgálata időszerűvé vált. A most bemutatott kutatásunkat több, rövidebb felmérés előzte meg: 2020 áprilisában a koronavírus megjelenése után készítettünk egy kérdőíves vizsgálatot, valamint 2020 nyarán, a könyvtárak újrainyitási tapasztalatait is összegyűjtöttük és megosztottuk³. Az ismeretek bővítése, a mélyebb és előremutató elemzés érdekében 2021 januárjában egy újabb, reprezentatív felmérést végeztünk a hazai könyvtárak járványhelyzethez való alkalmazkodásáról. Ennek apropója az volt, hogy mindeddig kevés reprezentatív, kvantitatív jellegű publikáció született, inkább benyomásokat összegző, kvalitatív jellegűek készültek. A kvantitatív kutatások legtöbbje is leíró jellegű volt, néhol a könyvtárak működésére, néhol pedig a könyvtárosok munkakörülményeire, benyomásaira kérdeztek rá.

2. A kutatás célja, kérdései és módszerei

A kutatásban többet között arra kerestük a választ, hogy milyen lépéseket tettek a könyvtárak a felhasználók segítségének, elérésének, megtartásának érdekében, a digitális oktatást, az otthoni tanulást és munkavégzést hogyan tudták támogatni, milyen változások történtek a könyvtárakban a szolgáltatások működtetését, minőségét tekintve a járványhelyzet miatt. A könyvtárvezetőket megkérdeztük arról is, hogy véleményük szerint mely szolgáltatások hiányoztak az olvasóknak, milyen a szolgáltatások megváltozott formáinak megítélése a fenntarthatóság szempontjából, továbbá, hogy hogyan változott a bezárások idején az egyes háttérmunkákra szánt munkaidő, illetve humánerőforrás-ráfordítás a könyvtárakban. Ezeket a kérdéseket vizsgáltuk könyvtártípusok szempontjából is. Részletesebben kérdeztünk rá a hazai könyvtárak járványidőszak alatt nyújtott táv- és online szolgáltatásaira, valamint a vizsgált szolgáltatások országos elterjedtségét is górcső alá vettük.

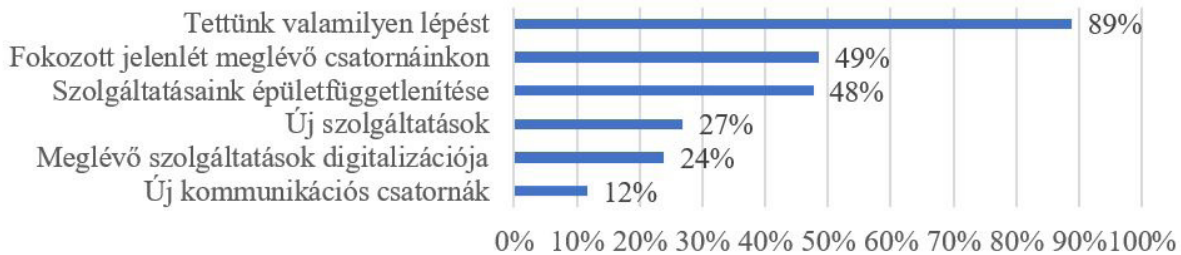
Az adatfelvételre online kérdőíves formában került sor 2021. január-február folyamán, minden, a 2019-es évről szóló könyvtári statisztikában szereplő könyvtárat és könyvtári szolgáltatóhelyet megszólítottunk. Az elemzés során összesen 689 választ tudtuk felhasználni. A minta nemválaszolásból adódó torzításait könyvtári funkciók és közigazgatási régiók szerinti súlyozással korrigáltuk. Az adatokat SPSS és PSCP programmal értékeltük ki.

3. A kutatás eredményei⁴

Kutatásunkban összegyűjtöttük azokat járványhelyzet idején hozott intézkedéseket, amelyekkel az olvasók találkozhattak, majd ezekből csoportokat alkottunk, és rákérdeztünk a válaszadóknál, hogy ezek közül melyeket alkalmazták a felhasználók elérésének, megtartásának érdekében 2020 márciusát követően.

3 A könyvtárak újrainyitása a koronavírus-járványt követően. Hozzáférés: 2021. 05. 31.
https://ki.oszk.hu/sites/default/files/dokumentumtar/covid_ujranyitas_0.pdf

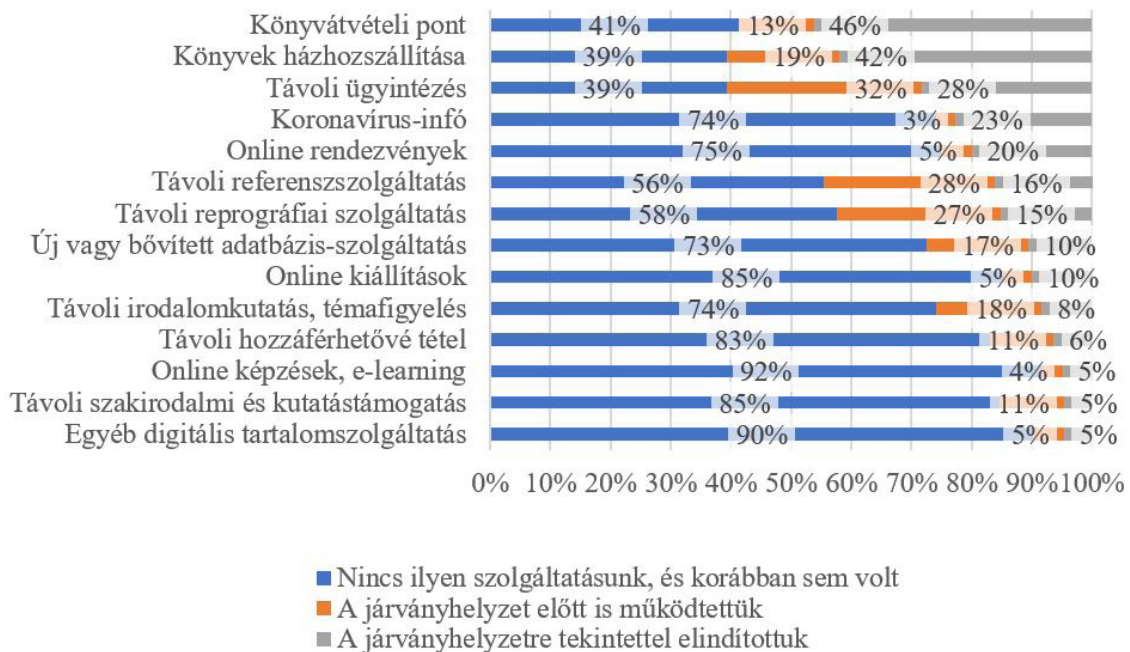
4 Előadásunkban előzetes adatokat közöltünk, ezért az eredmények pontosítását követően a jelen tanulmányunkban közzétett adatok kissé eltérhetnek az előadásban elhangzottaktól.



1. ábra. Lépések a felhasználók elérésének, megtartásának érdekében

A kapott válaszok alapján elmondható (1. ábra), hogy a könyvtárak 89%-a fejlesztette a szolgáltatásait vagy a kommunikációját ebben az időszakban, és mindössze csak 11%-uk nem tett semmilyen ez irányú lépést a járványhelyzet alatt. A legtöbben meglévő kommunikációs csatornáikon igyekeztek fokozni jelenlétüket, valamint megpróbálták szolgáltatásaikat épületfüggetlenné tenni.

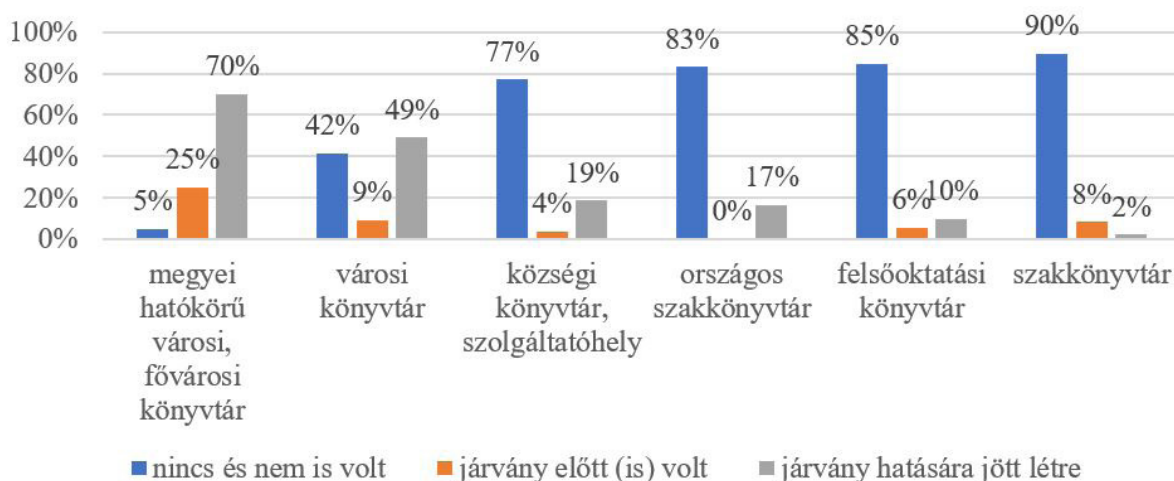
Az épületfüggetlen szolgáltatás a felmérésünk alapján 46%-ban a külső könyvtvételi pontok kialakítását jelentette (2. ábra). Ezek a könyvtárak épületén kívül elhelyezett átadó-átvételi pontok, ahol a megfelelő óvintézkedések betartása mellett kölcsönözhetnek az olvasók. A könyvtárosok előre összekészítik a kölcsönzéseket, az olvasóknak pedig csak át kell venniük azokat. A visszaérkezett könyveket néhány napra karanténba helyezik. Ezenfelül a könyvtárak 42%-a indított könyvházhozszállítás-szolgáltatást, amelynek keretén belül a könyvtárosok főként az idős vagy mozgásukban korlátozott olvasóknak viszik házhoz a kölcsönzéseiket. Továbbá a könyvtárak közel egyharmada a járvány hatására vezette be a távoli ügyintézés lehetőségét, ám itt meg kell jegyezni, hogy ennél sokkal több könyvtárban volt elérhető ez a mód, a kapott érték csak azokra a könyvtárakra vonatkozik, amelyknél ez egy új szolgáltatás volt.



2. ábra. Változások a szolgáltatások működtetését tekintve

Évek óta nagyon látványos tendencia, hogy a könyvtárak egyre több és több rendezvényt szerveznek. Ezeknek a száma az országos könyvtári statisztika⁵ alapján 2019-ben már átlépte a 122 ezret, és több mint 4,5 millió látogatót vonzottak. Ide sorolhatók többek között az IT-tanfolyamok, az író-olvasó találkozók, a mesedélutánok, a baba-mama klubok, sőt még a robotika foglalkozások is.

A járvány a rendezvényeknek a számát jelentősen visszavetette, viszont a felmérésünk alapján a könyvtárak egy része rendezvényeit továbbra is megtartotta, ám könyvtártípusonként eltérő arányban. Ez az arány az egyes könyvtártípusok vonatkozásában eltérő (3. ábra), a megyei hatókörű városi könyvtárak, köztük a fővárosi könyvtárral, több mint 70%-a, a városi könyvtáraknak pedig több mint 49%-a kezdett online rendezvényszervezésbe a bezárást követően. Természetesen rajtuk kívül is vannak könyvtárak, amelyekben tartottak online rendezvényeket, de esetükben nem új szolgáltatásról beszélünk, hiszen korábban is szerveztek programokat ilyen formában (ld. 3. ábra).



3. ábra. Online rendezvények a könyvtárakban

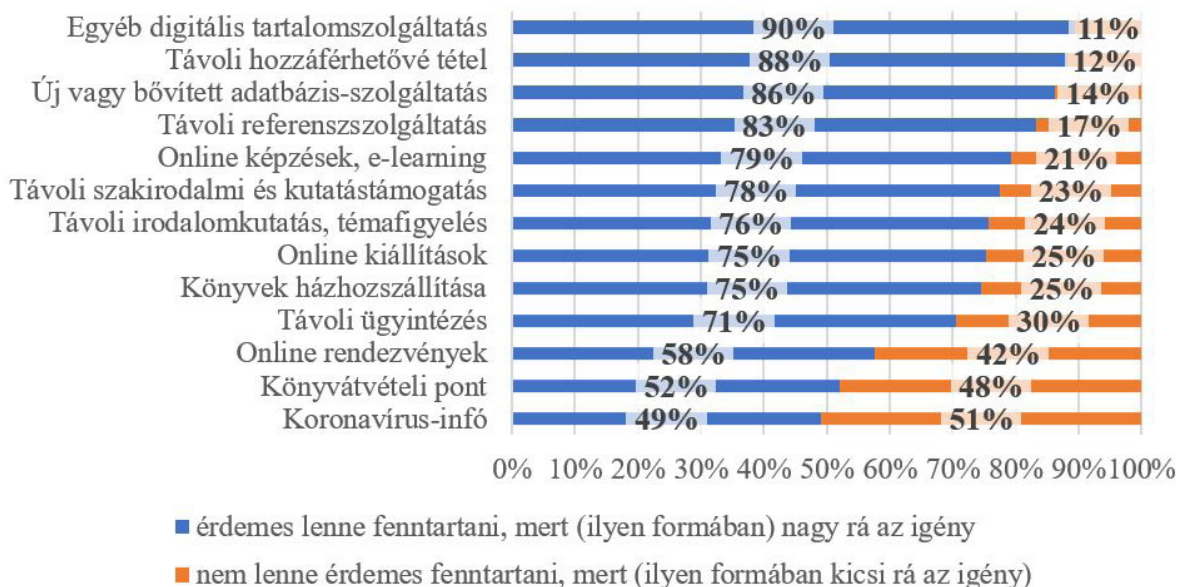
Már a járvány előtt is jól megfigyelhető folyamat volt, hogy az élet egyre több területe digitalizálódik, tehát vagy a meglévő szolgáltatások költöznek át a digitális térbe vagy teljesen újak jönnek létre. A pandémia ezt a folyamatot jelentősen felgyorsította, ám sok esetben kényszer szülte megoldásról van szó, ami nem biztos, hogy tartósan elérhető marad az olvasók számára. Megkérdeztük a kérdőívünket kitöltő könyvtárvezetőket, hogy ők erről mit gondolnak. Felsoroltuk a legnépszerűbb, járvány idején működő könyvtári szolgáltatásokat, és kértük, hogy ezekről döntsék el, mi lesz a sorsuk a pandémiát követően.

A legtöbben, a válaszadók 90%-a, úgy gondolja, hogy a viszonylag újszerű digitális tartalomszolgáltatásokat (pl. blog, podcast) érdemes lesz megtartani, mert nagy rájuk az igény (4. ábra). Hasonlóan vélekednek a távoli hozzáférés lehetőségéről (ide elsősorban a VPN-en és proxyn keresztüli tartalomelérés lehetőségét kell érteni), illetve a könyvtárak

5 Országos könyvtári statisztika, 2020. Hozzáférés: 2021. 05. 31.
<https://ki.oszk.hu/sites/default/files/dokumentumtar/stat19hiv.xlsx>



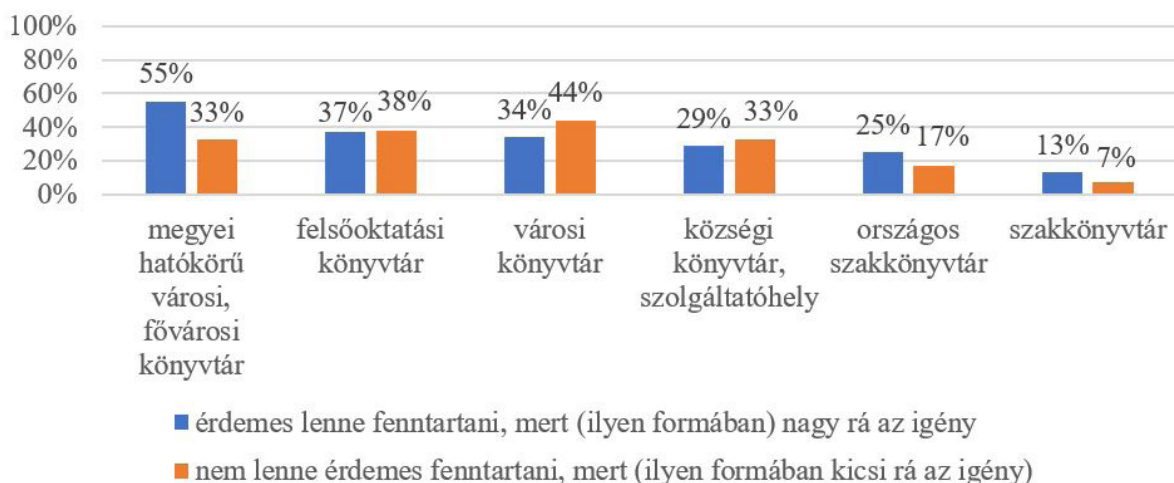
által épített adatbázis-szolgáltatásokról. A pozitív megítélést tekintve azonban ezekről nem sokkal maradnak el a távoli reprográfiai, a kutatástámogató, és a referenz szolgáltatások, valamint az online képzések sem.



4. ábra. Mit gondol, az Önök könyvtárában az alább felsorolt szolgáltatásokat érdemes lenne-e fenntartani a jelenlegi formájukban a járványhelyzet elmúltával is?

Talán a legkézzelfoghatóbb járvány szülte könyvtári szolgáltatások, amelyekkel a média is gyakran foglalkozott, a külső könyvtvételi pontok és a könyvházhozszállítás szolgáltatások voltak. A felmérésünk alapján jó esély van arra, hogy ezek a szolgáltatások – bár a kényszerű zárvarartás miatt jöttek létre – a könyvtárak egy részében a pandémia után is megmaradnak.

A könyvtvételi pontokkal kapcsolatosan a legnagyobb arányban a megyei hatókörű városi könyvtárakban és a fővárosi könyvtárban gondolják ezt így (55%), ám amint az 5. ábrán látható, minden könyvtártípusból érkeztek olyan válaszok, amelyek alapján azt gondolhatjuk, hogy a szolgáltatás sok helyen a jövőben is elérhető lesz.

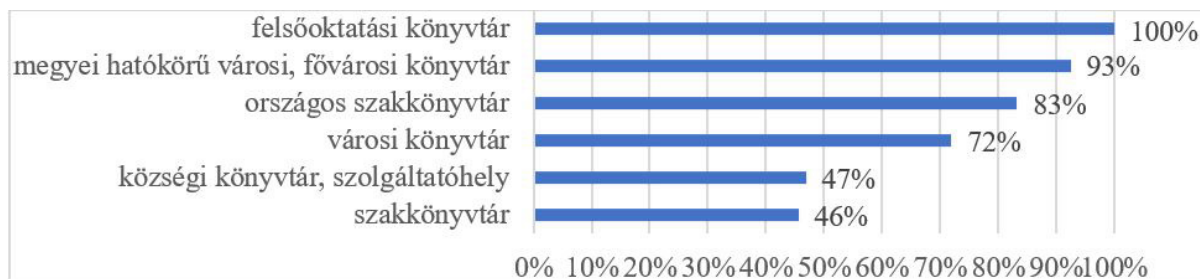


5. ábra. Érdemes lenne-e fenntartani a könyvtárvételi pontokat?

Az összes szolgáltatás megítélését összevetve tehát úgy tűnik, hogy sok könyvtárvezető, akiknek van már tapasztalata a távoli-, illetve online szolgáltatásokkal, hosszútávon is értékes szolgáltatásoknak tartja ezeket.

A könyvtárak kapcsán fontos kérdés volt, hogy miképp fogják tudni támogatni az oktatást, a tanulást és a munkavégzést a járványhelyzet idején. A könyvtárak e szerepe jelentős, amellett, hogy foglalkoznak formális és nonformális oktatással, teret biztosítanak a tanulásnak, munkavégzésnek, illetve mindezt egyéb szolgáltatásaikkal is támogatják. Felmérésünk alapján a könyvtárak jelentős része igyekezett segíteni az otthoni oktatást, tanulást és munkavégzést.

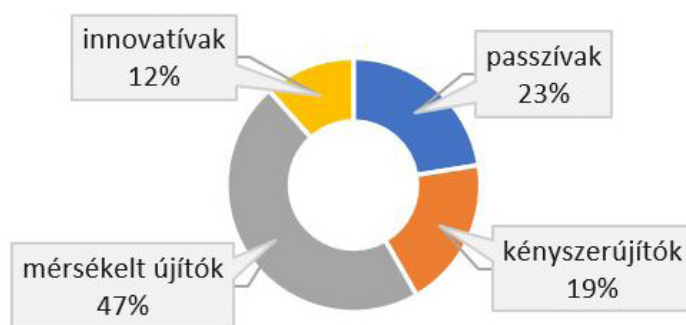
Amint az a diagramon látszik (6. ábra), a felsőoktatási könyvtárak 100%-a válaszolta, hogy tett lépéseket ez irányba, őket követték 93%-kal a megyei hatókörű városi könyvtárak és a fővárosi könyvtár, a sorban a következők az országos szakkönyvtárak voltak, majd a városi, a községi könyvtárak, végül az egyéb szakkönyvtárak. A válaszokból leszűrhetjük, hogy a legtöbben a tanulási felületekről szolgáltattak információt, dokumentumokat digitalizáltak és meglévő tartalomszolgáltatási felületeiket bővítették.



6. ábra. Tettünk lépéseket az online tanulás és munkavégzés támogatása érdekében.

