

# VALÓS TÉRBEN – AZ ONLINE TÉRÉRT

**Networkshop 31: országos konferencia**

2022. április 20–22.  
Debreceni Egyetem

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület  
Budapest, 2022



A kötet megjelenését támogatta az  
Energiaügyi Minisztérium

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Workshop

2022. április 20–22. Debreceni Egyetem, konferencia előadásainak közleményei

ISBN 978-615-82243-0-7

DOI: [10.31915/NWS.2022](https://doi.org/10.31915/NWS.2022)

Kiadja a HUNGARNET Egyesület  
az MTA Könyvtár és Információs Központ közreműködésével  
Budapest  
2022

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

## TARTALOMJEGYZÉK

<u>Előszó .....</u>	<u>5</u>
<u>Lencsés Ákos: A nyílt tudomány pénzügyi vonatkozásai .....</u>	<u>7</u>
<u>Farkas Katalin: Centenáriumi média-adattár és virtuális kiállítás létrehozásának tanulságai az SZTE Klebelsberg Könyvtárban .....</u>	<u>13</u>
<u>Bódog András: A nyílt archívumi információs rendszer (OAIS) szabványának honosítása.....</u>	<u>20</u>
<u>Perlaki Attila: Oktatást segítő gamifikációs alkalmazások, mint szakdolgozati témák .....</u>	<u>27</u>
<u>Csapó Noémi – Dani Erzsébet: APPropó fejlődés – A Bács-Kiskun Megyei Katona József Könyvtár mobilapplikációja.....</u>	<u>32</u>
<u>Simon András: Integrált könyvtári rendszerek tranzakciós rekordjainak vizsgálata, a könyvtári állomány digitalizálásának tervezésekor.....</u>	<u>41</u>
<u>Németh Márton: Az OSZK Webarchívum nemzetközi kapcsolatai.....</u>	<u>58</u>
<u>Antal Péter: A mesterséges intelligencia kihívásai a XXI. század társadalmára .....</u>	<u>70</u>
<u>Hajdu Csaba – Szilágyi Zoltán: Modern robotikai technológiai ismeretek oktatása „Teljes spektrumú” oktatási módszerrel .....</u>	<u>77</u>
<u>T. Nagy László – Boda István Károly – Tóth Erzsébet: E-tananyagfejlesztés virtuális 3D környezetben.....</u>	<u>84</u>
<u>Palencsárné Kasza Marianna: Digitális átállás – Minőség – lehetőségek az EQAVET terén.....</u>	<u>92</u>
<u>Nagy Gyula: Nemzetközi kitekintés a felsőoktatási könyvtárak világára: a EUGLOH könyvtári workshopja .....</u>	<u>99</u>
<u>Babocsay Gergely: Az európai természettudományi gyűjtemények digitális integrációja: határ a csillagos ég.....</u>	<u>108</u>
<u>Somorjai Noémi: Egyenlőtlenségek a tudományos kutatás területén. Az amatőr kutatók szerepe .....</u>	<u>114</u>
<u>Molnár Dániel – Dani Erzsébet: Robotok a könyvtárban: Hogyan válhat a robotika a könyvtári mindennapok részévé? .....</u>	<u>122</u>
<u>Horváthné Felföldi Helga: Digitalizáció a szakképzésben. A Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek felülvizsgálata .....</u>	<u>130</u>
<u>Kalcsó Gyula: Ne csak útra csomagoljunk! Miért fontos a csomagolás a digitális megőrzésben? .....</u>	<u>138</u>
<u>Karsa Zoltán István – Szeberényi Imre: A CIRCLE felhő elmúlt évtizede .....</u>	<u>146</u>
<u>Bobák Barbara – Kasza Péter: Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus <i>Rerum Ungaricarum</i> libri kéziratának digitális kiadása (esettanulmány) .....</u>	<u>154</u>
<u>Egyed-Gergely Júlia – Vajda Róza, Gárdos Judit – Horváth Anna – Meiszterics Enikő – Micsik András – Martin Dániel – Marx Attila – Pataki Balázs – Siket Melinda: Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok .....</u>	<u>161</u>
<u>Szemes Botond – Bajzát Tímea – Fellegi Zsófia – Kundráth Péter – Horváth Péter – Indig Balázs – Dióssy Anna – Hegedüs Fanni – Pantyelejev Natali – Sziráki Sarolta – Vida Bence – Kalmár Balázs – Palkó Gábor: Az ELTE Drámakorpuszának létrehozása és lehetőségei.....</u>	<u>170</u>