Felmérésünk egy másik érdekes pontja azzal foglalkozott, hogy a megváltozott helyzet hogyan hatott a munkaszervezésre és háttérmunkákra. Kitöltőinknél arra kérdeztünk rá, hogy mely háttérmunkára mennyi időt fordítottak a kényszerű zárva tartás idején, és a válaszok alapján elmondható, hogy olyan tevékenység nem volt, amelyre lényegesen kevesebb időt szántak, viszont voltak olyanok, amelyekre több jutott. Ilyen volt az állományrendezés, az állománygyarapítás, az apasztás és a leltározás, tehát tipikusan azok a tevékenységek, amelyeket könnyebb elvégezni, ha nincsenek olvasók az épületben.

A kutatás során létrehoztunk egy, a könyvtárak aktivitásának megállapítására szolgáló tipológiát is, amely egy változóban szemlélteti nemcsak a könyvtárak járványhelyzetre adott reakcióit, hanem a korábbi aktivitásukat a táv- és online szolgáltatások létrehozása terén (a kérdőívben adott válaszaik alapján, amelyek a már működtetett, az újonnan indított, illetve bővített szolgáltatásokra vonatkoztak). Négy elméleti típus különböztettünk meg egymástól: passzívak, kényszerűjítók, mérsékelt újítók, innovatívak.



7. ábra. A könyvtárak aktivitása a táv- és online szolgáltatások terén

Eszerint (7. ábra) a magyarországi könyvtárak 47%-a mérsékelt újító: tehát a járvány előtt is nyújtottak néhány táv-, illetve online szolgáltatást, de ezek számát a járványhelyzet hatására legfeljebb 2 új szolgáltatás bevezetésével egészítették ki. Kevésbé jellemző a könyvtárakra, hogy a járvány előtt egyetlen egyet sem biztosítottak az általunk felsorolt táv-, illetve online szolgáltatások közül és a járvány hatására is csak legfeljebb kettőt indítottak el (passzívak, 23%), valamint az is, hogy a járvány hatására hármat vagy akár többet is elindítottak (kényszerűjítók, 19%). A legkisebb százalékban mondhatjuk innovatívnak (12%) a könyvtárakat (a koronavírus járványt megelőzően is legalább 3 szolgáltatást biztosítottak, de a megváltozott környezet kihívásaira reagálva ezek számát még legalább 3 új bevezetésével bővítették).

4. Összefoglalás

A pandémia járulékos hatásai minden jel szerint felgyorsították a táv-, illetve online szolgáltatások terjedését, és olyan új szolgáltatások létrehozására ösztönözték a könyvtárakat, amelyek vagy csak jóval később, vagy egyáltalán nem jöttek volna létre. Annak érdekében, hogy a könyvtárak olvasóikat a zárva tartás idején is képesek legyenek elérni és kiszolgálni, nagymértékben fokozták online kommunikációjukat, valamint a kölcsönzést és az ezzel kapcsolatos ügyintézkést próbálták épületfüggetlenné tenni,

hagyományos szolgáltatásaik jelentős részét digitalizálták. Mindezek mellett fontos megjegyezni, hogy hiába fektettek a könyvtárak jelentős energiát újfajta szolgáltatásaik kialakítására, egyes funkcióik elkerülhetetlenül csorbultak. Ezek elsősorban a közösségi funkciók, és azoknak a könyvtárhasználóknak a kiszolgálása, akik számára nem, vagy csak korlátozottan érhetőek el IKT eszközök és az internet. Jó példa erre, hogy az aprófalvak könyvtári ellátásáért felelős könyvtárbuszok a pandémia egy jelentős időszakában nem közlekedtek. A kialakult helyzetnek minden bizonnyal vannak hosszútávon is pozitív járulékos következményei, viszont az alapvető könyvtári funkciók teljes körű ellátása nem lehetséges kizárólag a távoli elérésű szolgáltatások biztosítása révén.

5. Bibliográfia

1997. évi CXL. törvény a muzeális intézményekről, a nyilvános könyvtári ellátásról és a közművelődésről, 2021. Hozzáférés: 2021. 05. 31.

<https://net.jogtar.hu/jogszabaly?docid=99700140.tv>

Békésiné Bognár Noémi Erika [et al.]: *A könyvtárak és a világvilágjárványhullámjai: gyorsjelentés a magyarországi és a határon túli könyvtárak szolgáltatásairól.*

Hozzáférés: 2021. 05. 31. https://ki.oszk.hu/sites/default/files/hirfajlok/a_konyvtarak_es_a_vilagjarvany_hullamai.pdf

A könyvtárak újrainyítása a koronavírus-járványt követően. Hozzáférés: 2021. 05. 31.

https://ki.oszk.hu/sites/default/files/dokumentumtar/covid_ujranyitas_0.pdf

Országos könyvtári statisztika, 2020. Hozzáférés: 2021. 05. 31.

<https://ki.oszk.hu/sites/default/files/dokumentumtar/stat19hiv.xlsx>



Kutatási adatkezelést támogató szolgáltatás a Szegedi Tudományegyetemen

Zeller Rozália
SZTE Klebelsberg Könyvtár
rozalia.zeller@ek.szte.hu
ORCID: [0000-0003-2501-8760](https://orcid.org/0000-0003-2501-8760)

Hoczopán Szabolcs
SZTE Klebelsberg Könyvtár
szabolcs.hoczopan@ek.szte.hu
ORCID: [0000-0002-7892-9974](https://orcid.org/0000-0002-7892-9974)

Nagy Gyula
SZTE Klebelsberg Könyvtár
gyula.nagy@ek.szte.hu
ORCID: [0000-0002-8391-2851](https://orcid.org/0000-0002-8391-2851)

Following the national and international trends in mid-2020 the Klebelsberg Kuno Library of the University of Szeged has also started to deal with the issue of research data management. After thorough self-training the library staff studied the Hungarian and international best practices of managing research data. We tried to assess the needs of the institutional research data management habits and the opinion of the researchers of SZTE with a comprehensive questionnaire. We compiled a comprehensive questionnaire to assess the needs of our researchers, learn what they're thinking about RDM and what kind of practices regarding RDM already exist in the research community.

By evaluating the questionnaire we have determined the areas in which the library could provide professional assistance where there was a real need among researchers. Keeping in mind the needs of the research community of University of Szeged we have decided to develop the following services: copyright consulting, RDM trainings for PhD students, theoretical and methodological assistance for RDM, write institutional FAIR data management recommendations. The last four services have been successfully implemented. We also wrote a feasibility study to assess the possibilities of developing our own institutional data repository.

Keywords: research data management, data management plan, FAIR, data repository

1. Bevezető

A kutatási adatok kezelésének (Research Data Management, a továbbiakban RDM) egyre fontosabb szerep jut az Open Access mozgalom Open Science modellé való átalakulásában. Az akadémiai szféra szereplői felismerték, hogy napjainkban már nem csupán az egy-egy folyóiratcikkekben testet öltő tanulmány jelentheti a kutatások végtermékét, hanem az odavezető út során keletkező anyagok szisztematikus



dokumentálása, tárolása, megosztása, esetleges újr felhasználása legalább annyira fontos lehet az Open Science megközelítés szerint. Mindez azonban – mind egyéni, mind intézményi szinten – tudatosan végiggondolt kutatási adatmenedzsment tevékenység, avagy erre vonatkozó stratégia nélkül teljesen reménytelen vállalkozás. Egyre több európai (H2020, ERC) és hazai (NKFIH, OTKA) pályázati rendszerben követelmény a kutatási projektek során történő kutatási adatkezelés formális leírása.

2. RDM az SZTE Klebelsberg Könyvtárban

Egy EFOP 3.6.1. projektünk keretében az SZTE Klebelsberg Könyvtár munkatársaiként a bevezetőben vázolt problémakörre való hatékony válaszok kidolgozását tűztük ki célul a Szegedi Tudományegyetem (SZTE) polgárai számára, amely konkrét kutatási adatmenedzsment támogató szolgáltatások formájában öltött testet.

A kutatási adatok kezelésével kapcsolatos hasznos és naprakész tudnivalókat a Szerzői Eszköztár weboldalon, a kapcsolódó közösségi média platformjainkon, a könyvtár munkatársai által működtetett szakreferenci hálózaton és a további bejárattott kommunikációs csatornáinkon (kari hírlevelek, egyetemi honlap) keresztül kívántuk eljuttatni az SZTE kutatói közösséghez.

Első lépésben kijelölt munkatársaink online képzések, illetve a téma nemzetközi és hazai szakértőinek segítségével sajátították el a kutatási adatok kezelése kapcsán felmerülő legfrissebb nemzetközi trendeket, tudnivalókat. Kapcsolódtunk a HRDA¹ (Hungarian Research Data Alliance), a HUNOR (HUNGarian Open Repositories) és az NKFI Hivatal szakmai támogatásával tartott meetup sorozathoz.

Második lépésben, egy RDM témára vonatkozó kérdőíves kutatás keretében felmértük az egyetem kutatóinak körében tapasztalható elvárásokat, igényeket. Ezen felmérés eredményeire támaszkodva egészítettük ki az SZTE Klebelsberg Könyvtár eddigi szerzőtámogató szolgáltatásait: a Szegedi Tudományegyetem igényeihez adaptálva megterveztünk, majd elindítottunk egy, a kutatási adatok kezelésére vonatkozó szakértői tanácsadást.

Ennek keretében támogatjuk az SZTE szerzőit az egyre több pályázati kiírásban elvárt kutatási adatkezelési tervek (Data Management Plan) szakszerű elkészítésében. Segítséget adunk az egyetem szerteágazó kutatási portfóliójában keletkező, rengeteg különféle típusú kutatási adat hosszú távú, biztonságos megőrzése tekintetében, továbbá a kutatási adatok publikálására és repozitálására vonatkozó kérdések megválaszolásában.

A projekt keretében megvizsgáltuk egy SZTE-s intézményi adatrepozitórium felállításának lehetőségét, megvalósíthatósági tanulmányban jártuk körül ennek előnyeit-hátrányait, a megvalósítás kapcsán felmerülő nehézségeket, szóba jöhető megoldásokat. Egy ilyen adatrepozitórium esetleges létrehozása jól illeszkedhet az SZTE Klebelsberg Könyvtár tíz éve épülő Contenta repozitórium-rendszerébe, azonban a koncepció összetettsége és a felhasználói igények pontos felmérése miatt a konkrét kivitelezést mindenképpen alapos tervezésnek kell megelőznie.

¹ Hungarian Research Data Alliance. Hozzáférés: 2021.05.30. <https://hrda.hu/about>

3. Önképzés a kutatási adatkezelésben

A kutatási adatok kezelésével való ismerkedésünk alapos önképzéssel kezdődött. Egyfelől igyekeztünk feltérképezni a hazai és nemzetközi helyzetet ezen a téren: mely szervezetek és milyen minőségben tevékenykednek RDM vonalon, mik a kutatási adatok kezelésére kialakult jó gyakorlatok, milyen elvárásokat és igényeket fogalmaztak meg a tudományos élet szereplői a kutatási adatok kezelésével kapcsolatban. Ezzel párhuzamosan az elmúlt egy-másfél évben kollégáink részt vettek minden olyan hazai szervezésű workshopon és webináron, amely az RDM-hez és annak bármely részterületéhez kapcsolódott (pl.: HRDA meetup sorozat, NWS stb.). Mindemellett szükségesnek láttuk kibővíteni, illetve elmélyíteni tudásunkat nemzetközi kontextusban is, ezért a szakirodalomban való elmélyedésen túl elvégeztünk egy RDM témájú kurzust is.

A Research Data Management Librarian Academy (RDMLA)² egy, a Canvas platformon ingyenesen elérhető MOOC kurzus, amely átfogóan és könyvtáros-centrikusan mutatja be az RDM egyes részterületeit. A kurzus oktatóinak többsége neves amerikai egyetemi és kutatóintézeti könyvtárak kutatási adatkezeléssel foglalkozó szakértő munkatársa, illetve intézményvezetője. A képzés jelenleg 11 modulból áll, és mindegyik modul az RDM más-más részterületét mutatja be, többek között: a kutatási adatkezelés alapjai, kutatási ökoszisztémák, könyvtári RDM szolgáltatások kialakítása, projektmenedzsment, adatvizualizációs és analitikus eszközök, online RDM platformok, RDM képzések szervezése, kutatási adatokra vonatkozó licenzek és copyright, adatok archiválása és gondozása. A modulok videóleckékből, ajánlott olvasmányokból, opcionális gyakorlati feladatokból és önellenőrző tesztekkel állnak. Az összes modul teljesítése után vizsgadíj ellenében záróvizsga tehető, amely az amerikai felsőoktatási rendszerben kreditet jelent - ezzel a lehetőséggel azonban mi már nem élünk.

4. Igényfelmérés az SZTE kutatóinak körében

4.1. A kérdőív elkészítése

2020 év végén egy átfogó kérdőívet állítottunk össze, hogy az SZTE kutatóinak körében felmérjük, milyen elvárásaik és igényeik vannak a kutatási adatkezelés terén, illetve mi a véleményük az RDM egyes aspektusairól. A kérdőív elkészítéséhez több magyar és nemzetközi hasonló témájú felmérés kérdéseit, illetve eredményeit vettük alapul, és igazítottuk a saját, SZTE-s adottságainkhoz. A kérdőív kitöltésére nagyjából egy hónap állt a kutatók rendelkezésére, ez idő alatt többször is promotáltuk a felmérést a különböző kommunikációs csatornáinkon. Végül összesen 116 kutató töltötte ki a kérdőívet - ez a részvételi arány megfelelt az előzetes elvárásunknak, még valamennyire meg is haladta azt.

A kérdőívünk 5 fő kérdéscsoportból állt, ezek: 1) Személyes adatok; 2) Formátumok, adattípusok, etikai és jogi kérdések; 3) Adattárolás, archiválás és adatvesztés; 4) Az adatok hozzáférhetősége, megosztása és újrahasznosítása; 5) Adatkezeléssel

² Research Data Management Librarian Academy (RDMLA). Hozzáférés: 2021.05.30.
<https://learn.canvas.net/courses/2719>



kapcsolatos igények. Terjedelmi okokból itt csak az egyes csoportok legfontosabb kérdéseire adott válaszok kiértékelését tudjuk bemutatni, a teljes kérdőív kiértékelését önálló tanulmányban³ tettük közzé. Emellett fontosnak gondoltuk, hogy jó példával járjunk elől, ezzel is növelve mondanivalónk hitelességét, ezért a tanulmány alapjául szolgáló felmérés kutatási adatait szabadon hozzáférhető módon magunk is elhelyeztük egy adatrepozitóriumban⁴.

4.2. A kérdőív kiértékelése

A kitöltők legnagyobb része dolgozói státusza alapján docens (31%) és adjunktus (25%) volt. Ez okozott némi meglepetést, mert a PhD hallgatók nagyobb arányára számítottunk, ők viszont csak a kitöltők 12 százalékát jelentették. A másik meglepő eredményt a kitöltők tudományterületi megoszlásánál tapasztaltuk: a válaszadók fele a bölcsészettudományok területein tevékenykedik, ami tipikusan nem az első terület, ami a kutatási adatkezelés kapcsán bárkinek eszébe jut. Második és harmadik helyen a természettudományok és társadalomtudományok művelői képviseltették magukat, utánuk pedig az egészségtudományok, mindössze 8%-kal.

A tudományterületi megoszlás eredményeit tekintve nem meglepő, hogy a kitöltők leggyakrabban használt adattípusa szöveges (88%), illetve táblázatos (44%) formátumokból áll össze. A jövőre nézve biztatónak tekinthetjük, hogy a kitöltők igen nagy arányban (80%) saját magukat tartják felelősnek a kutatási adatok kezeléséért - bár ez is szorosan összefügg azzal, hogy a bölcsészettudományok területén leggyakrabban egyedül, vagy kevés társszerzővel publikálnak a kutatók. Jelentős mértékben tartják felelősnek a kitöltők az adott kutatás vezetőjét, illetve az egyetemet/intézetet a kutatási adatokért.

A leginkább használt adattárolási eszközök a várható módon oszlottak meg: a legtöbben (83%) saját számítógépen tárolják az adataikat; de a kitöltők csaknem fele jelölte meg a külső merevlemezt, az általános felhőalapú szolgáltatásokat (pl. Google Drive, One Cloud, stb.), illetve a pendrive-ot is. A speciálisan kutatási adatok tárolására szolgáló felhőalapú szolgáltatásokat csupán a kitöltők 4%-a használja, amely arány véleményünk szerint a közeljövőben várhatóan emelkedni fog. A kitöltők több mint fele rendszeresen készít biztonsági mentést a kutatása során (bár gyaníthatóan ennek egy része nem tudatos, hanem a különböző szövegszerkesztők automatikus mentési beállításából adódik). A kitöltők egyharmada ugyan csak néha, de készít biztonsági mentést; 3%-uk csak a kutatás lezárásakor, 6% pedig sosem. A válaszadók több, mint egyharmadánál fordult már elő adatvesztés a kutatás során, legtöbbször figyelmetlenség vagy hardverhiba miatt.

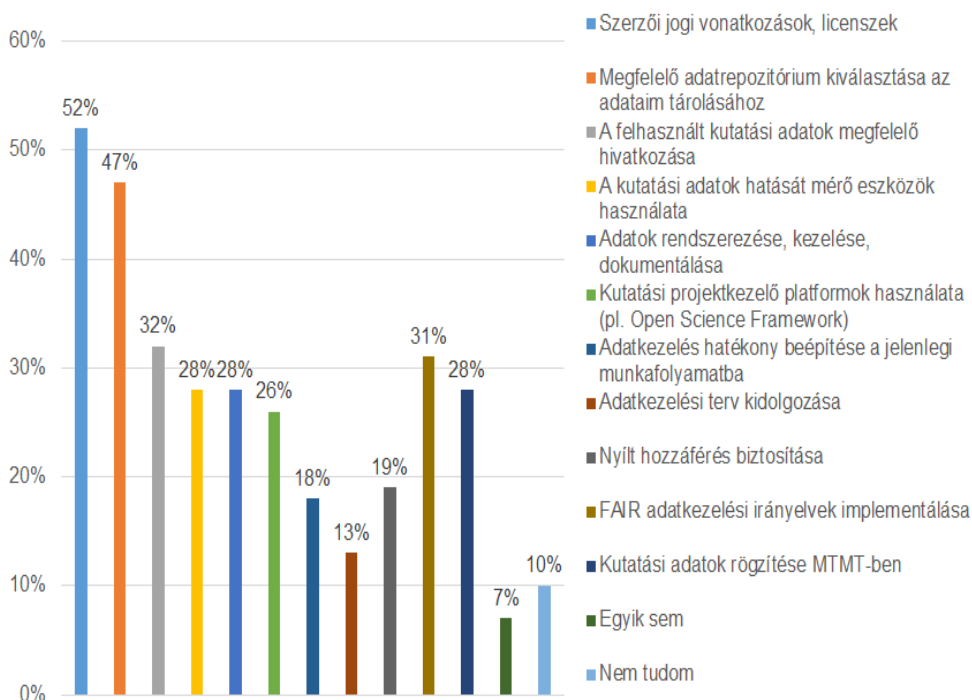
A kutatási adatok nyíltan elérhetővé tételére való hajlandóság teljesen szimmetrikus képet mutat: a kitöltők egynegyede mindig, fele néha, egynegyede pedig soha nem

3 Zeller Rozália, Hoczopán Szabolcs, Nagy Gyula. Kutatási adatkezelést támogató szolgáltatások előkészítése a Szegedi Tudományegyetemen. Tudományos és Műszaki Tájékoztatás, 68. 9. sz. (2021), 576-585.

4 Zeller Rozália, Hoczopán Szabolcs, Nagy Gyula. (2021). Kutatási adatkezelést támogató szolgáltatások előkészítése a Szegedi Tudományegyetemen kérdőív és válaszok [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5166625>

teszi nyíltan elérhetővé a kutatási adatait. A legtöbben (38%) a kutatói társadalom egésze számára teszik nyíltan elérhetővé az adatokat, de sokan jelölték meg a bárki, aki kérvényezi és a mindenki számára (20%) válaszlehetőséget is. Legkevésbé (3%) a saját intézményben dolgozó többi kutató számára tennék elérhetővé az adataikat, ami magyarázható talán az informális adatmegosztási módokkal, vagy a közeli konkurencia jelenlétével. A válaszokból kiderül, hogy a kitöltők a kutatási adatok megosztásának hagyományos formáit preferálják, vagyis a folyóiratcikkben való publikálást (88%); az intézményi (43%) vagy szakterületi (34%) adatrepozitóriumban való elhelyezést.

A kutatási adatkezelés mely területén/területein lenne szüksége segítségre?



1. ábra A kérdőív „Adatkezeléssel kapcsolatos igények” kérdéscsoportjának 5.1. kérdésére adott válaszok százalékos megoszlása

Könyvtáros szempontból az utolsó kérdéscsoport volt számunkra a legérdekesebb, mert itt mértük fel a kutatóink adatkezeléssel kapcsolatos konkrét igényeit, és a könyvtár irányában támasztott elvárásaikat. A fenti grafikonon jól látszik, hogy a válaszadók jelentős része szinte minden felkínált lehetőséget megjelölt. Örömmel láttuk azt is, hogy a kitöltők magas hányada fordulna tanácsért közvetlen kollégái (30%) mellett a könyvtárhoz (30%). Ezekkel szemben azonban ellentmondásos, hogy a válaszadók 15,5%-a egyáltalán nem venne részt kutatási adatkezeléssel kapcsolatos tréningen, amely ellentmondás áthidalása érdekes kihívást jelenthet a könyvtárosok számára.