Sebestyén Ádám: Az ELTEdata szemantikus adatbázis legújabb fejlesztései.....	179
Szlamka Erzsébet: Új trendek a tanulási eredmények tanúsításában .....	185
Tóth Máté – Héjja Balázs: Webshop indítása közkönyvtári környezetben .....	192
Etlinger Mihály – Hernády Judit: A kiadás hagyatéka / a hagyatéka kiadása:	
<i>A Régi Magyar Költők Tárának</i> hálózati kiadásáról.....	199
Varga Emese – Makkai T. Csilla: „Ki a fenének kell collstok?”	
A digitális szöveg rejtett mértékegységei .....	204
Dobás Kata – Fazekas Júlia: ITIdata – Egy irodalmi adatbázis fejlesztése Wikibase	
alapon és ennek hasznosítása Kosztolányi Dezső forrásjegyzékénél .....	211
Sörény Edina: Kézai Simon Program – digitális családi fotóarchívum.....	219
Fülöp Tiffany – Molnár Tamás – Hoczopán Szabolcs: Open Monograph Press	
e-könyvplatform a Szegedi Tudományegyetemen .....	227
Palkó Gábor: Mesterséges intelligencia, digitális bölcsészet, kulturális örökség:	
trendek és eredmények.....	235
Pergéné Szabó Enikő – Bátfai Mária Erika: A tudományos publikálás támogatása	
a Debreceni Egyetemi és Nemzeti Könyvtárban .....	241
Csirmazné Rezi Éva: Nemzetközi kiadványazonosítók és kötelempéldányok kezelése	
az OSZK OKP (Országos Könyvtári Platform) rendszerében .....	250
Alföldi István – Dióssy Anna Laura: Digitálisan született kutatási anyagok megőrzése:	
a relációs adatbázis mint born-digital objektum .....	262
Fekete Norbert: HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős	
szövegtörzsen. A Transkribus alkalmazása az Arany János hivatali iratokon.....	271
Horváth Péter – Kundráth Péter – Palkó Gábor: ELTE Népdalkorpusz – magyar	
népdalok gépileg annotált adatbázisa .....	276
Nagy György: IKT eszközök alkalmazása az alsó tagozatos	
környezetismeret órákon.....	284
Köpösdí Zsuzsa – Molnár Tamás: Multimédiás, interaktív és adaptív tananyagok	
létrehozásának lehetőségei H5P keretrendszerrel .....	289
Jankó Tamás: Munka 4.0 – Ipar 4.0 – Szakképzés 4.0 – :	
A digitális kompetencia jövőbeni fejlesztési útjai .....	296
Békésiné Bognár Noémi Erika – Nagy Andor: Megújuló könyvtári statisztika:	
az egységes adatstruktúra és a korszerű megjelenítés kialakításának útján .....	304
Bolya Mátyás: Kézírtos dallamjegyzések feldolgozása MI-vel támogatott	
digitális környezetben .....	310
Maróthy Szilvia – Seláf Levente – Vigyikán Villó: Régi magyar verskorpusz összeállítása	
stilometriai és számítógépes metrikai kutatásokhoz .....	324
Szűcs Kata Ágnes: Kézírtos források transzformációinak lehetőségei	
a közgyűjteményekben.....	330
Fellegi Zsófia: A digitális filológia infrastruktúrái. A DigiPhil megújulásáról. ....	338
Mihály Eszter: Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása.....	345
Nemeskey Dávid Márk – Palkó Gábor: Szemantikus névelim-azonosítás	
magyar nyelvű szövegeken (a HuWikifier bemutatása) .....	359



## ELŐSZÓ

Tisztelt Olvasó!

Ötödik szerkesztett Networkshop kötetünket az Open Conference Systems felületéről (vagy a REAL-ból, esetleg más repozitóriumból) letöltve elgondolkozhatunk azon, vajon „igazi” konferenciakötet kaptunk-e? Más érzés a nyomdából frissen kikerült kötetet lapozni, beszívni az új könyv illatát. A nyomtatott maradandó, a borítót grafikus tervezi, nem stockfotóból készül. És persze nagy kiadói-nyomdai szakembercsapat munkálkodik az elkészítésén. A hagyományos könyvkészítés feladatkörei mára összecsúsztak, és bizony, a technika (a helyesírás-ellenőrzők, kiadványszerkesztő programok) nem képesek még az íróasztalunkon (vagy az alatt) lévő számítógép dobozába sűríteni egy egész kiadó és nyomda szaktudását.