5. Tájékoztató honlap létrehozása

Természetes volt, hogy az RDM-hez kapcsolódó tájékoztató anyagaink és szolgáltatásaink a Szerzői Eszköztár weblapunkon kapjanak helyet, hiszen jól illeszkednek az SZTE-s szerzőknek kínált szolgáltatásaink portfóliójába. A tartalom kialakításakor figyelembe



vettük, hogy a hazai nagy egyetemi könyvtárak milyen példákkal járnak előttünk, illetve az RDMLA kurzus során látóterünkbe került neves külföldi (amerikai és angol) egyetemi könyvtárak régóta szolgáltatott RDM anyagaiból is merítettünk ötletet.

A Szerzői Eszköztár Kutatási adatkezelés menüpontja két, jól elkülöníthető részre osztható: alapvető elméleti tájékoztató anyagok az RDM fő részterületeiről, illetve az ezekhez kapcsolódó szolgáltatásaink leírása. Az RDM kalauz a kutatási adatok életciklusát különbözőképpen leíró modelleket mutatja be, illetve a kutatási adatok kezeléséhez - úgymint fájlok elnevezése, mappastruktúra kialakítása, verziókezelés, biztonsági mentés, stb. - kínál praktikus tanácsokat a szerzőknek. A kutatási adatokra vonatkozó FAIR irányelvekről is készítettünk egy tömör összefoglalót, hasznos linkekkel kiegészítve.

6. Új szolgáltatások

A tájékoztató honlapok elkészítésével egy időben mérlegelnünk kellett, melyek azok a területek, ahol az igényfelmérő kérdőív szerint az egyetem kutatói számítanának a könyvtár segítségére, ugyanakkor rendelkezésünkre is állnak a megfelelő eszközök, valamint szaktudás.

A következő témákban indult el szolgáltatásunk az egyetem kutatói számára az SZTE Klebelsberg Könyvtár által, az MTMT és a publikációs támogató szolgáltatások miatt már jól ismert Szerzői Eszköztár⁵ felületünkön:

- Kutatási adatmenedzsment tanácsadás
- Kutatási adatkezelési terv konzultáció
- Adatrepozitórium ajánló
- Eszköztár Chat

6.1. Kutatási adatmenedzsment tanácsadás

Mivel a tudatos kutatási adatmenedzsment még viszonylag friss téma a magyarországi kutatói körökben, a szolgáltatás keretében általános kutatási adatkezelési témákban kívánunk segítséget nyújtani az egyetem kutatóinak. A szolgáltatást konzultációs jelleggel terveztük, hiszen nem minden esetben adhatóak egyszerű válaszok, a kutatónak és a könyvtárosnak mindenképpen közösen érdemes keresnie a megoldást. A kutató ismeri saját diszciplínájának specifikus adattípusait, munkafolyamatait és mérőműszereit. A könyvtáros pedig a konzultáció során a fenti információk birtokában segítséget tud nyújtani a kutatónak a kutatási adatok rendszerezésében, optimalizálásában, a metaadatok és kutatási dokumentáció létrehozásában, a kutatási adatok FAIR alapelveknek való megfeleltetésében, az ajánlások és jó gyakorlatok megtalálásában, valamint a biztonságos adattárolási módok kialakításában.

⁵ Szerzői Eszköztár. Hozzáférés: 2021.05.30. <http://szerzoknek.ek.szte.hu>

6.2. Adatkezelési terv konzultáció

A szolgáltatás keretében a pályázatokhoz kötelezően benyújtandó adatkezelési tervek kitöltésében kívánunk segítséget nyújtani a kutatóinknak. Kezdetből fogva lendületet adott a szolgáltatásnak, hogy meghirdetése környékén volt esedékes az OTKA és az Élvtal pályázatok benyújtása, melyeknek 2021-től már kötelező eleme lett az adatkezelési terv készítése. Csatlakoztunk az NKFIH-EKK-HUNOR tíz alkalmas meetup sorozatához⁶, mely során az adatkezelési terv kitöltéséhez szükséges elméleti és gyakorlati tudást próbáltuk átadni a pályázat kitöltőinek. A meetup sorozat jó alkalom volt számunkra, hogy sokféle tudományterület, változatos adatkezelési problémáival találkozassunk, és így saját ismereteinket is bővíthessük.

Ezt követően hamar felmértük, hogy a tíz alkalmas sorozatról nagyon sok kutató lemaradt a létszámkorlát miatt, így átvéve a debreceni kollégák kidolgozott előadásmenetét, az OTKA pályázatok beadási határideje előtt immár saját szervezésben indítottunk két létszámkorlát nélküli online találkozót kifejezetten a Szegedi Tudományegyetem számára. Amellett, hogy ilyen módon sikerült konzultációs lehetőséget biztosítani minden érdeklődőnek, jó alkalom volt, hogy a könyvtár új szolgáltatását éles körülmények között is bemutassuk az egyetemnek.

A saját szervezésű meetupokat követően – bár többnyire nem a meghirdetett szolgáltatási email címen – de folyamatosan érkeztek a megkeresések a véglegesnek szánt adatkezelési tervek utolsó ellenőrzésére és a még kétséges pontok tisztázására.

Az OTKA - Élvtal pályázatok, a még nem túlzottan összetett DMP-vel⁷, számunkra is ideális gyakorlási, betanulási lehetőséget biztosítottak. Nagy öröm volt, amikor az NKFIH pályázatokon túl, az első Horizon és FP7 adatkezelési tervvel is megkerestek minket kutatók és azt tapasztalhattuk, hogy itt is releváns segítséget tudunk nyújtani. Ugyanakkor az is teljesen nyilvánvalóvá vált, hogy az európai pályázatok adatkezelési tervei nem kezelhetők rutinból, a kutatóknak és könyvtárosoknak közösen kell összeállítaniuk azokat.

6.3. Adatrepozitórium ajánló

Az adatrepozitórium ajánló volt a legkézenfekvőbb szolgáltatás, amelyet el szerettünk volna indítani, hiszen ezt akár az általános publikációtámogató szolgáltatásaink közé is sorolhattuk volna. A publikációkhoz kapcsolódó, azokat igazoló datasetek repozitálása mellett, hogy egyre inkább pályázati, kiadói elvárás, a szerzőnek is alapvető érdeke. Így a publikáció és az adatcsomag is idézhető a kiadott DOI-k révén, ráadásul a két objektum azonosító össze is van kapcsolva, az adatcsomagra érkező hivatkozás számít a publikációnál is, és ugyanez érvényes fordítva is.

6 Kutatási adatkezelés meetupok – NKFIH-EKK-HUNOR szervezésében. Hozzáférés: 2021.05.30. <https://openscience.hu/events/kutatasi-adatkezeles-meetupok-nkfi-h-ekk-hunor-szervezeseben-10-exkluziv-alkalom>

7 Kutatási adat-kezelési terv „OTKA” K_21, FK_21, PD_21, ANN_21, SNN_21 pályázatok benyújtásához. Hozzáférés: 2021.05.30. <https://nkfi.gov.hu/palyazoknak/nkfi-alap/kutatasi-temapalyazat-k21/palyazati-csomag/kutatasiadat-kezelesi-terv-k21-fk21-pd21-ann21-snn21>



A megfelelő adatrepozitórium kiválasztása mégsem triviális feladat. Nemcsak a kutatók barátkoznak még a feladattal, de az adatrepozitórium infrastruktúra is igencsak épülőfélben van, így bizonyos tudományterületek még nincsenek lefedve repozitórium szolgáltatással. Az intézményi adatrepozitóriumok ugyancsak gyerekcipőben járnak még Magyarországon, így azokra sem támaszkodhatunk olyan mértékben, mint a publikációk esetében.

Mindezeket figyelembe véve elsősorban a legmeghatározóbb pályázati, fenntartói és kiadói ajánlások alapján próbálunk repozitóriumot ajánlani a kutatónak. Ha ezek nem adnak semmilyen támpontot, a DataCite DOI szolgáltató ügynökség által összeállított Re3Data repozitórium kereső⁸ segítségével folytatjuk a keresést, ahol a szűkítő facetták segítségével juthatunk el a kívánt találatokhoz. Ha esetleg nem találunk megfelelő diszciplináris repozitóriumot, akkor egy megbízható és minden igényt kielégítő multidiszciplináris adatrepozitóriumot javasolunk, mint például a Zenodo, vagy a Figshare.

Arra is volt példa, hogy szó szerint vett adatrepozitórium nem állt rendelkezésre, viszont a szűk tudományterületnek volt egy szinte kizárólagos adattára, melyet megkerülni a láthatóság érdekében úgysem lett volna érdemes. Ezen esetben fontos volt a kutató és a könyvtáros közötti együttműködés, mert célzott adatrepozitórium keresése közben az említett adattár nem került volna a látókörünkbe, a kutató pedig nem tudta, hogy az említett adattár repozitóriumként is felfogható.

6.4. Chat az Eszköztár honlapon

Az SZTE Klebelsberg Könyvtár fő honlapján⁹ már használatban lévő chatalkalmazást frissítést követően a Szerzői Eszköztár honlapunkon is bevezettük öt különböző témakörben, amelyek gyakorlatilag lefedik a teljes szolgáltatási körünket:

- MTMT szerzői támogatás
- MTMT adminisztrátori támogatás
- Kutatástámogatás
- Publikációs támogatás
- Szerkesztőségi támogatás

Elsődleges célunk a chatszolgáltatással a kutatókkal való kommunikáció felélénkítése és közvetlenebbé tétele volt. Különös tekintettel arra, hogy az utóbbi időben számos új szolgáltatással bővült a portfóliónk.

Be kell ugyanakkor látni, hogy ennek a tájékoztatási „műfajnak” is megvannak a maga korlátai. Az érintett témakörök gyakran pontos, személyre szabott útmutatásokat, több körös konzultációt igényelnek, melyre más csatornák sokszor alkalmasabbnak tűnnek. Ebben a formában a chat leginkább „forgalomirányító” funkciót tölthet be. Mivel mindegyik témakörben való tájékoztatás speciális szaktudást igényel és egyetlen kolléga

⁸ Registry of Research Data Repositories. Hozzáférés: 2021.05.30. <https://www.re3data.org>

⁹ Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára. Hozzáférés: 2021.05.30. <http://www.ek.szte.hu>

nem tudja ellátni egy-egy témában a teljes aktív chat időt, a szolgáltatás működtetése a rendelkezésre álló kollégák szélesebb körének bevonását teszi szükségessé.

7. Intézményi adatrepozitórium megvalósíthatóságának vizsgálata

Az elkészült megvalósíthatósági tanulmányban tárgyalt lehetőségeket összegezve elmondható, hogy a kutatási adatok repozitálása néhány éven belül elkerülhetetlenné válik a szegedi kutatók széles köre számára is. A tudományági jellegzetességeket és bevett gyakorlatot figyelembe véve azonban ez sok esetben nem egy intézményi repozitóriumban, vagy elsődlegesen nem ott fog megvalósulni. Bizonyos tudományágakban az adatok repozitálásának kérdése jóval előrehaladottabb állapotban van, mint másoknál, illetve egyes folyóiratok kiadói máris kész elvárásrendszerrel rendelkeznek. Ezekben az esetekben látható módon az intézményi repozitóriumnak csak másodlagos szerep juthat.

Azonban mindezen szempontok figyelembevételével is úgy gondoljuk, hogy lehet létjogosultsága egy helyi telepítésű, saját üzemeltetésű intézményi kutatási adatrepozitórium létrehozásának, hiszen egy ilyen rendszer lehetőséget biztosít a fent nevezett esetekben is a kutatási adatok másodlagos elhelyezésére. Ennél sokkal fontosabb azonban, hogy hiánypótló csatornát jelenthet az egyetem kutatói számára a kutatási adatok közzétételére azokban az esetekben, amikor erre nincs konkrét kiadói, szerkesztőségi elvárás és iránymutatás. Ezekben az esetekben a célokat szem előtt tartva jobb helyen lehetnek a kutatóink adatai egy saját üzemeltetésű, minőségbiztosított intézményi repozitóriumban összegyűjtve, mint szétszórva a világban ezerféle (ingyenes és fizetős) változó minőségű szolgáltatás között. Az intézményi adatrepozitóriumban gyűjtött adatok egyetemünk láthatóságát, presztízsét egyértelműen növelni tudnák a kutatási adatok kezelésének területén.

Mindezek miatt javasoltuk az SZTE számára egy intézményi kutatási adatrepozitórium létrehozását, az ehhez szükséges technikai-infrastrukturális és személyi feltételek biztosításával. A lehetőségek további tanulmányozása érdekében saját könyvtárunk számára pedig azt javasoltuk, hogy kerüljön kialakításra egy tesztkörnyezet, ahol a legígéretesebbnek tűnő Dataverse és/vagy Invenio RDM szoftvert alaposabban meg lehetne vizsgálni.



Intézményi, támogatói szabályozások kialakítása a nyílt tudomány kutatási adatkezelési céljaihoz

Holl András
MTA Könyvtár és Információs Központ
holl.andras@konyvtar.mta.hu
ORCID: [0000-0002-6873-3425](https://orcid.org/0000-0002-6873-3425)

A nyílt tudomány (Open Science) kulcsterülete a kutatási adatok kezelése. Az immár a kutatási gyakorlat részévé vált nyílt hozzáférés (Open Access) tekintetében léteznek hazai szabályozások, továbbá rengeteg nemzetközi ajánlás is segíti ezek finomítását, bővítését, azonban a kutatási adatkezelés területén még előttünk áll a szabályozások kialakítása.

Kulcsszavak: nyílt tudomány, kutatási adatok, FAIR, szabályozások

0.) A FAIR kritériumrendszer

A kutatási adatkezelés ma elfogadott paradigmáját a FAIR mozaikszóval jellemzik: Findable (megtalálható), Accessible (hozzáférhető), Interoperable (szabványos) és Reusable (újrafelhasználható)¹. A kutatási adatok nyilvánosságának korlátait felismerve (pl. személyes adatok védelme) a kutatási adatkezelés javasolt megközelítése a „nyílt, amennyire lehetséges; zárt, amennyire szükséges”.

A jól dokumentált, rigorózus adatkezelésen és archiváláson túl két alapvető oka van a kutatási adatok FAIR kezelésének: egyfelől égetően szükség van a kutatási eredmények reprodukálhatóságára, másfelől az adatok újra felhasználásával időt és pénzt takaríthatunk meg.

Két kulcs-elemet emelünk ki az adatok megfelelő kezeléséhez: az egyedi azonosítók (DOI, ORCID, projekt- és intézményazonosítók), valamint a széles körben elfogadott sémák, szótárak, ontológiák használatát.

1.) Kötelezettség szabályozások kialakítására

Jól működő kutatási adatkezelési gyakorlat csak megfelelő szabályozások mellett képzelhető el. Mivel minden tudományterület képviselői jelentősen eltérő adatokkal dolgoznak, a magasabb szintű – országos, egyetemi, kutatóhálózati, finanszírozói – szabályozásoknak szükségszerűen általánosnak kell lenniük: gyakran csak deklarálják a FAIR alapelvek betartásának szükségességét, az említett „nyílt, amennyire lehetséges...” alapelveket és előírják az alacsonyabb szintű szabályozások (angolul mandate) elkészítését. Ilyen lehet az intézményi, kutatócsoporti szabályozások vagy az adatkezelési tervek megkövetelése. A tudományos folyóiratok is szabályozhatják – és egyre gyakrabban szabályozzák is – a publikált kutatásokat alátámasztó adatok kezelését.

1 FAIR: <https://www.force11.org/group/fairgroup/fairprinciples>



Minden szabályozási szinten érdemes figyelembe venni a releváns nemzetközi mandátumokat (H2020², ERC³). Az európai országos szintű szabályozásokat a SPARC Europe és a Digital Curation Centre anyaga mutatja be (2020). Magyarországon az NKFIH Open Science Szakértői Bizottsága⁴ készített ajánlást a szabályozásra. Az adatok nyílt elérhetőségének biztosítására ajánlani kell minősített repozitóriumok használatát. Amennyiben a szabályozást készítő szervezetnek létezik nyílt hozzáférésre vonatkozó szabályozása, a kutatási adatkezelésre vonatkozó szabályozás beépíthető ebbe, vagy egységes nyílt tudomány szabályozás készítése javasolt.

2.) Adatkezelési tervek

Adatkezelési tervek témakörben több nemzetközi minta, ajánlás⁵ szakcikk és útmutató is található (Michener, 2015; Jones, 2011). Magyarországon az NKFIH pályázatok kiírói megkövetelik a nyertes pályázatokhoz az adatkezelési tervek készítését⁶.

Az adatkezelési terv élő dokumentum, amit a projekt során karban kell tartani, és a projekt befejezése után csatolni kell a projekt beszámolóhoz a keletkezett adatok listájával és az elérésükre is használható azonosítóval (DOI).

Az adatkezelési tervnek ki kell térnie az alábbiakra:

- adatok jellege, formátuma, mennyisége;
- adatok leírása;
- alkalmazott adatfeldolgozási eljárások;
- hosszú távú megőrzés, egyedi azonosítók használata;
- kockázatok (etikai, személyes adatok kezeléséből adódó, stb.);
- hozzáférhetőség, felhasználási licenc;
- adatkezelés költségei;
- adatok felhasználhatósága, esetleges értékesíthetősége;
- nyilvános, projekten kívül keletkezett adatok használata;
- ... és minden egyéb releváns körülményre.

Az adatkezelési terv alternatívája a szakterületi adatkezelési protokoll (Domain Data Protocol, DDP)⁷. Az intézményi szabályozások hivatkozhatnak a DDP-re, vagy – a kutatócsoportok munkáját megkönnyítendő – adhatnak intézményi alap/keret adatkezelési tervet is.

2 H2020: <https://www.openaire.eu/how-to-comply-to-h2020-mandates-for-data>

3 ERC: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

4 NKFIH: <https://nkfi.gov.hu/hivatalrol/open-science-szakertoi/open-science-szakertoi>

5 H2020: DOI: [10.5281/zenodo.2635768](https://doi.org/10.5281/zenodo.2635768); ERC: <https://erc.europa.eu/content/erc-data-management-plan-template>

6 OTKA: <https://nkfi.gov.hu/palyazoknak/nkfi-alap/kutatasi-temapalyazat-k21/palyazati-csomag/kutatasiadat-kezelesi-terv-k21-fk21-pd21-ann21-snn21>

7 DDP: https://www.scienceurope.org/media/nsxdyvqn/se_guidance_document_rdmps.pdf

Az adatkezelési tervek szabályozásához javasoljuk – az amúgy igen különböző filozófiájú – ERC és HorizontEurope adatkezelési terv űrlapok⁸ tanulmányozását, de léteznek online eszközök is, amik az adatkezelési terv elkészítésében jelentős segítséget nyújtanak⁹.

3.) Publikus adatok hivatkozása, jelentése, nyilvántartása

A publikus adatok hivatkozása, jelentése, nyilvántartása tekintetében az eddigieknél jóval kevesebb támpontot találhatunk. Alapvetés, hogy a kutatási adatokat el kell látni állandó azonosítóval (pl. DOI), a nyilvános adatok felhasználásakor pedig a szükséges hivatkozásoknak az adatállomány azonosítóját is tartalmaznia kell. Amennyiben az adatokra alapuló publikációt közlő folyóiratnak van erre vonatkozó előírása, azt kell követni, egyéb esetben a cikkben hivatkozni kell a más kutatásokból, adatbázisokból felhasznált adatokra, amennyiben van, DOI azonosítóval. Ha a kutatáshoz felhasznált adatok egy dinamikus frissülő publikus adatbázisból való SQL lekérdezéssel keletkeztek, meg lehet fontolni a lekérdezett adatok archiválását és DOI-val való ellátását. Ez a kutatás reprodukálhatósága miatt is lényeges.

A saját kutatási adatoknál hivatkozni kell a nyilvános repozitóriumban, adatbankban elhelyezett adatok DOI azonosítójára. Mind a saját, mind az újra felhasznált adatoknál javasoljuk a hivatkozások elhelyezését a cikk irodalomjegyzékében, ahonnan a hivatkozáskapcsolatok gyűjtése könnyebben megvalósítható, mint a szövegtestből, ráadásul az irodalomjegyzékek kezelése sok adatbázisban automatikus. Más kutatók adatainak felhasználásánál általános gyakorlat az adatokra épülő cikkekre való hivatkozás, pedig a közvetlen adathivatkozások használatának vannak előnyei.

A kutatási jelentésekben a publikációk mellett a közzétett kutatási adatok elérhetőségét is fel kell sorolni a DOI azonosítókkal együtt. A felsorolást hazai beszámolási kötelezettség esetén célszerű az MTMT-ben tárolt adatokra való hivatkozással megtenni, amire az MTMT lehetőséget is ad: a publikus kutatási adatok felvihetők a rájuk kapott hivatkozásokkal együtt, majd az egyéni kutatók összesítő táblázatában ezek az adatok is megjelennek.

4.) Tudományértékelési szerep, mutatók

Amíg a kutatási adatok (általában a publikációkon túlmenő egyéb, nyilvánosságra hozott kutatási termékek) mennyisége és hatásmutatói (hivatkozások) nem számítanak a tudományos tevékenység értékelésében, addig a kutatók nem lesznek ösztönözve ezen termékek (esetünkben az adatok) nyilvánosságra hozatalára (bár a finanszírozók és a folyóiratok esetleges ilyen irányú követelményei segíthetnek). Mind a kutatási alapoknak, mind a folyóiratoknak célszerű a kutatási adatok kezelésének kérdését szabályozniuk, hozzájárulva a publikált kutatási adatok és a rájuk kapott hivatkozások értékének megalapozásához.

8 ERC: <https://erc.europa.eu/content/erc-data-management-plan-template>
H2020: https://zenodo.org/record/2635768#.YEsVEP4o_0o

9 Mint a Digital Curation Centre által készített DMPTool: <https://dmptool.org/>



Célszerű mind egyéni, projekt és intézményi szintű értékeléseknél figyelembe venni az MTMT-ből kinyerhető teljesítménymutatókat, továbbá a kutatási adatokra vonatkozó hivatkozásokat ugyanúgy értékelni, mint a publikációkét. Az értékelési rendszerről az Európai Bizottság dokumentumot készített (2017).

5.) Infrastruktúrák biztosítása

A kutatási adatok megőrzése és szolgáltatása hosszú távra biztosítandó feladat. Infrastruktúrákat és szervezeti egységeket kell létrehozni, amik hosszú távú működését megfelelő szabályozással biztosítani kell. A létrejött adatrepozitóriumokat el kell helyezni az intézményi struktúrában és költségvetésben. Működésüket szabályozni kell: a kutatók csak akkor tudják a FAIR követelményeket teljesíteni, amennyiben ennek megvannak az intézményes feltételei, amik feltételei nyilvánosak, szabályozottak. Az adatrepozitórium használata szerződés létrejöttével jár a kutató és a repozitóriumot működtető intézmény között, még ha konkrét aláírás nem is történik. A repozitóriumnak közzé kell tennie a megőrzésre vonatkozó vállalásait, az adatelhelyező számára kötelező nyilatkozatot, melyet az adatrepozitórium használatának kezdetekor el kell fogadni.

A repozitóriumok felé való bizalom megteremtéséhez/fenntartásához szükség van a repozitórium hosszú távú stabilitásának biztosítására, ami megjelenhet a fenntartó intézmény szabályozásaiban, pl. az SZMSZ-ben és egy külső, minőségbiztosító szervezet által kiadott tanúsítványban. Ilyen tanúsítványt Magyarországon az MTMT Repozitórium-minősítő bizottság¹⁰ ad, általánosan elfogadott nemzetközi tanúsítványt pedig a CoreTrustSeal¹¹.

6.) Kivételkezelés, embargók, fokozatosság

Minden szabályozás a kivételek, felmentések kezelésével lesz teljes. Lehetőséget kell teremteni a szabályozás alól való felmentésre indokolt esetben, megfelelő nyilvánossági embargók alkalmazását kell biztosítani, adott esetben a nemzetközi gyakorlattól eltérő türelmi időszakok megadásával, és a szabályozásokat fokozatosan kell bevezetni.

Köszönetnyilvánítás: a szerző hálás Harnos Noéminek és a tanulmány másik bírálójának a szöveg javításáért.

¹⁰ MTMT: <https://www.mtmt.hu/repozitoriumminosito-szabizottsag>

¹¹ CoreTrustSeal tanúsítvány: <https://www.coretrustseal.org/>

Irodalom

EC „Evaluation of Research Careers fully acknowledging Open Science Practices; Rewards, incentives and/or recognition for researchers practicing Open Science”, 2017, DOI: [10.2777/75255](https://doi.org/10.2777/75255)

Jones, Sarah „How to Develop a Data Management and Sharing Plan.” DCC, 2011. <https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>

Michener, William K. „Ten Simple Rules for Creating a Good Data Management Plan.” PLoS computational biology vol. 11,10 e1004525. 22 Oct. 2015, DOI: [10.1371/journal.pcbi.1004525](https://doi.org/10.1371/journal.pcbi.1004525)

SPARC Europe & DCC „An Analysis of Open Science Policies in Europe.” v6. DOI: [10.5281/zenodo.4005612](https://doi.org/10.5281/zenodo.4005612)



Born digital workflow tervezése a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában

Kalcsó Gyula

Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ (Budapest)

Eszterházy Károly Egyetem (Eger)

kalcsogyula@dbk.pim.hu

The Petőfi Literary Museum's Digital Humanities Centre, established last year, is in the process of creating a digital humanities platform, dHUpla. One of its sub-areas is the planning of the management of born digital materials. The development of a procedure for born digital materials is an urgent need: in addition to the existing collection items in the PIM, it is expected that in the future more and more content of this kind will be created, which will need to be managed professionally. The presentation will describe the research work carried out and the details of the workflow that is being outlined. It will present the workflow developed on the basis of the OAIS model, which covers the process from ingestion of such material to making it searchable and publishable. It covers not only the more manageable text file formats of born digital content, but also the more difficult tasks of emails or social media sites, and how to professionally extract and archive data from external media.

Keywords: born digital, OAIS, BagIt, dHUpla, digital humanities platform

1. A born digital objektumok gyűjteményi kezelése Magyarországon aránylag új terület. Részletesen kidolgozott born digital workflow-val egyelőre csak a Magyar Nemzeti Levéltár büszkélkedhet, az ő gyűjteményi profiljuk azonban egy kissé speciális, a levéltári iratok egy jó része kötelezően szolgáltatandó az iratképzők által, ezért 2016-tól kormányrendelet szabályozza a levéltári elektronikus iratbeszolgáltatás rendjét¹. A kulturális közgyűjtemények born digital anyaga más jellegű, már ezen a területen is van kialakult gyakorlat, de elméletileg és tudományosan kevésbé megalapozott. A nagyobb közgyűjteményi szereplők természetesen rendelkeznek a born digital anyagokra vonatkozó, többé-kevésbé szabályozott eljárásrenddel: léteznek elektronikus könyvtárak, intézményi repozitóriumok, ahol kezelnek born digital objektumokat. Ugyanakkor ezek kialakításának az elvei, a használt módszerek, szoftverek közel sem egységesek, ebből fakadóan sajnos sok esetben egymással kölcsönösen nem (teljesen) átjárhatóak, valamint a köztük lévő szemantikus kapcsolatok sem gyakorlatiak.

A Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központja 2020-ban alakult meg, alapító okirata szerint "a hazai közgyűjteményekben zajló digitalizáció, tartalom- és szolgáltatásfejlesztések támogatására, koordinálására és dinamizálására" jött létre².

- 1 Magyar Közlöny, „34/2016. (XI. 30.) EMMI rendelet az elektronikus formában tárolt iratok közlevéltári átvételének eljárásrendjéről és műszaki követelményeiről”, elérés 2021. május 15., <https://magyarkozlony.hu/dokumentumok/afc7a7bc8a8b8c9e22db85179ec5932bc5fa05e9/megtekintes>.
- 2 <https://pim.hu/hu/digitalis-bolcseszeti-kozpont/rolunk>, elérés 2021. május 15.



A PIM-DBK kettős indíttatásból szándékozik a born digital objektumkezeléssel intenzíven foglalkozni: szükséges egyfelől a PIM gyűjteményezési tevékenységét támogatnia, a már meglévő gyűjteményi elemek feldolgozását elvégeznie, másfelől egy elméletileg és tudományosan megalapozott kulturális közgyűjteményi jó gyakorlat kialakítását is célul tűzi ki.

Az SZMSZ-e szerint „a PIM gyűjtőköre kiterjed a magyar irodalom egészének tárgyi, képi (képzőművészeti, fotó, videó és film), valamint kéziratos és nyomtatott, digitálisan keletkezett, illetve könyv formában található emlékeire, ill. hangzó dokumentumaira, valamint a magyar színházművészet és színházi kultúra muzeális tárgyi, képi, írásos és egyéb (hang-, film-, digitális anyag, stb.) emlékeire és az ezzel összefüggő képzőművészeti alkotásokra; illetve a tagintézmények vonatkozásában a kulturális javak mindazon körére, amelyet a tagintézmények muzeális működési engedélye gyűjtőkörként meghatároz. [...]

Fő feladata a gyűjtőkörébe tartozó muzeológiai forrásanyag felkutatása, gyűjtése, raktári megőrzése, műtárgyak kölcsönzése, visszaszorolása, szakszerű nyilvántartása, kezelése, állagmegóvása és védelme, restaurálása; továbbá tudományos feldolgozása és rendezése, mindezek kiállításokon és más formákban történő bemutatása, a közművelődését segítő hasznosítása.”³

A born digital objektumokkal kapcsolatos összetett feladatrendszer tehát egyfelől a közgyűjteményi hatáskörből adódik, másfelől viszont felszínre hoz olyan megoldandó problémákat és elvégzendő teendőket is, amelyek korábban nem voltak a fókuszban. Ennek a szerteágazó tevékenységnek a támogatására szükséges a PIM-DBK-nak kidolgoznia egy eljárásrendet, amely megfelel a Közgyűjteményi Digitalizálási Stratégiában⁴ megfogalmazott követelményeknek, összhangban van a PIM elvégzendő feladataival, és amelyet más közgyűjteményi szereplők számára is ajánlhat (szolgáltatásként is) abban a reményben, hogy létrejön egy egységesebb, átgondoltabb, szabványosabb, nemzetközileg is elfogadott normáknak megfelelő born digital-kezelési gyakorlat. A kidolgozandó eljárásrend a DBK által fejlesztett dHUpla⁵ nevű digitális bölcsészeti platform workflow-jának része lesz.

2. A born digital kifejezés digitálisan született, eleve digitálisan létrejött objektumokat⁶ jelöl, azaz tulajdonképpen minden olyan fájlt vagy több fájlból álló egységet, amely nem valamely korábban már analóg módon létező objektum digitalizálásával keletkezik. Ugyanakkor a born digital nem jelenti azt, hogy ezek az objektumok kizárólag digitálisan létezhetnek (készülhet analóg másolat róluk), viszont a kezelésükkel kapcsolatos problémák a digitális létmódjukból adódnak. A born digital objektum legalább egy, de sokszor több fájlból áll, amelyek lehetnek könyvtárakba szervezve. Fontos és problémás

3 https://pim.hu/sites/default/files/book/file/pim_szmsz_2020.10.14.pdf, elérés: 2021. május 15.

4 <https://digitalisjoletprogram.hu/hu/tartalom/kds-kozgyujtemenyi-digitalizalasi-strategia>, elérés: 2021. május 15.

5 A Digital Humanities Platform kifejezésből alkotott betűszó.

6 A digitális objektum kifejezés a Fehér könyv ajánlása a PREMIS alapján. Bánki Zsolt és Kómár Éva, szerk., *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez* (Emberi Erőforrások Minisztériuma, 2019), 123. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf. Elérés: 2021. május 15.

esetet jelentenek azok a “zárványok”, amelyek elavult adathordozókon (esetleg elavult formátumban) találhatóak, ezek megmentése elemi közgyűjteményi érdek.

A born digital objektumok kezelésének egyik fő nehézsége a heterogenitásukban rejlik. Ez jelenti egyrészt azt, hogy nagyon sokféle digitális objektumtípus lehetséges, másrészt egy-egy digitális objektum is állhat informatikailag eltérő típusú összetevőkből. Ugyanakkor kezelésük nehézségei közé tartoznak a digitális archiválás általános problémái, amelyek nem csak a born digital objektumokra vonatkoznak: a nagy mennyiség kezelése, a változatosságból adódó problémák megoldása, az elavulás kiküszöbölése, az értelmezhetőség fenntartása.

2.1. A born digital tárgykörébe tartozó objektumok meghatározása nemzetközi szinten sem egységes. A Közgyűjteményi Digitalizálási Stratégia⁷ alapidokumentumának számító *Fehér könyv*⁸ szerint “a digitális objektumok lehetnek szövegek, álló- és mozgóképek, hanganyagok, adatbázisok, adathalmazok, 3D objektumok és a felsoroltak konténerai”.⁹ Ez alapvetően tartalmi kategorizáció, a közgyűjteményi válogatás szempontjainak meg is felelhet, de a digitális megőrzésben általában inkább a fájlformátum alapján szokás osztályozni a digitális objektumokat. Ugyanakkor a fájlformátumok szerinti kategorizáció sem visz közelebb a born digital meghatározásához, hiszen bármilyen típusú fájl(ok együttese) lehet born digital tartalom. A born digital objektumok azért jelentenek speciális problémát, mert vannak olyan tulajdonságaik, amelyek eleve a digitális közeghez kötődnek, illetőleg analóg másolat előállításánál ezek a digitális közeghez kötődő sajátosságok torzulnak, megsemmisülnek, továbbá akár az is lehetséges, hogy nem is lehet analóg másolatot készíteni róluk. A legegyszerűbb born digital szövegfájl is több, mint a szöveges tartalmának nyomtatott változata: a fájlban tárolva vannak olyan metaadatok, amelyek az értelmezhetőség szempontjából fontos kontextuális információt tartalmazhatnak (pl. egy Word-dokumentum létrehozásának az időpontja).

Mindezek alapján a dHUpla rendszere eleve nem zár ki semmilyen formátumot, ugyanakkor kutathatóvá tenni, megjeleníteni, publikálni mindenekelőtt a szöveges tartalommal bíró digitális objektumok meghatározott körét tudja: szövegfájlokat, prezentációkat, e-maileket, egyéb szöveges üzeneteket (amennyiben szabványos formátumban állnak rendelkezésre), webes tartalmakat (blog-, weboldal-, közösségimédia- és egyéb exportokat, amennyiben szabványos formátumban állnak rendelkezésre). Ez utóbbiakkal kapcsolatban fontos újdonságot jelent, hogy a PIM-DBK rendszere alkalmas lesz a PIM gyűjtőkörébe vágó élő webes források szabályozott archiválására: kortárs írók, művészek által előállított, élő born digital tartalom befogadására és kezelésére, valamint publikálására és kutathatóvá tételére.

7 <https://digitalisjoletprogram.hu/hu/tartalom/kds-kozgyujtemenyi-digitalizalasi-strategia>, elérés: 2021. május 15.

8 Bánki Zsolt és Kómár Éva, szerk., *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez* (Emberi Erőforrások Minisztériuma, 2019), 123. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf. Elérés: 2021. május 15.

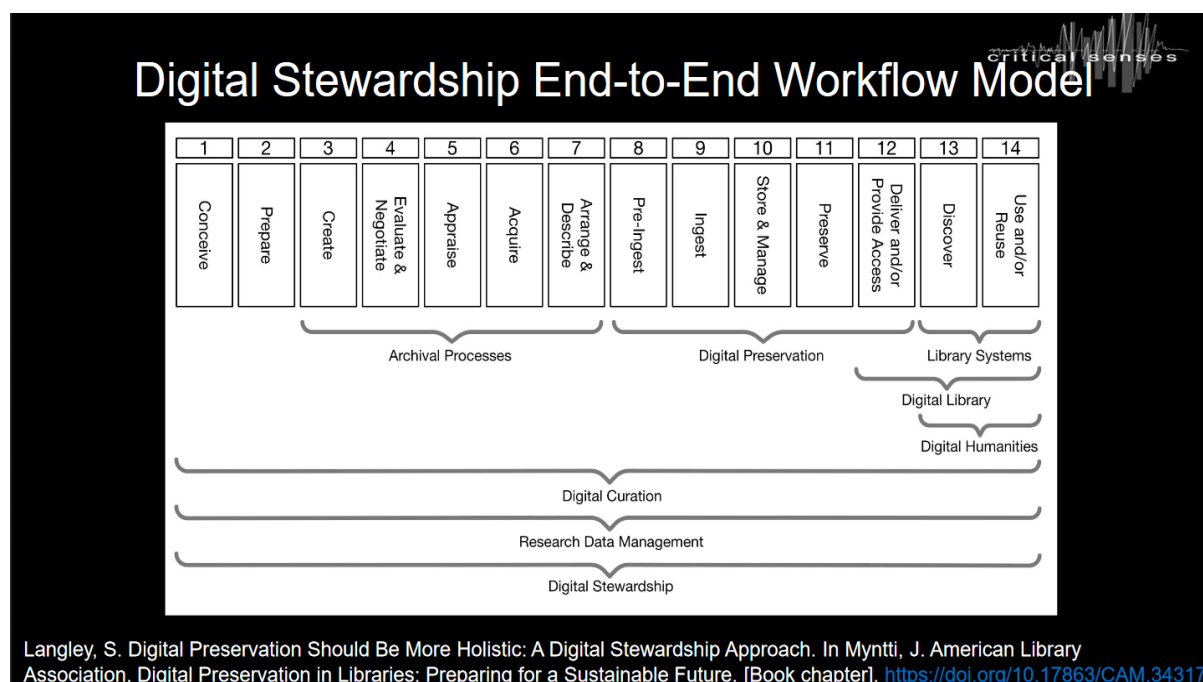
9 Uo., 123.



3. A PIM-DBK born digital workflow-jának kidolgozása előtt el kellett végeznünk a kutatómunkát: fel kellett tárnunk a vonatkozó szakirodalmat, valamint a hazai és nemzetközi jó gyakorlatokat. Bár a túlnyomórészt angol nyelvű szakirodalmi listánk hamar tekintélyesre duzzadt, összességében mégis inkább általában a digitális objektumok (és nem annyira speciálisan a born digital) kezelésére (digital curation v. digital stewardship) vonatkozó tételek dominálnak.¹⁰ A magyar nyelvű szakirodalomból kiemelendő a *Fehér könyv*, amely első változatában ugyan nem tartalmaz kifejezetten a born digital objektumok kezelésére vonatkozó részeket, a digitális megőrzésre vonatkozó általánosabb alapelvei viszont természetesen mérvadóak és követendőek. Számunkra alapvetően fontos volt a VII. fejezet (A digitális tartalmak menedzsmentje), amely többek között fontos terminológiai javaslatokat is tartalmaz. Ennek alapján használjuk pl. a digitális objektum terminus technicust.¹¹

3.1. Ugyancsak a *Fehér könyv*, valamint a nemzetközi szakirodalom, továbbá a tanulmányozott jó gyakorlatok alapján határoztuk meg az elméleti modelleket, amelyeket a born digital objektumok kezelésében is eredményesen lehet alkalmazni.

3.1.1. Bár a *Fehér könyv* a brit Digital Curation Centre Curation Lifecycle Modeljét ajánlja,¹² saját használatra egy bővebb modellt, Somaya Langley Digital Stewardship End-to-End Workflow Modeljét vettük alapul.¹³



Langley, S. Digital Preservation Should Be More Holistic: A Digital Stewardship Approach. In Myntti, J. American Library Association, Digital Preservation in Libraries: Preparing for a Sustainable Future. [Book chapter]. <https://doi.org/10.17863/CAM.34317>

1. ábra. Somaya Langley Digital Stewardship End-to-End Workflow Modelje

10 <https://www.zotero.org/groups/2532329/pim-dbk-dh-born-digital/library>. Elérés: 2021. május 15.

11 Bánki és Kómár, *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez*, 123.

12 Uo., 124. L. továbbá <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>. Elérés: 2021. május 15.

13 Langley Somaya, „Digital Preservation Should Be More Holistic: A Digital Stewardship Approach” (Apollo - University of Cambridge Repository, 2018), <https://doi.org/10.17863/CAM.34317>.

Ez a modell a teljes digitális objektumkezelés folyamatát 14 fázisra bontja.¹⁴

Langley cikkében kiemeli, hogy a digitális objektumok kezelésében alkalmazott modellek hiányosak, nem terjednek ki a teljes életciklusra (amelynek fontos része pl. a digitális tartalom létrejöttének a megtervezése, vagy éppen a felhasználás, hasznosítás szempontjainak a figyelembevétele). Pedig a körültekintő tervezés segíthet a digitális megőrzés problémáinak (mennyiség, változatosság, elavulás, értelmezhetőség; l. fentebb) enyhítésében.

3.1.2. A digitális tartalmak menedzsmentjének, a megőrzés általános feladatainak a modellezésére világszerte a CCSDS által javasolt **Open Archival Information System (OAIS)** modell, valamint a hozzá kapcsolódó szabványcsalád a legelterjedtebb.¹⁵ Ez egy nemzetközi standard (ISO 14721:2012), amelyet a *Fehér könyv* is ajánl, valamint Magyarországon is van már példa az implementálására: a Magyar Nemzeti Levéltár born digital objektumkezelési gyakorlata is ezen alapul.¹⁶ A modell három fő ágenszt különböztet meg: a tartalom előállítója, kezelője és felhasználója. Az ágensek közötti adatcserére vonatkozó ajánlásokat fogalmaz meg: a digitális objektumokat ún. csomagokban javasolja kezelni, amelyekben kötelezően előírt elemek vannak. A SIP-csomag (Submission Information Package) tartalmazza az előállító által létrehozott tartalmat, valamint a létrehozás és a kezelőnek átadás körülményeit dokumentáló metaadatfájlokat. Ebből a kezelő AIP-csomagot (Archival Information Package) hoz létre, amely a SIP-csomagon kívül tartalmazza a megőrzési terv alapján előírt további kötelező elemeket: pl. a megőrzési formátumra konvertált reprezentációkat, azok metaadatait, mindenféle, a megőrzést megelőző művelet (pl. vírusellenőrzés, hashing stb.) dokumentációját, metaadatait. Az AIP-csomag célja kettős: egyrészt az archiválás, másrészt ebből jön létre a szolgáltatási csomag, a DIP (Dissemination Information

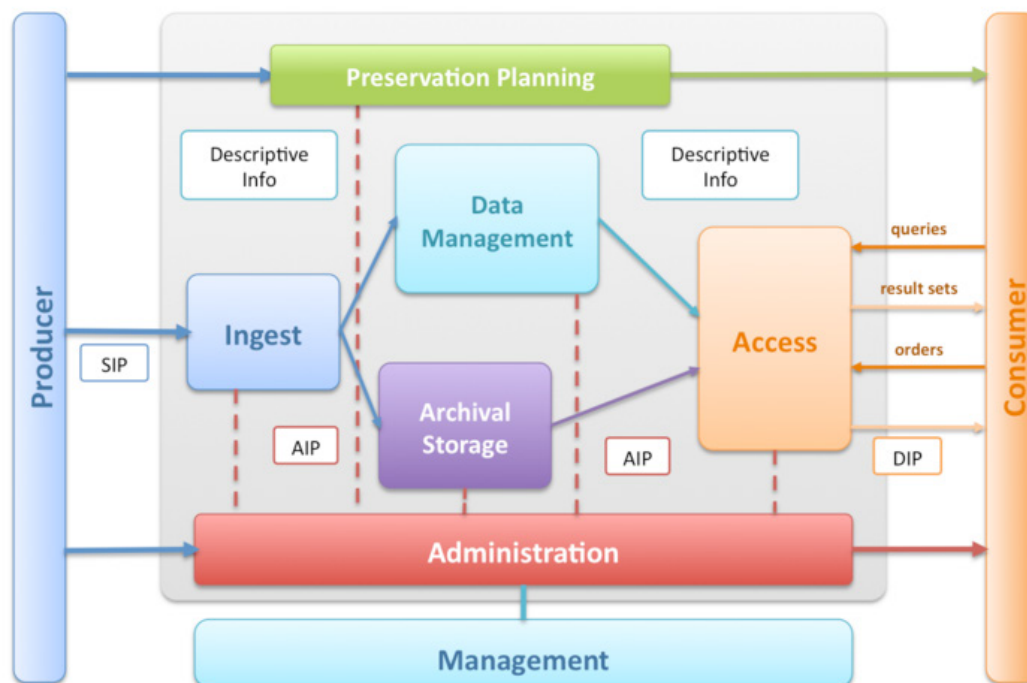
14 Nem a magyar terminológia egységesítésének, ill. bizonyos esetekben kialakításának a szándékával, hanem csupán a jelen cikk szóhasználatának bevezetése céljából közöljük ezen fázisok általunk használt magyar megfelelőjét:

- 1 Conceive: tervezés
- 2 Prepare: előkészítés
- 3 Create: létrehozás, előállítás
- 4 Evaluate & Negotiate: kiértékelés és megvitatás
- 5 Appraise: értékelés, besorolás
- 6 Acquire: beszerzés
- 7 Arrange & Describe: elrendezés és leírás
- 8 Pre-Ingest: a gyarapítás (átvétel) előkészítése
- 9 Ingest: gyarapítás (átvétel)
- 10 Store & Manage: tárolás és kezelés
- 11 Preserve: megőrzés
- 12 Deliver and/or Provide Access: szolgáltatás és/vagy hozzáférés biztosítása
- 13 Discover: feltárás
- 14 Use and/or Reuse: használat és/vagy újrahasznosítás.

15 CCSDS, *Reference Model for an Open Archival Information System (OAIS)*, *Magenta Book* (CCSDS - Consultative Committee for Space Data Systems, 2002), <http://public.ccsds.org/publications/archive/650x0b1.pdf>, Elérés: 2021. május 15. További szabványok: PAIMAS (<https://public.ccsds.org/Pubs/651x0m1.pdf>, elérés: 2021. május 15.); PAIS (<https://public.ccsds.org/Pubs/651x2g1.pdf>, elérés: 2021. május 15.)

16 <https://mnl.gov.hu/mnl/szkk/szabvanyok>, elérés: 2021. május 15.

Package), amelyet a felhasználók rendelkezésére lehet bocsátani. Ez a csomag az AIP bármely reprezentációját, valamint szükség esetén további konvertált reprezentációkat, azok dokumentációját, metaadatait tartalmazza.



2. ábra. Az Open Archival Information System (OAIS) alapfogalmai

A „csomagolás” az OAIS-modellben korántsem öncélú, számos haszna mellett két fontos mozzanat emelendő ki: a digitális objektumokra vonatkozó dokumentáció, ill. metaadataik együtt kezelhetők, mozgathatók magával az objektummal, továbbá azon digitális objektumok esetében, amelyek több különálló fájlból vagy akár könyvtárak struktúrájából állnak, lehetséges az objektum részeinek egyben tartása, valamint az eredeti felépítés megőrzése.

3.1.3. A csomagok tartalmára nézve is létrejöttek ajánlások, amelyek beépültek az OAIS-modellt alkalmazó szoftverekbe is, megkönnyítve a digitális kurátorok dolgát a csomagkészítés terén. Az egyik ilyen ajánlás a Library of Congress által kidolgozott BagIt, amely az Internet Engineering Task Force 8493-as ún. RFC-je (Request for Comments).¹⁷ A BagIt előírása alapján szükséges egy tetszőleges elnevezésű gyökérmappa (base directory). Ebben kötelezően 3 elem van: egy data elnevezésű mappa, ebben van a tartalom (payload). A tartalom tetszőlegesen összetett lehet (maga is egy vagy több mappa további almappákkal és fájlokkal stb.). Szükséges legalább egy manifest-<algorithm>.txt fájl, ahol az <algorithm> a fájlnevében helyettesítendő a használt checksumkészítési algoritmus nevével (pl. MD5). Ebben a fájlban benne kell

¹⁷ Kunze John A., Littman Justin, Madden Liz, Scancellia John és Adams Chris, *The BagIt File Packaging Format (V1.0)*, Request for Comments 8493 (RFC Editor, 2018), <https://doi.org/10.17487/RFC8493>. Request for Comments 8493 (RFC Editor, 2018 Elérés: 2021. május 15. Itt említendő még az elsősorban levéltári használatra kifejlesztett E-Ark sepcifikáció, amely a BagIt-nél jóval összetettebb csomagokat ír elő. <https://www.eark-project.com/>. Elérés: 2021. május 15.

lennie a csomag összes fájljára vonatkozó checksumértéknek. Szükséges végül egy bagit.txt fájl, amelyben két információt kell megadni: a Bagit csomagolási eljárás verzióját (jelenleg 1.0), és a TXT-fájlok karakterkódolását (pl. UTF-8, ez utóbbi az ajánlott). A csomag további címkefájlokat is tartalmazhat opcionálisan, amelyek elnevezésére és tartalmára, valamint a formátumára és kódolására a Bagit tesz javaslatokat. Pl. bag-info.txt ilyen adatokkal: Source-Organization, Contact-Name, Contact-Email stb. Lehetséges még címkefájlokat (metaadat-fájlokat) elhelyezni külön mappába is (tetszőleges mappanévvvel). Minden opcionális fájl checksumjának szerepelnie kell a manifestfájl(ok)ban. A Bagit csomagolóeljárást számos szoftver támogatja (Bitcurator, RODA).

4. A szakirodalom és a nemzetközi jó gyakorlatok alapján a fentebb kifejtett alapelveknek megfelelő workflow kidolgozásába kezdtünk a PIM-DBK-ban. Kiindulópontunk az volt, hogy a born digital objektumok kétféle csatornán kerülhetnek a rendszerünkbe: adathordozón, valamint hálózatosan (e-mailben elküldve, felhőtárhelyes hozzáférés megadásával stb.). Biztosítani kívánjuk továbbá, hogy a csatlakozó partnerek (akár magányszemélyek, akár más közgyűjtemények vagy egyéb szervezetek) képesek legyenek az általunk fejlesztett rendszer segítségével a saját born digital tartalmaik kezelésére. Rendszerünket tehát úgy kellett megtervezni, hogy mind a PIM belső, már meglévő gyűjteményi elemeit (akár adathordozón, akár szerveren tárolt anyagról van szó), mind a kívülről (akár adathordozón, akár hálózaton) érkező anyagokat fel tudjuk dolgozni.

4.1. Az adathordozón tárolt anyagok kezelése speciális hardveres és szoftveres infrastruktúrát igényel. Ennek oka az, hogy az adathordozón található ideiglenes vagy akár törölt fájlok, magának az adathordozónak a fájlstruktúrája stb. fontos, megőrzendő kontextuális információt jelentenek, továbbá szükséges biztosítani azt is, hogy a tartalom másolása szakszerűen, az fájlok integritásának megőrzésével történjen. Ebből következően az adathordozók gyűjteményi archiválásához szükség van megfelelő számítógépre, amely fel van szerelve az ún. digitális törvényszéki módszerek alkalmazásához elengedhetetlen hardveres és szoftveres kiegészítőkkal (írásblokkoló, lemezképkészítő, fájlvalidáló, vírus- és malware-ellenőrző stb.). A kialakult nemzetközi standardoknak megfelelően a PIM-DBK is az USA-beli Digital Intelligence termékét, a FRED-et (Forensic Recovery of Evidence Device),¹⁸ vagy azzal ekvivalens nagy teljesítményű PC-t kíván szolgálatba állítani, amelyre telepíteni kell a Bitcurator nevű Linux-disztribúciót.¹⁹ A FRED és a Bitcurator segítségével minden olyan feladat megoldható, amely a born digital tartalmak gyűjteményezéséhez elengedhetetlen. Az adathordozó tartalmáról biztonságosan bitszintű másolat készülhet, a tartalom szűrhető (eltávolíthatók az archiválni nem kívánt vagy az archiválásból jogilag kizárt fájlok, akár a donor közreműködésével is) stb., továbbá minden műveletről részletes és szabványos dokumentáció készülhet.

4.2. A nem adathordozón tárolt tartalom esetében is szükséges a Bitcurator használata, hiszen a born digital objektumok gyakran tartalmaznak kiszűrendő érzékeny információkat (pl. védett személyes adatokat, a GDPR által a felhasználáshoz hozzájárulást előíró

¹⁸ <https://digitalintelligence.com/products/fred/>, elérés: 2021. május 15.

¹⁹ <https://bitcurator.net/>, elérés: 2021. május 15.



adatokat). A Bitcurator megfelelő konfigurációval jelentősen megkönnyíti ezek kezelését, továbbá tartalmaz egy Bagger nevű eszközt is, amellyel (akár adathordozóról származik a tartalom, akár online forrásból) elvégezhető a Baglt specifikációnak megfelelő csomagolás, azaz előállhat a gyűjteményezési folyamat bemeneteként szolgáló SIP-csomag. Ez a csomag kerül be a PIM-DBK által fejlesztett digitális bölcsészeti platform (dHUpla) rendszerébe.

4.3. A SIP-csomag tartalma igen változatos lehet, a data mappa (l. fentebb) egyetlen fájl, de akár sokszintű könyvtárstruktúrát is tartalmazhat. A leglényegesebb alapelv, hogy a SIP-csomagnak tartalmaznia kell az eredeti (beadási) formátumot, továbbá minden dokumentációt, amely a bekerüléssel kapcsolatos (a donorral kötött szerződéseket, a Bitcurator által előállított jelentéseket stb.).

5. A born digital tartalmak a dHUpla rendszerében. A PIM-DBK digitális bölcsészeti platformja a born digital tartalmak kezelésére egy OAIS-kompatibilis teljes körű repozitóriumszolgáltatást nyújtó nyílt forráskódú szoftvert, a portugál Keep Solutions RODA nevű termékét²⁰ szándékozik használni. Ez a szoftver képes befogadni az előállt SIP-csomagokat, amelyekből tetszőleges tartalmú AIP-csomagokat állít elő. A csomagokkal minden, a repozitóriumokban megszokott művelet elvégezhető (akár API-n keresztül, automatizáltan és tömegesen is): fájlok átcsoportosítása, könyvtárstruktúra kialakítása, megváltoztatása, a fájlok törlése, betekintés stb. Mivel számos nyílt forráskódú eszközt magába integrál, a RODA képes az AIP-csomagok kialakításához műveleteket elvégezni: fájlkonverziók, (újra)hashelés, validálás stb. Hosszútávú megőrzésre szabványos formátumra konvertáljuk a bemeneti formátumokat, az AIP-csomagok az eredeti mellett ezeket is tartalmazzák. A dHUpla rendszere ezeknek a szabványos formátumoknak a megjeleníthetőségét, kutathatóságát is biztosítja.

A RODA kimenetként ugyancsak tetszőlegesen összeállított DIP-csomag letöltését is lehetővé teszi. Azon born digital tartalmak esetében, amelyek tudományos szempontból kitüntetett figyelmet érdemelnek (írói kéziratok, feljegyzések, levelezések stb.), a dHUpla forráskiadási rendszere²¹ tudományos kiadást állít elő, amely külön felületen megtekinthető és kutatható, valamint nem utolsósorban olyan szemantikus annotációt kap, amely lehetővé teszi különféle szemantikus kapcsolatok létrehozását (pl. jelentős nemzetközi adatbázisokkal vagy éppen a Nemzeti Névtérrel).

²⁰ <https://www.keep.pt/en/producs/roda-long-term-digital-preservation-repository-solution/>, elérés: 2021. május 15.

²¹ A forráskiadás rendszere külön részrendszer a dHUplán belül, amelynek működése nem képezi jelen előadás tárgyát.

Bibliográfia

- Bánki, Zsolt és Kómár, Éva, szerk. *Fehér könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez*. Emberi Erőforrások Minisztériuma, 2019. https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf.
- CCSDS. *Reference Model for an Open Archival Information System (OAIS)*, Magenta Book. CCSDS - Consultative Committee for Space Data Systems, 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- Magyar Közlöny. „34/2016. (XI. 30.) EMMI rendelet az elektronikus formában tárolt iratok közlevéltári átvételének eljárásrendjéről és műszaki követelményeiről”. Elérés 2021. június 15. <https://magyarkozlony.hu/dokumentumok/afc7a7bc8a8b8c9e22db85179ec5932bc5fa05e9/megtekintes>.
- Kunze, John A., Littman, Justin, Madden, Liz, Scancella, John és Adams, Chris. *The BagIt File Packaging Format (V1.0)*. Request for Comments 8493. RFC Editor, 2018. <https://doi.org/10.17487/RFC8493>.
- Langley, Somaya. „Digital Preservation Should Be More Holistic: A Digital Stewardship Approach”. Apollo - University of Cambridge Repository, 2018. <https://doi.org/10.17863/CAM.34317>.



TANULMÁNYOK A MŰSZAKI TERÜLETEKRŐL

OSINT technológiák és alkalmazási lehetőségeik a felsőoktatási rendszerek ellen

Koczka Ferenc
Eszterházy Károly Egyetem
ORCID: [0000-0002-7541-6495](https://orcid.org/0000-0002-7541-6495)

Absztrakt

A tudomány és eredményeinek szabadsága, az egyetemek szabad szellemisége, a nyílt működés és az alkotói szabadság évszázadok óta fontos jellemzője az akadémiai szférának. A tudományos eredmények publikálása számos, akár nemzetközi szintű szakmai csoportok kialakulását segítette, s hozzájárult a tudományos élet színvonalának emelkedéséhez. Az Internet szolgáltatásaiban rejlő lehetőségeket a tudományos világ az elsők közt használta fel, új csatornákon tette közzé kutatásainak folyamatát és eredményeit, ami korábban nem látott mértékben segítette az azonos érdeklődésű kutatók kapcsolatépítési lehetőségeit. A közzétett információk nem korlátozódnak a tudományos környezetre, a tudományos eredmények gyakorlati felhasználása értéket jelent a gazdasági és a bűnözői körök számára is. E szereplők elsődleges célja azonban más, szinte kizárólag a gazdasági előny megszerzésére korlátozódik. Cikkemben rámutatok, hogy mások mellett az akadémiai szféra intézményei is túlzott mennyiségű információt tesznek közzé magukról, azok védelmét magasabb szinten kellene ellátniuk, és javaslatot teszek arra, hogy nagyobb hangsúlyt fektessenek a nyílt forrású felderítés elleni védekezésre.

Abstract

The freedom of science and its results, the free spirit of universities, openness and creative freedom have been an important feature of the academic world for centuries. The free access to university courses and the publication of scientific results have helped the development of many professional groups, even at the international level, and have contributed to raising the quality of academic life. The scientific world was among the first to take advantage of the Internet's services, using new channels to publish their research and results, which greatly facilitated networking opportunities for researchers with similar interests. The information published is not limited to the scientific environment, the practical use of scientific results is of value to economic and criminal circles. Although the primary objective of these roles is different, in most cases, they seek to gain an economic advantage. In my article, I point out that academic institutions also publish too much information about themselves, that they should have a higher level of protection, and I suggest that they should put more emphasis on protecting themselves against open-source intelligence.



Bevezetés

A nyílt forrású felderítést (OSINT – Open Source Intelligence) az amerikai kormányok már a múlt században is alkalmazták, elsősorban a külföldi média által közölt információk begyűjtésére és rendszerezésére. Pearl Harbor megtámadása után a módszert rendszerszintűvé tették, 1941-ben felállt az FBIS (Foreign Broadcast Information Service) mely néhány év múlva már a CIA részeként működött az USA hírszerzési rendszerében [1]. Az OSINT azóta is fontos eleme a hírszerzésnek, definíciója olyan publikus, vagy korlátozott körben elérhető adatok legális eszközökkel történő megkeresését, összegyűjtését, szintetizálását és felhasználását határozza meg, melynek eredményeképp új információ hozható létre. A legalitás lényeges elem a meghatározásban, ez kizárja a minősített információk megszerzésére irányuló célzott támadásokat, és más olyan eseteket, melyek során az információ megszerzése valamilyen forrás legális felhasználásán túl jött létre. Bár eredetileg kormányzati, hírszerzési, katonai, bűnüldözési és belbiztonsági szervezetek alkalmazták, de az OSINT már jelen van az üzleti szférában is¹.

Mivel célzott támadások előkészítésére az egyik legegyszerűbb módszer az OSINT, a metodikát kiberbűnözők is alkalmazni kezdték. Alkalmazásával elkerülhetők az olyan kockázatos műveletek, mint pl. az illegális belépési kísérletek, és nem szükségeses social engineering módszerekkel személyes kapcsolatokat építeni, megtéveszteni vagy megvesztegetni kulcspozícióban levő alkalmazottakat sem. Annak ellenére, hogy az OSINT metodikák egy része kifejezetten a manuális adatgyűjtésre épül, egy részük technikai szempontból könnyen automatizálható és naprakészen tartható. A mesterséges intelligencia fejlődésével pedig úgy tűnik, az automatizálási lehetőségek tovább erősödnek majd. Nem tétlenek a bűnözői körök sem, a technikai újításokat és tudományos eredményeket felhasználva olyan új típusú bűncselekményeket valósíthatnak meg, melyek ma még szinte kivédhetetlennek bizonyulnak. Miközben vállalatok a humán telefonos ügyfélszolgálatuk mesterséges intelligencián alapuló kiváltásával kísérleteznek, utóbbiak ugyanezzel a technológiával már sikeresen tévesztettek meg munkatársakat és vették rá őket fiktív kifizetések indítására [2].

Az adatszerzés forrásai

Az OSINT technológia alkalmazásának kiszélesedését az adatforrások körének kiszélesedése, jogszabályi követelmények és társadalmi elvárások változásai is segítik.

Az üzleti szférában a szolgáltatók a meglévő és a potenciális klienseik tájékoztatására minél több információt igyekeznek közzétenni, melyek begyűjtését az ellenérdekelt oldal a megfigyeléstől az internetes kereséseken át, a különféle adatbázisokra történő előfizetéssel, szabadon elérhető információhordozókkal, publikációk begyűjtésével, könyvek, tanulmányok, napilapok, televízió- és rádióadások, podcastok, blogok, vblogok, konferenciák, beszámolók elemzésével végzi [3].

1 Egy norvég OSINT adatokat szolgáltató cég pl. nyílt forrású adatgyűjtéssel teszi lehetővé biztosítók számára, hogy olyan kelet-európai országok mezőgazdasági termelők számára is biztosítást nyújthassanak, amelyekben nem áll rendelkezésre hivatalos adat a kockázatok kiszámításához. Forrás: <https://www.osintanalytics.com>.

A közösségi média és alkalmazások népszerűségének felértékelődése különösen értékes információforrást biztosít a nyilvános forrású adatgyűjtéshez, mivel az eredeti adatközlésen túlmutató, másodlagos adatforrást is jelentenek. Előfordul, hogy munkahelyi adatok publikálását végző munkatársak nincsenek tisztában azzal, hogy olyan információkat is közzétesznek, amely eredetileg nem állt szándékukban². A Facebook megosztások kockázatának ismerete általánosan mondható, de gyakorlatozó amerikai katonák számára már nem volt nyilvánvaló, hogy a sporttevékenységeik megosztásával a katonai egységeik helyzetére és mozgására engednek következtetni³.

Nagy tömegű anonim GPS adat összegyűjtésével kapcsolatos nemzetbiztonsági kockázatra a New York Times vizsgálatai hívták fel a figyelmet. Az újságíró számára egy adatbázist juttattak el, mely mobiltelefonok egyedi azonosítóit és azok mozgását leíró GPS adatokat tartalmazott. Bár abban személyes adat nem állt rendelkezésre, a koordináták helyét és idejét más, publikus forrásokkal összevetve az eszközök tulajdonosai könnyűszerrel azonosíthatóvá váltak. A vizsgálatban védett személyek mellett sikeresen azonosították az amerikai elnök telefonját is és bizonyították mozgásának további követhetőségét [4] [5].

Bár az egyes rendszerek és komponenseik sérülékenységeit közlő források⁴ többsége alkalmazható a nyílt forrású felderítésben, a sérülékenység kiaknázásához szükséges eszköz kifejlesztéséhez etikai okokból nem nyújtanak elégséges információt. Ezt a hiányt a Darkneten elérhető források fedik le.

Jogsabályok és eszközök

Az Európai Parlament és a Tanács 2016/679 rendelete (közismert nevén GDPR) az EU állampolgárai számára nyújt védelmet a személyes adataik kezelése terén. A jogszabály számos ponton részesíti előnyben a magánszemélyek személyes adatainak védelmét az üzleti szférával szemben, ugyanakkor a jogszabályban meghatározott adatkezelési folyamatokat természetesen nem akadályozza. Bár a jogszabály bonyolult és a gyakorlati alkalmazása is nehézkes, súlyos szankciókat határoz meg az adatkezelés tömeges megsértése esetén. A GDPR kizárólag magánszemélyek adatainak kezelését szabályozza, hatálya nem terjed ki cégek, szervezetek védelmére [6].

Az OSINT alkalmazási lehetőségeinek korlátozásában a GDPR-nak fontos szerepe volt az EU-ban. Az EU-n kívüli államokban a magánszemélyek adatainak védelme korántsem olyan egységesen szigorú, a személyes adatok sok esetben árucikkek, melyekkel a hatályos jogszabályok betartásával kereskedni is lehet.

2 A weblapokon, közösségi hálózatokon és más forrásokban közzétett képek metaadatainak elemzése bizonyos esetekben lehetővé teheti azok készítési helyének és idejének megállapítását.

3 <https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases>

4 A Common Vulnerabilities and Exposures adatbázisa a <https://cve.mitre.org> címen érhető el, és nyújt információt az ismertté vált sérülékenységekről, de nem nyújt részletes információkat ahhoz, hogy a sérülékenységek kihasználására alkalmas eszköz kifejleszhető legyen.



Az Egyesült Államokban számos szolgáltatás kínál nyílt forrású adatok alapján készült jelentéseket magánszemélyekről is, melyek összehasonlítása során kimutatható, hogy az alapszolgáltatásaik jórészt azonos forráson alapulnak. Egy ilyen jelentés részleteinek megismerésére, és a közzétett adatok érvényességének ellenőrzésére egy interjút készítettem egy amerikai állampolgárral.

Az ő előzetes engedélyével kb. 20 dolláros áron vásároltam meg egy jelentést róla, melyhez kiinduló információként pusztán a nevét és egy lehetséges lakhelyét kellett megadnom. A szolgáltatás számára ennyi már elég volt ahhoz, hogy azonosítsa és egy 24 oldalas jelentést generáljon róla. A dokumentum a leíró adatok mellett részletezte a munkahelyeit, felsorolta rokoni és egyéb lehetséges kapcsolatait, megnevezte a vele korábban közös lakcímen lakó személyeket és annak időtartamát, a SSN (Social Security Number) forrását, korábbi telefonszámait, e-mail címeit, lakcímeit és szomszédjait, korábban birtokolt gépjárműveit. A hatósági ügyekkel kapcsolatos információk közt bírósági meghallgatások, szabálysértések szerepeltek. Önálló részt kapott a környezetében nyilvántartott szexuális bűncselekmény elkövetésével kapcsolatba hozott személyek fényképes listája. Az érdeklődési köréről és attitűdjéről nem csak a legismertebb közösségi hálózatok konkrét azonosítói adtak tájékoztatást, ebből vallási hovatartozása is egyértelművé vált.

Az interjú alanya ellenőrizte a dokumentum tartalmát, egy kontakt személy és néhány, a hatósági nyilvántartásban szereplő adat kivételével megerősítette azok helyességét.

A nyílt forrású felderítés támogatására első látásra bőséges eszköztár érhető el az Interneten⁵. Ezek a szoftverek jól dokumentáltak, bár megítélésem szerint számos olyan van köztük, melyek elsődleges célja nem az OSINT támogatása, legfeljebb bizonyos körülmények közt adhatnak használható információt. A valóban jól használható, hatékony és mély műszaki ismereteket nem igénylő megoldások ingyenesen használható verziói rendszerint csak erősen korlátozott funkcionalitással érhetők el, és jellemzően adott időszakokra szóló előfizetéssel használhatók.

Az OSINT és az akadémiai szféra

Az OSINT a mindennapi életben hatékonyan segítheti a különféle szervezetek munkáját, pl. a munkáltatók korlátozásokkal ugyan, de megtekinthetik és döntésükhöz felhasználhatják az adott munkakör betöltésére jelentkezők nyilvános profilját [7]. Ugyanakkor nem csak az intézményt érő célzott támadások elkerülésének érdekében kellene minimalizálniuk a publikusan elérhető adatok körét.

2021-ben több magyar egyetem esett áldozatul a bankszámlaszám változását elérő csalásnak, mely során a csalók egyetemi ügyintézők számára telefonon is megerősítették a változás tényét. Az eset bizonyítja, hogy az akadémiai szféra Magyarországon is lehet célpontja olyan jellegű támadásoknak, melynek megtervezésében az OSINT is szerepet kaphatott, és bár az nem kapott széles körű publicitást, az érintett intézmények megítélését is negatívan befolyásolhatja.

⁵ Az OSINT eszközök egyfajta rendszertana a <https://osintframework.com> oldalon érhető el.

Anyilvános telefonkönyvek, dolgozói katalógusok, e-mail címjegyzékek és organogramok segítenek feltérképezni a kulcspozícióban levő döntéshozókat. Az egyetemek működésére vonatkozó adatok közérdekből nyilvánosak, könnyen megtalálhatók a szerződött partnerek és szerződés részletei⁶. Ezek birtokában sokkal könnyebb rést találni a szervezet működésében, főleg, ha a dolgozók nem feltétlenül ismerik egymást. A vezetők által aláírt és közzétett dokumentumok nagyban leegyszerűsítik olyan hamisított iratok előállítását, amelyek a belső munkatársak számára is hitelesnek tűnhetnek⁷.

A nyilvánosság ilyen formája kifejezetten negatív hatású a munkaerő megőrzésének területén. Önéletrajzok, tudományos eredmények közzététele OSINT forrást jelent a fejtárgysz cégek számára, de az adathalász és megtévesztő támadások kiindulópontja lehet.

A munkatársainak kapcsolati adatainak egy részét a legtöbb felsőoktatási intézmény valamilyen formában elérhetővé teszi. Csak néhány esetben találkozhatunk olyan stratégiával, melyben nem személyek, hanem szerepkörök ismerhetők meg az e-mail címek alapján. Bár általában a telefonkönyvek fejlesztői igyekeznek megakadályozni a kapcsolati adatok tömeges leszüretelését, tapasztalataim szerint ez könnyen kijátszható. Erre vonatkozó országos mérés nehezen végezhető, mivel a munkahelyi e-mail cím is személyes adat, és tulajdonosának hozzájárulása nélkül nem kezelhető. Sajnos a felkeresett egyetemi informatikai vezetők nem támogatták egy ilyen vizsgálat elvégzését, így csak kerülő megoldással lehet megállapítani, hogy pl. dolgozói személyes adatok leszüretelhetők-e az adott intézmény esetében. Az alábbi anonimizált példa demonstrálja, hogy egy ilyen feladat néhány soros scripttel is megoldható úgy, hogy a forrás kilétének rejtését a Darknet infrastruktúrája adja.

```
#!/bin/bash
URL="https://www.uni-ANON.hu/telefon/index.php?menu_id=2&myaction=details&
key_field_value=XXX"
for I in $(seq 250 1 10000) ; do
  echo $URL | sed „s/XXX/$I/g”
  torify wget $(echo $URL | sed „s/XXX/$I/g” ) -q -O - | grep -EiEio
  ‚\b[A-Z0-9._%+~]+@[A-Z0-9.-]+\.[A-Z]{2,4}\b’
done
```

6 Az EKE jelentősebb szerződéseit és beszerzéseit a https://uni-eszterhazy.hu/public/uploads/5-millio-ft-feletti-erteku-szerzodesek-2019-5e382fcbd6e7e_5e4ec05f04e0b.pdf oldalon tette közzé. Az Államkincstár az egyetem élő szerződéseinek listáját a http://www.allamkincstar.gov.hu/files/A%20Kincstárról/Üvegzszeb/2021/ÜVEGZSEB_20210312.pdf oldalon publikálta.

7 Érdemes egy Google keresést indítani a “Szervezeti és Működési Szabályzat” filetype:pdf kifejezéssel, a találatok közt alig van olyan dokumentum, amelyben ne szerepnének eredeti aláírások.



Összegzés

A magyar felsőoktatás informatikai rendszereinek működését csak általános jogszabályok szabályozzák, bár az elmúlt időszakban néhány egyetem kutatási egysége nemzetbiztonsági védelem alá került [8]. A teljes felsőoktatás stratégiai besorolásának megváltoztatására már van nemzetközi példa [9], így várható, hogy idővel azt más országok is követik. Ezen a téren a magyar jogi helyzet nem koherens: miközben az esetenként több tízezres létszámú egyetemek szabadon határozzák meg az informatikai rendszereik működtetésének szabályait, egy kis vidéki önkormányzatnak sokkal szorosabb jogszabályi keretek közt kell működnie.

Műszaki szempontból tekintve a felsőoktatási informatikai rendszerek nem speciálisak, a fő eltérések a finanszírozásban, a nyílt működés támogatásában és a szektor speciális szoftvereiben jelennek meg. Több eset igazolja, hogy az OSINT eszközök a felsőoktatási informatikai rendszerekkel szemben is alkalmazhatók, és a publikus információs felületeken megjelenő tartalmak mennyiségét jelentős mértékben szűkíteni kellene. Különösen kerülendő a munkatársak személyes és kapcsolati adatainak teljes körű begyűjthetősége, a belső szabályzatok, és egyéb dokumentumok korlátozás nélküli hozzáféréseinek biztosítása. Ezek publikálására célszerű védett területet kialakítani, az intézményi publikus felületek szerkesztését pedig olyan centralizált szervezeti egységre bízni, melynek tagjai ismerik az OSINT célokat módszereket, így azok képesek a publikált tartalmakat a nyílt forrású felderítés számára a lehető legkevésbé ellene fordítható mértékben és formában megjeleníteni.

A jogalkotás szintjén célszerű lenne újragondolni azokat az előírásokat, melyek a kiberbűnözők információhoz jutását segítik elő. Bár az üvegseb programmal járó részletes adatszolgáltatási kötelezettségek és a pályázati források felhasználását bemutató weboldalak a közpénzek felhasználását, így a közintézmények átláthatóságát teremtik meg, egyúttal gazdag információforrást nyújthatnak előre megtervezett, célzott támadások kivitelezéséhez.

Jogi támogatás hiányában az felsőoktatás döntéshozói nem tudnak ezen a területen hatékony védelmi lépéseket tenni, mivel saját belső szabályzataik nem kerülhetnek ellentmondásba az érvényes jogszabályokkal. Az adatszolgáltatási kötelezettség gyakorlata felülírja az OSINT védelmi szabályokat, elsődleges szempontként a jogszabályok betartása és az esetleges szankciók elkerülése érvényesül. Általános intézményi jó gyakorlatok megítélésem szerint legfeljebb szigetszerűen léteznek, az egyetemek informatikai vezetésének jelenleg nincs olyan működő platformja, melyben tapasztalataikat megoszthatnák, illetve egy esetleges informatikai incidens esetén a társintézmények számára információt nyújthatnának.

Az intézményi informatikai biztonsági szabályzatok és az azokhoz kapcsolódó oktatások megtervezése során célszerű az nyílt forrású felderítés kérdéskörére is kitérni. Ennek eredményeként az OSINT által alkalmazott technikák ismeretében egy olyan területen javítható az intézmény információbiztonsága, melyet manapság az informatikai szakemberek szinte alig felügyelnek.

Irodalomjegyzék

- [1] F. Schaurer és J. Störger, „The Evolution of Open Source Intelligence (OSINT),” *Journal of U.S. Intelligence Studies*, 2013, 19. kötet, 3. szám, 53. o.
- [2] C. Supp, „Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case,” 30 08 2019. [Online]. Elérhető: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>. Hozzáférés dátuma: 2021.01.01.
- [3] V. Deák, „A nyílt forrású információszerezés szerepe a kibertámadások végrehajtása során,” *Hadmérnök*, 2018. XIII. évf. 3. szám, 393. o.
- [4] S. A. Thompson és S. Warzel, „welve Million Phones, One Dataset, Zero Privacy,” *The New York Times*, 19 12 2019. [Online]. Elérhető: <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>. Hozzáférés dátuma: 2021.01.12.
- [5] S. A. Thompson és C. Warzel, „How to Track President Trump,” *New York Times*, 20 12 2019. [Online]. Elérhető: <https://www.nytimes.com/interactive/2019/12/20/opinion/location-data-national-security.html>. Hozzáférés dátuma: 2021.01.15.
- [6] Európai Parlament és a Tanács, *Az Európai Parlament és a Tanács 2016/679 Rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról*, 04: 27, 2016.
- [7] Nemzeti Adatvédelmi és Információszabadság Hatóság, „A Nemzeti Adatvédelmi és Információszabadság Hatóság tájékoztatója a munkahelyi adatkezelések alapvető követelményeiről,” [Online]. Elérhető: https://naih.hu/files/2016_11_15_Tajekoztato_munkahelyi_adatkezelesek.pdf.
- [8] 187/2015. (VII. 13.) Korm. rendelet az elektronikus információs rendszerek biztonsági felügyeletét ellátó hatóságok, valamint az információbiztonsági felügyelő feladat- és hatásköréről, továbbá a zárt célú elektronikus információs rendszerek meghatározásáról, Budapest: Magyarország Kormánya, 2015.
- [9] „Australian Government Department of Home Affairs,” 11 2020. [Online]. Elérhető: <https://www.homeaffairs.gov.au/reports-and-pubs/files/exposure-draft-bill/exposure-draft-security-legislation-amendment-critical-infrastructure-bill-2020-explanatory-document.pdf>. Hozzáférés dátuma: 2021.03.11.



KIFÜ nemzetközi együttműködései és projektjei

Mohácsi János

ORCID: [0000-0002-1847-4437](https://orcid.org/0000-0002-1847-4437)

KIFÜ, nemzetközi kapcsolatokért felelős vezető

mohacsi.janos@kifu.gov.hu

Abstract

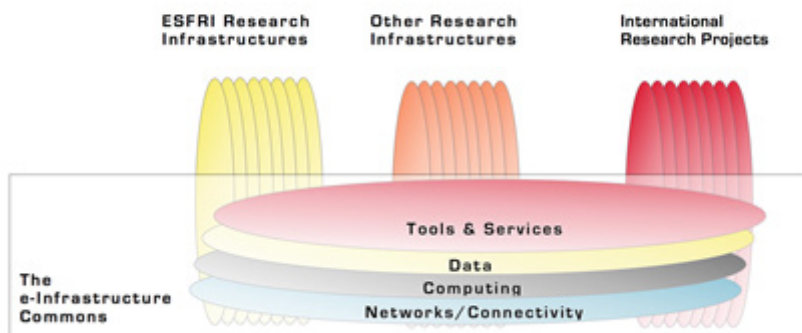
The professional work under the NIIF Program has been a key enabler for Hungarian research and education since 1986. The KIFÜ's infrastructure is part of the GÉANT (European ICT Infrastructure) and EuroHPC and PRACE (European Supercomputing Infrastructure) infrastructures, multimedia information management, resource and service virtualization, federated authentication, etc. ensure effective and full connectivity with all European research institutions and researchers, including all European research infrastructures (ESFRI, ERIC and others). This paper presents the international cooperation and projects of the KIFÜ and the NIIF Program.

Keywords: infrastructures, networking, computing, NREN

Bevezetés

Az NIIF (korábban IIF) Program keretében folyó országos jelentőségű szakmai munka 1986 óta meghatározó kiszolgálója a hazai kutatásnak és oktatásnak, de egyúttal jelentős speciális szegmense a magyarországi informatikának is. Ez a munka, amely a 90-es évek eleje óta nemzetközi együttműködésben a nemzetközi élvonallal együtt haladva végez az I²F, majd NIIF közösség. Nemzetközi szinten is jellemzően a kutatói hálózatokra épülő e-Infrastruktúrák¹ világszerte virágzanak, fejlesztésük és használatuk egyre intenzívebb és szélesebb körű (kutatásbázisú innováció, ipari együttműködések, nyitott tudomány, adat-infrastruktúrák, szuper-számítástechnika, virtuális környezetek és szervezetek, globális kollaboráció stb.). Az e-infrastruktúrák horizontális módon biztosítanak adatmozgatási, adathozzáférési, adatfeldolgozási és egyéb együttműködést támogató szolgáltatásokat minden diszciplináris kutatásnak, tudományos projekteknek beleértve Európa valamennyi (ESFRI, ERIC és egyéb) kutatási infrastruktúráját is.

1 <https://ec.europa.eu/digital-single-market/en/e-infrastructures>



1. ábra: e-infrastruktúrák viszonya az EC e-IRG tanácsadó testület szerint

Az elmúlt 35 év forradalmi új szakaszt indított be és alapozott meg a tudományos kutatás technológiája, metodikája és innovációs hatása, ezen keresztül a gazdasági versenyképességre és a társadalmi jólétre gyakorolt hatása tekintetében.

A KIFÜ nemzetközi együttműködései

A kutatás-oktatás-közgyűjtemények magyarországi IKT infrastruktúrája (az utóbbi években szokásossá vált európai szóhasználattal e-infrastruktúrája vagy digitális infrastruktúrája) szempontjából a nemzetközi kapcsolatoknak továbbra is kiemelkedő a jelentőségük. A kapcsolatokat három jellemző alapján értékelhetjük:

1. a külföldi partnereinkkel együttműködésben üzemeltetett szolgáltatások,
2. a fejlesztési irányok tekintetében meghatározó jelentőségű szervezeti együttműködések,
3. a nemzetközi együttműködésben megvalósított projektek.

E három láb elsősorban személyes és elektronikus információcsere keretében támogatja kapcsolatainkat. 2020 év elején Magyarországon és Európában is megjelent a COVID-19 vírus, így a személyes kapcsolattartással szemben az elektronikus kapcsolattartás (videokonferencia, e-mail, különböző webes technológiák) vált meghatározóvá.

A program nemzetközi együttműködése stabil maradt és az oktatás, kutatás és közgyűjteményeknek nyújtott szolgáltatások fenntarthatósága szempontjából pedig jól működtek. A COVID-19 járvány által okozott jelenlegi helyzet mindannyiunk számára példátlan, a mindennapi élet európai és globális szintű megzavarásával szembesülünk. 2020-ban kihívásokkal teli időkben a GÉANT munkatársai, partnerei és az NREN közösség tagjai- köztük a KIFÜ összefogtak, hogy hasznos információkat és forrásokat gyűjtsenek és osszanak meg, biztosítsák az oktatás és kutatás számára alapnak tekintett szolgáltatások biztonságos nyújtását, és olyan gyorsan új javaslatokat dolgozzanak ki, amelyek támogatják kutatási és oktatási közösségünket a pandémiás válságban².

2 https://www.geant.org/People/COVID-19_Community_Support

A HUNGARNET és a KIFÜ együttműködése hozzájárult a felhasználók és a fejlesztés-szolgáltatás szempontjait is érvényesítő kapcsolatrendszerünk stabilitásához, ezen kapcsolatrendszerünknek is köszönhető, hogy sikerült továbbra is megtartani a nemzetközi együttműködések szempontjából is meghatározó jelentőségű integráltságot.

Az integráltság különösen fontos a kutatás és oktatás hazai e-infrastruktúrájának feltételrendszere, szolgáltatási spektruma és főként alkalmazói-alkalmazási teljessége terén. Mindez – a 2016-ban kialakult helyzetből adódóan – 2020-ban is túlnyomórészt a KIFÜ ernyője alatt valósult meg, kapcsolódva a HUNGARNET által is képviselt felhasználói érdekekhez-szempontokhoz, de jó kapcsolatokat ápolva a kutatás számára az e-infrastruktúrába illeszkedő szolgáltatásokat nyújtó KIFÜ-n kívüli partnerekkel is, mint például a OpenAire³ és egyéb adatrepozitóriumok.

Az oly fontos integráltság, mely szerencsés módon az elmúlt évben is Európa-szerte irigyelt ritka előnye és stabilitási tényezője maradt az NIIF Program céljaihoz illeszkedően működtetett itthoni e-Infrastruktúrának, jelentős mértékben hozzájárult nem csak ahhoz, hogy a szolgáltatások gyakorlatilag "one stop shopping" elvet követően legyenek elérhetők, hanem valóban, maguknak a nemzetközi kapcsolatoknak a fenntartásához-fenntarthatóságához is. Itt érdemes megjegyezni, hogy egyre jellemzőbbé kezd válni, az e-infrastruktúrák összekapcsolódása, illetve egyre szorosabb együttműködése. Az előbb leírt helyzetet kell továbbra is fenntartanunk, a korábbiakhoz hasonló tudatos építkezés útján, az infrastruktúra fejlesztése és alkalmazása, valamint a szolgáltatások nyújtása terén is, maximálisan kihasználva – mind az üzemeltetői, mind pedig a felhasználói oldalon – a benne rejlő lehetőségeket.

Igen kedvező alapot jelentett 2020-ban is az említett kedvező helyzet fenntartása szempontjából, hogy a nemzetközi kapcsolatok "három lábon álló" stabil működésében továbbra is többé-kevésbé egyenlő súlyú, kiegyensúlyozott szerepet játszott a projekt-kapcsolatok, a szervezeti kapcsolatok, valamint a (hálózati és egyéb) szolgáltatási kapcsolatok hármassága.

A KIFÜ nemzetközi projekt részvételei

Ahogy korábban, a projekt kapcsolatokat tekintve 2020-ban is kulcsfontosságú volt nemzetközi együttműködéseink szempontjából az EU Horizon2020 programja keretében futó GÉANT és PRACE "zászlóshajó" projekteken való KIFÜ részvétel és ezeken keresztül kapcsolódásunk Európa kutatói hálózati és HPC tárgyú együttműködéseinek az élvonalához. Emellett megjelent a Nyílt Tudomány támogatása is mely több projektben, több különböző résztvevővel valósult meg (Ni4OS-Europe – Debreceni Egyetem, KIFÜ; OCRE- KIFÜ GÉANT kedvezményezettként; OpenAIRE- Debreceni Egyetem; ExPaNDS – SZTAKI EGI alvállalkozójaként, EOSC-Future - KIFÜ alvállalkozóként).

3 <https://www.openaire.eu/>



2. ábra: EC támogatású Nyílt Tudományosságot támogató projektek

2019-ben elindultak a 2022-ig tartó GN4-3⁴ és GN4-3N projektek, melyek többek között a gerinchálózat új generációjának a kiépítését tűzték ki célul. Mindez lehetővé teszi olcsóbb, illetve NREN-ekre épülő megoldások (Spectrum Connection Service, Whitebox router, DCI) alkalmazását. A finanszírozást tekintve viszont kedvező, hogy ez a projekt is az európai NREN-ek közössége és az EC által aláírt FPA (Framework Partnership Agreement, azaz együttműködési keretmegállapodás) feltételrendszerének megfelelően működik. A GN4-3 és a GN4-3N projekt hagyományos célja, hogy a lehető legmodernebb és költséghatékonyabb hálózati technológiák alkalmazásával a legfejlettebb adathálózati szolgáltatásokat lehessen kialakítani Európában.

KIFÜ GN4-3(N) Hálózat építés

- 7 fázis
- 14 országról → 24+ ország – sötétszál technológiára építve
- Hálózati eszközök
 - Keretprogram
 - Line System – DCI
 - Kapacitás 5x növekedés
 - Helyigény 5. része
- GN 5. generációjának létrehozásában GÉANT implementációs csoport szorosan együttműködik az NREN-ekkel és a külső szolgáltatókkal a telepítések kapcsán

Map legend:
- Procured (green line)
- Out for tender (orange line)
- NAC endorsed (yellow line)
- First option (dashed blue line)
- Alternative options (dotted blue line)
- Replacing current with long term (dashed red line)
- Additional routes (dotted red line)
- DWDM only PoP (blue circle)
- DWDM/IP MPLS PoP (orange circle)
- Non GEANT PoP (grey circle)
- GEANT PoP (red circle)
- Fibre attaching points (blue square)
- Alternative PoPs / Fibre attaching points (orange square)

Map shows network routes across Europe, including countries like France, Germany, Spain, Italy, UK, and others.

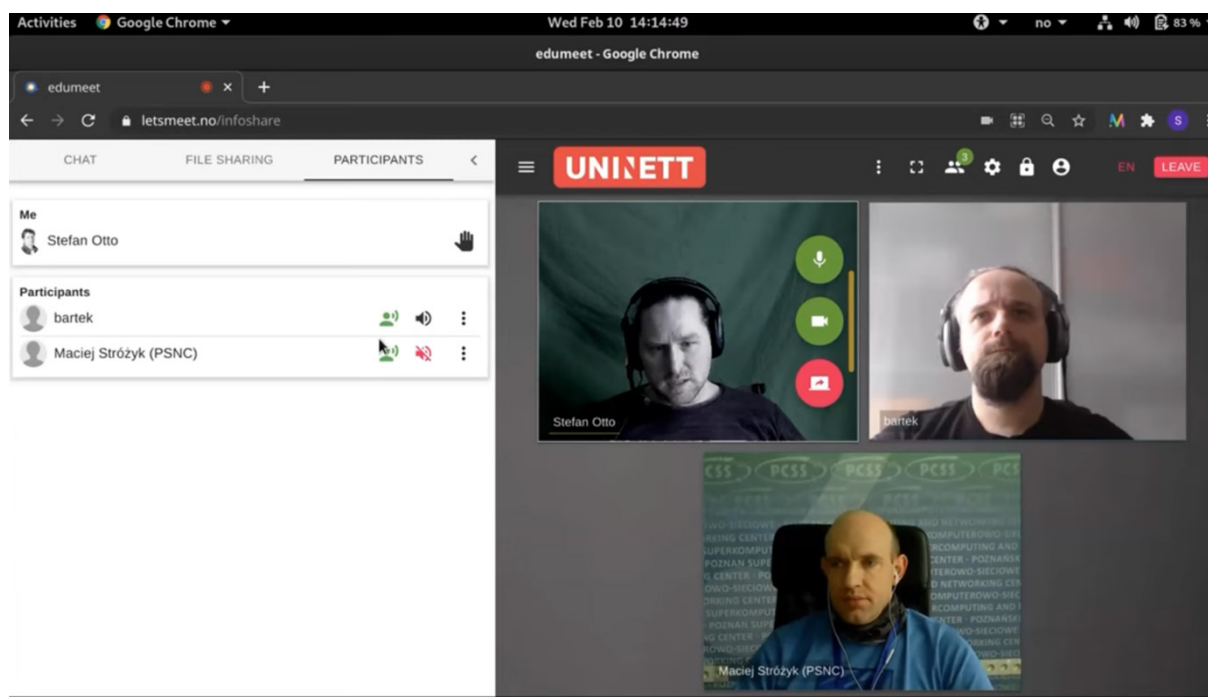
GÉANT

3. ábra: GN4-3(N) hálózat építés

4 https://www.geant.org/Projects/GEANT_Project_GN4-3

A GN4-3 egyébként folytatta a hagyományos GÉANT szolgáltatások mellett a 2018-ben beindított cloud-infrastruktúra szolgáltatást is, mely a legfontosabb kereskedelmi felhő- és storage-szolgáltatókkal kötött speciális megállapodások keretében, kedvezményes feltételek között érhető el az NREN-eken keresztül az európai kutatási és oktatási közösségek számára. Magyar részről a KIFÜ keretében folytatódott a hazai fejlesztésű oktatási felhő szolgáltatás és a GÉANT szolgáltatás közötti együttműködés minél kedvezőbb (minden tekintetben kölcsönösen előnyös) kidolgozására irányuló munka. A GÉANT fejlesztések kapcsán kiemelendők a T&I (Trust and Identity) fejlesztések, különös tekintettel az eduGAIN-re és eduTEAMS-re, valamint a rendkívül erős multimédia szolgáltatás-fejlesztés, és a white-box routereket támogató freertr portolás, hiszen magyar részről ezek keretében szintén jelentős hozzájárulások voltak a projektben.

Érdeemes külön kiemelni az edumee^{5t} fejlesztést, amely egy nyílt forrású videokonferencia rendszer, amely hozzájárult a COVID-19 időszakban sok oktatási és kutatási intézményben ahhoz, hogy tovább folytatódhasson a munka.



4.ábra: edumeeet felhasználói felület

A PRACE mellett egyre meghatározóbb az EuroHPC együttműködésben való koordinációs-kooperációs kapcsolódás. A korábbi évek NIIF Intézeti és később KIFÜ erőfeszítéseinek eredményeként a HPC terület NKFIH támogatási ígérvényt kapott, mely lehetővé tette az EuroHPC csatlakozóknak- köztük Magyarországnak- a nemzetközi pályázást. Magyarországon a KIFÜ vezetésével HPC kompetencia központ jött létre, 2020 nyarától kezdve. Ennek a hazai munkának a nemzetközi koordinációját biztosítja a 2020 szeptemberétől működő EuroCC projekt. A PRACE2 Együttműködési Program (Partnership for Advanced Computing in Europe) 2020-ben megfelelő alapot biztosított a GÉANT-hoz hasonlóan magas prioritású részvételhez a PRACE együttműködésben,

5 <https://edumeeet.org/>



ezzel kiváló HPC környezetet biztosítva a magyar e-Infrastruktúra fejlesztő és működtető, valamint alkalmazó közösségnek. 2020-ban folytatódott a PRACE-6IP⁶ projekt, mely európai szuperszámítógép központok együttműködésének fokozását tűzte ki célul.

AGÉANT-ra épülő európai hálózati kapcsolatok és az ezekre támaszkodó e-Infrastruktúra szolgáltatások (cloud, HPC, VoIP, VC, WebRTC, data) terén – bár lényeges változásra a szolgáltatásokban általában nem került sor – a nemzetközi kapcsolatok tovább gazdagodtak, szélesedtek és mélyültek, de legalábbis szinten maradtak, ezáltal kedvező módon bővültek az együttműködési lehetőségek. E lehetőségek kiaknázására a KIFÜ-ben az elkövetkezőkben is komoly figyelmet kell fordítani.

A 2020-as év során folytatódtak a munkák az adat-infrastruktúrák és az infrastruktúra globalizáció terén. A 2019 szeptemberében elindult új Horizon2020 projekt, a NI4OS-Europe⁷, melynek célja, hogy az Open-Science együttműködések erősítse a Nyugat-Balkántól egészen Kelet-Európáig, folytatódott.



5.ábra: NI4OS-Europe partnerek

Ennek a projektnek a fő célkitűzése a helyi nemzeti és regionális lehetőségek kiaknázása mellett, az Open Science irányelvek kialakítása, az EOSC – Európai Nyílt Tudományos Együttműködés szolgáltatásaihoz (on-boarding) történő csatlakozás támogatása, a FAIR - Findability, Accessibility, Interoperability and Reusability (megtalálható, hozzáférhető, szabványosak és újrafelhasználható) adatkezelési alapelvek adaptálásának megkönnyítése, valamint konzultációk, tréningek szervezése az érintett csoportok számára. Ezt egészíti a 2021 áprilisában indult EOSC-Future project, melynek az a célja, hogy működő integráló EOSC platformot („Rendszerek szövetsége”) hozzon létre, amely adatokból, szakirányú tudományos szolgáltatásokból, valamint nyitott kutatási termékekből és infrastruktúrából

6 <https://prace-ri.eu/>

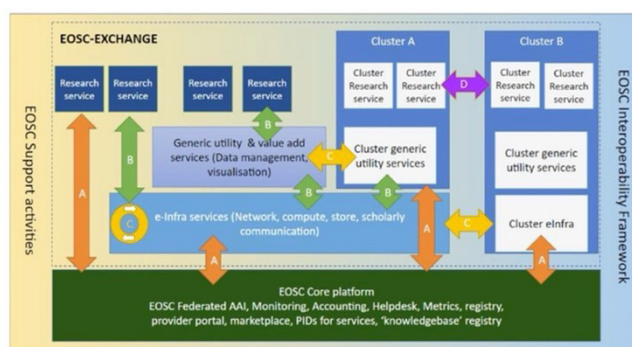
7 <https://ni4os.eu/>

áll, amelyekhez a résztvevő Európai kutatók támogatást és képzést kapnak az EOSC erőforrásainak és megoldásainak felhasználásához.

KIFÜ Tudományos klaszterek támogatása

ESFRI Science Clusters :

- **ENVRI-FAIR** – environmental sciences
- **EOSC-Life** – life sciences
- **ESCAPE** – astronomy and particle physics
- **PaNOSC** – photon and neutron sciences
- **SSHOC** – social sciences and humanities



6.ábra: EOSC-Future projekt struktúrája

Az EOSC által is kiemelten kezelt Open Science nemzeti iniciatíváinak (open science cloud fejlesztéseknek és nemzeti és tudományterület EOSC „portálok” kialakításának, valamint FAIR elvet követő adat-infrastruktúra építéseknek) az összehangolására, a szolgáltatások integrálására törekvő projekthez kapcsolódó tényleges munkák elindultak. 2020 folyamán az RDA (a globális Research Data Alliance) magyarországi node-ja (HRDA) folytatta a munkáját – meetupokat tartott. HRDA-ban, melyben az alapító tagok között szerepel a KIFÜ is, képviselve az adat-infrastruktúrák háttérét biztosító hazai e-infrastruktúra fejlesztő és szolgáltató közösséget.

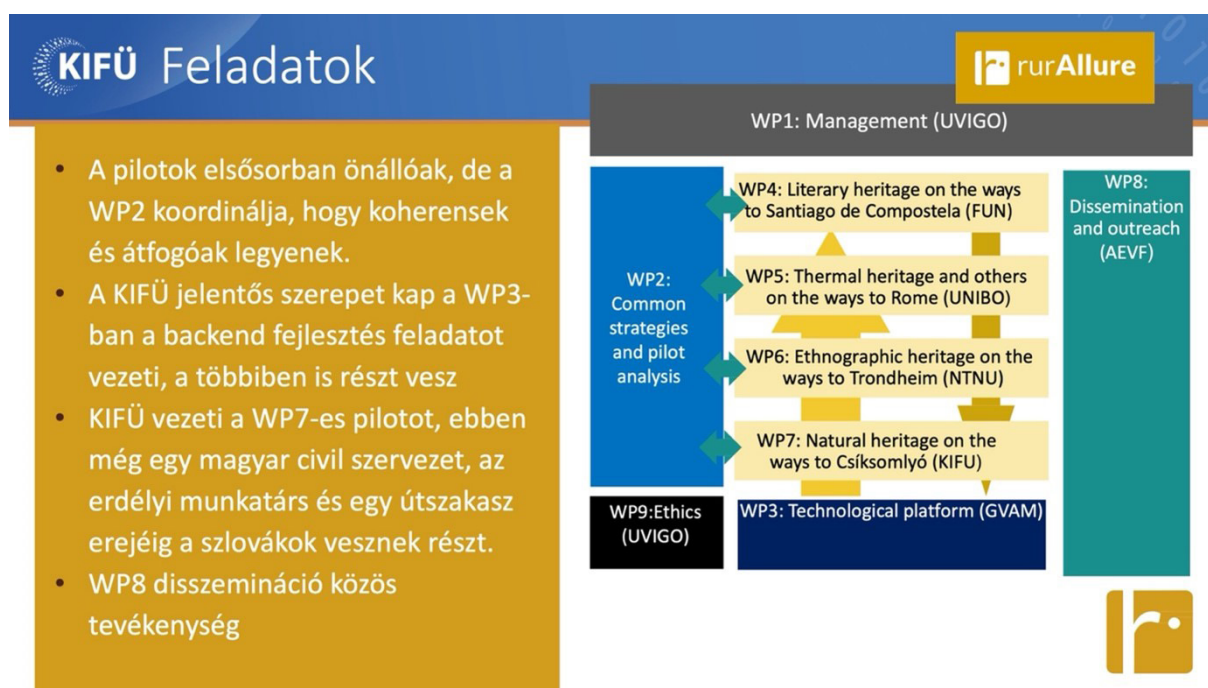
Az oktatási infrastruktúra és szolgáltatások fejlesztésének fontos projektje az Up2U 2020-ban befejeződött. Fő feladatként az európai partnerekkel együttműködésben a KIFÜ egy olyan oktatás támogatási eszközcsoportot fejlesztett ki, mely a közoktatás számára elérhetővé teszi és megújítja az NIIF Program szolgáltatásait.

Az előbb felsorolt projektek ugyanis – bár kevésbé voltak átütőek a nemzetközi kapcsolatok szempontjából, mint az említett ”zászlóshajó” projektek, hozzájárultak a nemzetközi forrású hazai bevételek szerény növeléséhez, és ami még fontosabb, segítették szemléletünk és felkészültségünk gazdagítását. Egyrészt jelentős mértékben hozzájárultak a felfutóban lévő adat-infrastruktúra témakörbeli tapasztalataink bővítéséhez, másrészt egy valóban globális kitekintést tettek lehetővé. Eközben több fejlődő régió támogatásában vehettünk részt és a jövőt tekintve újabb ismeretekre tehattünk szert olyan térségekben, melyek egy része néhány év alatt ledolgozhatja hátrányát és kiváló együttműködési partner lehet a fejlesztéseinkben, itthon is és európai



léptékben is. Érdeemes itt megemlíteni az EAP és CAREN – a kelet-európai és közép-ázsiai régiók e-infrastruktúra fejlesztéseinek támogatására irányuló – törekvéseket, melyekben szerény hozzájárulással ugyan, de amennyire módunk van rá, szintén igyekszünk fenntartani a közreműködésünket.

Ezt egészíti ki egy 2021 januárjában indult rurAllure⁸ projekt amelynek célja a kulturális együttműködés és a fenntartható turizmus elősegítése az európai történelmi zarándokútvonalak térségében. A célok eléréséhez a projekt korszerű informatikai megoldásokat alkalmaz a vidéki múzeumok és örökségi helyszínek népszerűsítésére.



7.ábra: rurAllure projekt feladatai

Összegzés

Mint az előző években, a munkák most, 2020-at követően is folytatódnak, a feladatok bővülnek, az e-Infrastruktúra közösségek felelőssége nemzeti és nemzetközi szinten is egyre nő. A KIFÜ infrastruktúrája a GÉANT (az európai IKT infrastruktúra) valamint az EuroHPC és PRACE infrastruktúra (a szuperszámítástechnika európai infrastruktúrája) részeként, a multimédia információkezelés, az erőforrás és szolgáltatás-virtualizáció, a szövetségi azonosítás stb. eszközeire építve hatékony és teljes értékű kapcsolódást biztosít az európai kutatás összes intézményéhez és kutatójához, beleértve Európa valamennyi (ESFRI, ERIC és egyéb) kutatási infrastruktúráját is. Ezek a korábbiakban is említett 3 komponensű rendszer egyrészt garantálja a nemzetközi kompatibilitást, másrészt lehetőségeket teremt a versenyképes hazai kutatás számára a nemzetközi kutatásba való bekapcsolódásra. Ebben a soknemzetiségű kapcsolatrendszerben továbbra is fontos a hazai együttműködés biztos alapokra épülő megszervezése és működtetése, egyebek mellett az NIIF Program végrehajtásáért felelős KIFÜ, valamint a

⁸ <https://rurallure.eu/>

felhasználókat, a felhasználói közösséget tömörítő és annak igényeit közvetítő szakmai érdekvédelmi szervezet, a HUNGARNET Egyesület együttműködésével, a közöttük kialakított és folyamatosan fejlesztendő kapcsolatból eredő lehetőségek kiaknázásával.



Komplex informatikai megoldás természettudományos ismeretek bemutatására az egri Varázstoronyban

Horváth Krisztián
hkrisztian1997@gmail.com

Keller Zsófia
kellerzsofi2@gmail.com

Oravecz Zsolt
zsolti1012@gmail.com

Sass-Gyarmati Norbert
sassnorbert99@gmail.com

Geda Gábor
geda.gabor@uni-eszterhazy.hu

Eszterházy Károly Egyetem, Informatikai Tanszék

Absztrakt: Magyarországon az 1970-es évek elején vette kezdetét a múzeumokban az a változás, melynek eredményeként a gyermek- és az ifjúsági korosztály kiemelt szerepet kapott. Célja a múzeumokban felhalmozott gyűjtemény és szellemi érték népszerűsítése az említett korosztályok számára. A fentieket alapul véve, azt tűztük ki célul, hogy az egri Természettudományos Pályaorientációs és Módszertani Központot választva, az ott működő Varázstorony eszközeit újítjuk meg, és ott alakítjuk ki az általunk tervezett és fejlesztett rendszert, egyben javaslatot téve a látogatási rend módosítására is. Egy olyan komplex hardver és szoftver rendszert hoztunk létre, melynek segítségével a látogatók a saját mobileszközük segítségével tudják bejárni és megismerni a Varázstorony épületét, kiállításait, valamint kísérleteket végezni olyan módon, hogy az arra alkalmas eszközöket például telefonjuk segítségével vezérlik.

Kulcsszavak: Komplex IT rendszer, Tudománytörténet, Múzeumpedagógia, Interaktív kísérletezés

A complex IT solution for presenting scientific knowledge in the Varázstorony in Eger

Abstract: In Hungary, a conscious activity with children and young people in museums began in the early 1970s. It aims to promote the collection and intellectual value accumulated in museums. By the basic concept, we set the goal of choosing the Eger Science Career Orientation and Methodology Center, more precisely renewing the tools of the Varázstorony operating there, and developing the system we planned and developed there, at



the same time proposing changes to the visit order. We have created a complex hardware and software system that allows visitors to explore and learn about the Varázstorony's building, exhibitions, and experiments on their own mobile devices by controlling the appropriate devices using their phones.

Keywords: Complex IT system, History of science, Museum pedagogy, Interactive experimentation

1. Bevezetés

Az Varázstorony[1] az egri Líceum épületének része, amely az Eszterházy Károly Egyetem fő épülete is egyben. 1776-ban kezdte meg működését, mint csillagvizsgáló és azóta is az oktatás és a tudományok szolgálatában áll. Ennek köszönhetően rendkívül sokszínű múzeumi anyag gyűlt itt össze az elmúlt évszázadok során. Mindezzel együtt úgy gondoltuk, hogy szükséges lehet a kiállítási anyag korszerűsítése, hogy elsősorban interaktivitás szempontjából jobban megfelelhessen a 21. századi elvárásoknak.

2. Célkitűzéseink

Alapvető célunk az ismeretátadás hatékonyságának fokozása, az interaktivitás és ezen keresztül az élményszerűség biztosítása. Már ezt megelőzően is rendelkezünk néhány interaktív kiállítási eszközzel, azonban szerettük volna az interaktivitást magasabb szintre emelni. Amennyiben a régi eszközök alapkoncepcióknak megfelelő átalakítása valamilyen oknál fogva nem megvalósítható – például műszaki paraméterek, állapot vagy gazdaságossági szempontok miatt – akkor a szóban forgó eszközzel az alapfunkciók szempontjából egyenértékű új eszközt állítunk elő és annak funkcionális kiegészítését végezzük el a továbbiakban is.

A funkcionális kiegészítés alatt azt értjük, hogy az alapfunkciók megtartása mellett olyan lehetőségeket biztosítunk, amelyek révén a korábban is megszerezhető ismeretek mellett újabb ismeretek elsajátítását teszik majd lehetővé, vagy a korábbi funkcióinak lesz képes pontosabban, esetleg gyorsabban, a felhasználó és/vagy az eszköz szempontjából biztonságosabban, vagy egyszerűen csak látványosabb formában megfelelni.

Továbbá célként jelöltük meg azt is, hogy a rendelkezésre álló eszközzel olyan szempontok figyelembevételével bővítsük, hogy azokkal lehetővé váljon további tudományterületekhez (elsősorban kémia, matematika, informatika, esetleg biológia) sorolható ismeretek közvetítése is. Az egyszerűbb megvalósítás és az eszközök gazdaságosabb üzemeltetése miatt elhatároztuk, hogy a reverzibilis változásokon alapuló kísérletek bemutatását fogjuk előnyben részesíteni.

3. Múzeumpedagógia

Ha iskolán kívüli tanulmányi kirándulásokat szervezünk egy interaktív kiállításra, akkor a diákoknak olyan élményt biztosíthatunk, amelyre nem biztos, hogy lenne lehetőségünk az iskola falai között [3].

Jól tudjuk továbbá, hogy a látogatók tényleges bevonása egy-egy eszköz, berendezés kipróbálásába azért fontos, mert így hosszabb ideig fenntartható az érdeklődők figyelve, mint amikor például oktatófilm segítségével közvetítjük a kívánt ismereteket.

4. Korábbi kutatások

A fentebb megfogalmazott szempontoknak megfelelően elsőként a jelenleg is rendelkezésre álló eszközöket vizsgáltuk meg. Ezzel kapcsolatban arra az eredményre jutottunk, hogy bizonyos, már hosszú évek óta működő, éppen ezért kissé „elhasználódott” eszközök átalakítását azok összetettsége miatt, valamint gazdaságossági szempontokat is figyelembe véve, későbbre halasztjuk. Ugyanakkor például a Cartesius-búvár [2] megújítását úgy látjuk lehetségesnek, ha egy új eszközt hozunk létre, amely műszaki paramétereit tekintve jobban megfelel majd az új funkcióknak is.

A fentiek mellett szükségesnek és lehetségesnek tartottuk további jelenségek bemutatására alkalmas, új eszközök fejlesztését és azok rendszerbeállítását is, bővítve ezzel a szemléltethető ismeretanyagok körét.

Az új eszközöket úgy kellett kialakítani, hogy használatuk az interaktivitáson alapuljon, a szemléltetni kívánt jelenség egyszerű bemutatásán túl (kvalitatív információk közvetítése) alkalmasak legyenek mérési feladatok, mérő kísérletek kivitelezésére is.

Kutatásaink során más interaktív kiállításokkal, létesítménnyel is találkoztunk (Mobilis Győr, Csodák palotája, Szépművészeti Múzeum...).

Bár napjainkban már számos interaktív múzeumot tekinthetünk meg, melyek minden korosztály érdeklődését felkelthetik, olyan komplex rendszer, amely az általuk tervezett módon segíti a látogatót, még nincs az országunkban.

5. A rendszer tervezése

A rendszer kialakításánál fontos szempontnak tartottuk az alacsony költségvetést is, ezért terveztük az interaktív vezérlést különböző, azonban pénztárca barát mikrokontrollerekre építve.

A rendszer másik fontos alkotóeleme a szerver, aminek segítségével a felhasználói tevékenységeket kezelhetjük, valamint összeköthetjük őket az eszközökkel. Erre a célra egy Raspberry Pi nevű apró számítógépet választottunk [4], mely ára és kis mérete ellenére rendkívül jól használható erre a célra is.



Célul tűztük ki továbbá, hogy ahol ez lehetséges és/vagy szükséges, ott az eszközökkel való kommunikációt a látogatók saját mobileszközeikkel valósíthatják majd meg. Ez a lehetőség – megítélésünk szerint – újszerűségénél fogva is nagyobb érdeklődésre tarthat majd számot, és nem utolsó sorban költséghatékonyabb megoldást is jelent egyben a rendszer megvalósítása szempontjából is, hiszen így nincs szükség költséges kezelői felület kialakítására az egyes eszközök esetében.

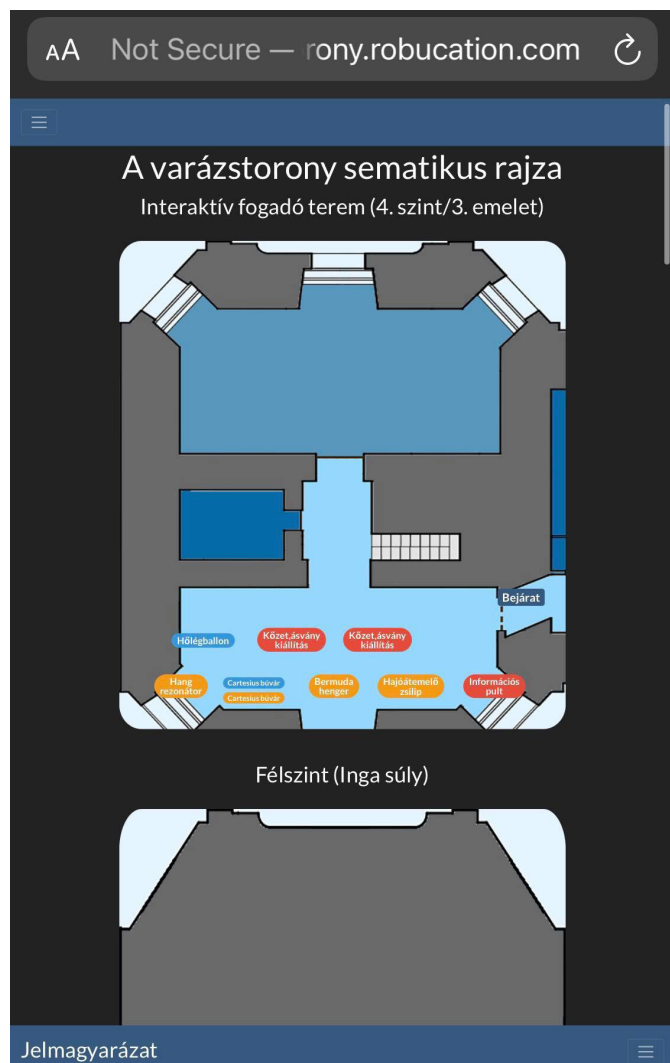
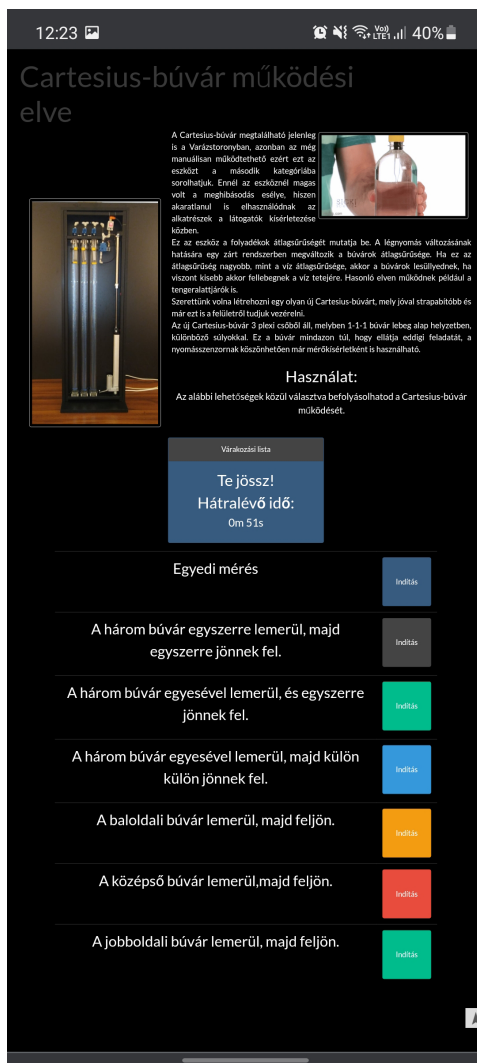
5.1. Megvalósítandó funkciók

A korábban ismertetett szempontok alapján a fejlesztésre kerülő eszközökkel kapcsolatosan az alábbi néhány, egymástól jól elválasztható feladatkör jelölhető ki:

- kísérletek és a nekik megfelelő kísérleti eszközök tervezése, kifejlesztése,
- a kísérleti eszközök elektronikai fejlesztése, valamint azok vezérlő szoftverének programozása,
- a látogatók számára a kísérleti eszközök elérését biztosító mobil alkalmazások fejlesztése,
- a rendszer hatékony alkalmazását biztosító módszertani ajánlások kidolgozása, használat során szerzett tapasztalatok alapján történő módosítása.

5.1.1. A keretrendszer funkciói

- Adminisztrációs funkciók: Miután a látogatók bejelentkeztek és kapcsolódtak a Varázstorony vezeték nélküli hálózatához, várólistákon keresztül érhetik el a vezérelni kívánt eszközöket.
 - Csoportok létrehozása:
Előfordulhat, hogy egy osztály (vagy más csoport) látogat el a múzeumba. A kísérő tanár létrehozhat egy csoportot, majd hozzáadhatja a gyerekeket.
 - Pontok gyűjtése:
A látogatók különböző kísérletek elvégzésével, vagy tesztek kitöltésével pontokat gyűjthetnek, melyeket különböző tárgyak vásárlására fordíthatnak, de a pontok révén van lehetőség verseny meghirdetésére is a csoport tagjai között.
- Interaktív vezérlés: A rendszer fő előnye, hogy a látogatók a saját mobileszközük segítségével tudnak kísérleteket végezni. Az egyes interaktív eszközökhöz, egy-egy speciális felület tartozik, ahol különböző paraméterek megadásával tudják végrehajtani az általuk eltervezett kísérletet. Egyes eszközök esetében rendelkezésükre állnak különféle demó programok is, amelyekkel előre megtervezett kísérleteket hajthatnak végre (1. ábra 1. kép).
- Idegenvezetés: Bejelentkezés után idegenvezető igénybevétele nélkül járhatjuk be a Varázstorony termeit, ami az alkalmazásba integrált térképek segítségével valósítható meg (1. ábra 2. kép).



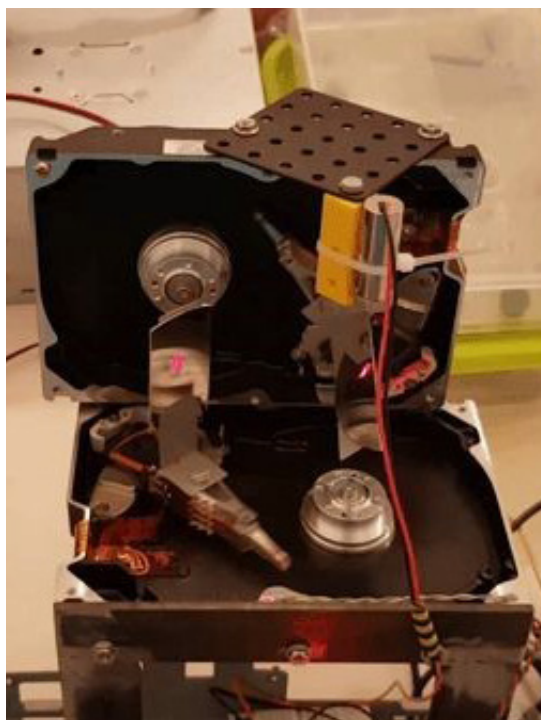
1. ábra A Cartesius kezelői felülete (bal oldalt) és az idegenvezetéshez szükséges térképek (jobb oldalt)
(Forrás: saját szerkesztés)

Az eszközök fejlesztése során fontos szempontként kellett figyelembe vennünk, hogy csak olyanokat alakítsunk át, melyek nem veszítenek tudományt népszerűsítő tulajdonságukból. Erre a legjobb példa a Cartesius-bűvár, mely esetében új eszközt készítettünk, és így már az eddigi funkcióin túl mérőkísérletként is szolgálhat a használata, valamint felruháztuk játékos funkcióval is. Ezt az elvet a többi eszköz tervezése és megvalósítása során is igyekszünk szem előtt tartani.

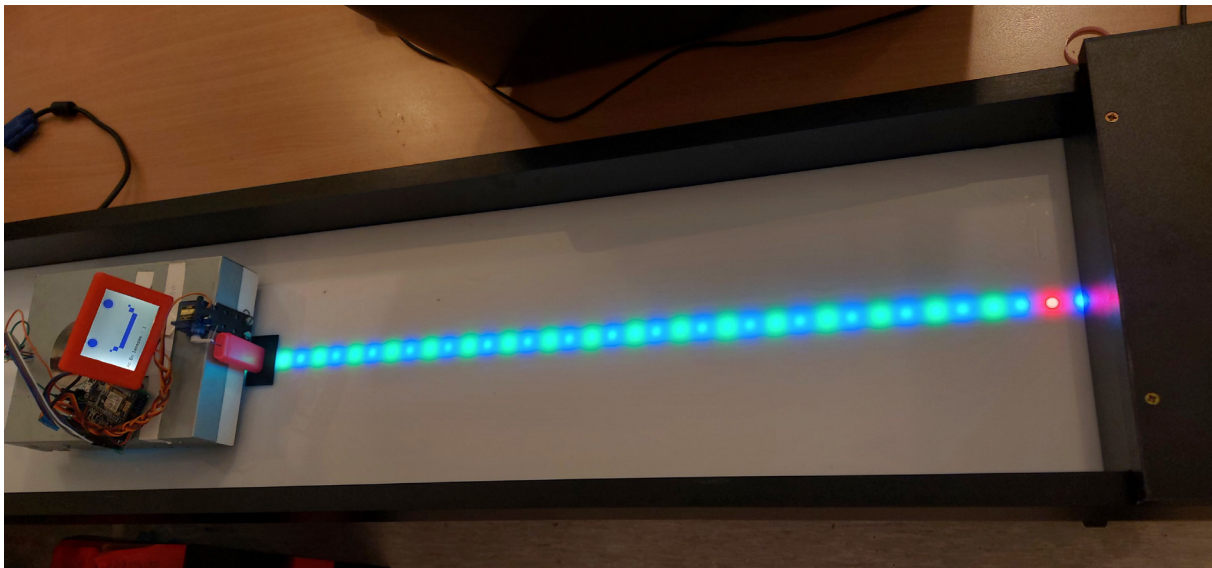
6. Általunk fejlesztett eszközök bemutatása

A rendszer tervezése és fejlesztése során az eszközöket az alábbi kategóriákba soroltuk. Az első csoportba azok a hagyományos eszközök tartoznak, melyeket nem szerettünk volna felokosítani, mert veszítettek volna tudományt népszerűsítő funkciójuk hatékonyságából. Ezeknél az eszközöknél, csak egy leírás jelenik meg az alkalmazásban, vagy esetleg egy rövid kis kvíz, mellyel pontszerzésre van lehetőség. A második csoport a továbbfejlesztett eszközök csoportja, melyeken változtatásokat hajtottunk végre, de funkciójánál fogva nem volt értelme vezérelhetővé tenni (például Hell-harang [8] [9]). Ebben az esetben is egy leírás jelenik meg az adott eszközről, illetve a fejlesztésről. A harmadik csoport pedig az interaktív eszközök csoportja, melyről a leíráson túl egy vezérlő felület is megjelenik. Az ebbe a csoportba tartozó eszközöknél egy mikrokontrolleres vezérlőn keresztül tudjuk egy adott kísérlet elvégzésére utasítani az eszközt. Ebben a csoportban már vannak olyan eszközök, amelyek már korábban helyet kaptak a Varázstoronyban valamilyen formában, illetve olyanok is, melyeket mi fejlesztettünk új eszközként.

A Varázstoronyba számos új eszköz kerül bevezetésre, mint például az általunk továbbfejlesztett Hell-harang [6][7][10][11] (2. ábra 1. kép), a Cartesius bűvár (4. ábra 2. kép), Robi a robot (3. ábra), a lézer-show (2. ábra 2. kép), a terepasztal vagy a lejtő [5].



2. ábra A megújult Hell-harang (bal oldalt) és a lézer-show mechanikus szerkezete (jobb oldalt)
(Forrás: saját szerkesztés)



3. ábra Robi a robot játék közben (Forrás: saját szerkesztés)

Ezek közül az eszközök közül a Cartesius bűvár bemutatásával szeretnénk szemléltetni az okos interaktív eszközök csoportját.

6.1 Cartesius bűvár

A Cartesius-bűvár egy manuálisan működtethető változata megtalálható volt már korábban is a Varázstoronyban (4. ábra 1. kép). Az eszköz maga, különféle fizikai jelenségek, mennyiségek és törvényszerűségek tanulmányozását teszi lehetővé. Ilyenek például az átlagsűrűség, úszás, lebegés jelensége, Pascal törvénye.

A légnyomás változásának hatására egy zárt rendszerben megváltozik a bűvárok átlagsűrűsége. Ha ez az átlagsűrűség nagyobb, mint a víz sűrűsége, akkor a bűvárok lesüllyednek, ha viszont kisebb, akkor felemelkednek a víz felszínére. Hasonló elven működnek például a tengeralattjárók is. Itt különösen nagy volt a meghibásodás gyakorisága, hiszen akaratlanul is elhasználnának az alkatrészek a látogatók kísérletezései során.

Szerettünk volna létrehozni egy olyan új Cartesius-bűvárt (4. ábra 2. kép), mely jóval strapabíróbb és már ezt is a webes felületről tudjuk vezérelni.

Az új Cartesius-bűvár 3 plexi csőből áll, melyben 1-1 bűvár lebeg, különböző súlyokkal „hangolva”, ami azt teszi lehetővé, hogy különböző nyomásértékek esetén kezdenek alámerülni. Ez az eszköz mindazon túl, hogy ellátja eddigi feladatát, a nyomá szenzornak köszönhetően már kísérleti mérést is lehetővé tesz. Az új eszköz képes automatikus kalibrációt végezni minden újraindításnál, majd ezek után egy zárt rendszert hoz létre. Így évszaktól és az aktuális légköri nyomástól függetlenül tudjuk biztosítani azt, hogy minden bűvár ugyanolyan nyomásérték mellett süllyedjen le, amit mi beállítottunk.



4. ábra A régi (bal oldalt) és az új (jobb oldalt) Cartesius bűvár (Forrás: saját szerkesztés)

7. Összegzés

Az 1700-as évek végén még kézzel kellett megszólaltatni a Hell-harangot az egi dél időpontjában. Akkor el sem tudták volna képzelni, hogy ma már egy kis mikrokontroller fogja figyelni a meridián vonalat és amikor a vonal közepére ér a fényfolt, emberi beavatkozás nélkül szólal meg a harang. Nemcsak a fiatalok számára lehet ez nagyon jó visszatekintés, hanem az idősebbeknek is, hiszen annyi minden változott az évek alatt. További terveink között szerepel a Varázstorony vezérelhető eszközeinek bővítése, valamint a fejlesztett rendszernek köszönhetően a felhasználói adatok és a mérési eredmények elemzésével egy olyan statisztika készítése, mellyel még jobban optimalizálhatjuk a rendszert a látogatók igényeihez, ezzel is elősegítve a múzeumok népszerűsítését.

8. Irodalomjegyzék

- [1] Vida József 2016. MEGÚJULT AZ EGRI LÍCEUM VARÁZSTORNYA. *MultiScience - XXX. microCAD International Multidisciplinary Scientific Conference University of Miskolc* <http://docplayer.hu/36041031-Megujult-az-egri-liceum-varazstornya-magic-tower-of-egerlyceum-renewed.html>
- [2] Almádi Péter - Bózvály Bence - Nagy Szabaszián Vilmos - Szabó Dávid 2017. Kísérletezés bárhonnan, bármikor. XXXIII. Országos Tudományos Diákköri Konferencia. Győr.
- [3] Bodnár Dorottya – Jászberényi Melinda - Ásványi Katalin 2017. Az új muzeológia megjelenése a budapesti múzeumokban. *TURIZMUS BULLETIN XVII. évfolyam 1-2. szám* 46-55.
- [4] Simon Monk 2014. *Raspberry Pi Cookbook*. O'Reilly Media. Massachusetts.
- [5] Geda Gábor 2009. Matematikai modellezés és számítógépes szimuláció. *Debreceni Egyetem Természettudományi Doktori Tanács Matematika- és Számítástudományok Doktori Iskola*. https://dea.lib.unideb.hu/dea/bitstream/handle/2437/90900/Geda_G_ertekezes.pdf;jsessionid=AB60C83444F814A13ED1A855A3A3D0C3?sequence=9
- [6] Csaba György Gábor 1997. *A csillagász Hell Miksa írásaiból*. Magyar Csillagászati Egyesület. Budapest.
- [7] Vasné Tana Judit 2001. *Albert Ferenc és az egri csillagásztorony*. Egri Csillagásztorony Védelmében Alapítvány. Eger.
- [8] Horváth Krisztián - Ködmön Zoltán 2020. Traditional timing in the modern age. *Proceedings of the 11th International Conference on Applied Informatics*.
- [9] Horváth Krisztián - Keller Zsófia 2020. Hagyományos időmérés a modern korban. *Agria Média 2020 - ICI 16*.
- [10] Geda Gábor - Vida József 2004. Observation of mechanical movements through virtual experiments. *Proceedings of the 6th international conference on applied informatics (ICAI)*. 67-72. <http://icai.ektf.hu/pdf/ICAI2004-vol2-pp67-72.pdf>
- [11] Biró Csaba - Geda Gábor 2010. Computer-aided quantitative observation of a crystallization process. *Proceedings of the 8th International Conference on Applied Informatics*. 455- 460. <http://icai.ektf.hu/pdf/ICAI2010-vol1-pp455-460.pdf>