Nem hisszük, hogy a digitális csak a nyomtatott silány pótléka lehet. A maradandóságról máshogy kell gondoskodnunk, mint azt annak idején a papír kötetekkel tettük. A Creative Commons licenc megteremti a lehetőségét, hogy a digitális kötet több helyen is archiválva lehessen – az MTA Könyvtár és Információs Központ szerverein túl például a szegedi Contenta-ban is. A könyvkészítést tanulni kell – az olvasó e téren is megfigyelhet fejlődést a korábbi kötetekhez képest: mindannyian rutinosabbak lettünk, akik a konferenciakiadvány készítésén dolgoztunk. A cikkek színvonaláról pedig a lektorálás gondoskodik. Minden eddiginél több cikk került a kötetbe – úgy tűnik, a szerzők jobban értékelik ezt a publikálási lehetőséget.

Köszönet illet minden szerzőt, és különösen a szakmai bírálókat, akik névtelen munkája nélkül nem működne a tudományos folyóirat (és tanulmánykötet) kiadás.

Budapest, 2022. november 26.

A szerkesztők



## A nyílt tudomány pénzügyi vonatkozásai

Lencsés Ákos

*Kormányzati Informatikai Fejlesztési Ügynökség (ROR: 01s0v4q65)*

ORCID: [0000-0003-4461-1105](https://orcid.org/0000-0003-4461-1105)

[lencses.akos@kifu.gov.hu](mailto:lencses.akos@kifu.gov.hu)

### Absztrakt

A tudományos világban egyre több szó esik a nyílt tudományról (open science) és annak társadalmi és kutatási hasznosságáról. Magyarországon 2021 októberében a Nemzeti Kutatási Fejlesztési és Innovációs Hivatal (NKFIH) vezetésével 12 hazai szakmai szervezet fogalmazott meg a nyílt tudományra vonatkozó állásfoglalást, ami a téma hazai fontosságát jelzi.

A nyílt tudomány előnyei gyakran megjelennek a szakirodalomban: a kutatási eredmények hozzáférhetősége, ellenőrizhetősége, felhasználhatósága egyértelműen hozzájárulnak a tudomány fejlődéséhez. Ez vizsgálható a letöltések/megtekintések számával vagy a hivatkozások számának növekedésével.

Viszonylag kevesebb szó esik azonban arról, hogy a nyílt tudomány milyen pénzügyi ráfordítást igényel, illetve mekkora megtérülés remélhető a „hagyományos” kutatási módszerekhez képest. Az előadásban az eddigi kutatási eredmények alapján elemezzük az open access publikálás költségeit, az egyedi azonosítók használatának pénzügyi vonatkozásait, a nyílt tudomány egyes infrastrukturális kérdéseire szükséges ráfordításokat.

### Abstract

There is more and more talk in the scientific world about open science and its social and research benefits. In October 2021, 12 Hungarian professional organizations, led by the National Research, Development and Innovation Office (NKFIH), signed a position paper on open science, which indicates the importance of the topic in Hungary.

Benefits of open science are regularly discussed in the literature: the accessibility, reproducibility, and usability of research results clearly contribute to the development of science. This can be easily studied by using the number of downloads/views or the increase in the number of citations.

However, there is relatively less talk about the financial costs of open science and the expected return of investment compared to ‘traditional’ research methods. In the presentation, based on the research results so far, we analyse the costs of open access publishing, the financial implications of the use of persistent identifiers (PIDs), and the expenditures required for certain infrastructural issues in open science.



## Nyílt tudományos irányelvek 2021 második félévében

A nyílt tudomány napjainkra kétségkívül kiemelten fontossá vált a tudománypolitikai gondolkodásban. 2021 második felében számtalan fórumon láttak napvilágot olyan állásfoglalások, szabályozások, iránymutatások, amelyek a kutatókat és a kutatási munkafolyamatokat a nyílt tudomány irányába igyekeznek elmozdítani.

Magyarországon ezek közül az irányelvek közül talán a legfontosabb a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal 2021 októberében közzétett nyílt tudományos állásfoglalása.<sup>1</sup> A dokumentumot kialakító 12 hazai kutatáspolitikai és -támogató szervezet a teljes magyar kutatói szférát lefedi. Ahogyan a dokumentum fogalmaz, a „nyilatkozat célja a nyílt tudományra vonatkozó közös, szakmai konszenzuson alapuló, a tudomány világában jelenleg zajló paradigmaváltásra válaszoló állásfoglalás közzététele [...] A nyilatkozat felhívja a hazai tudományos közösség figyelmét az új szemlélet jelentőségére és időszerűségére, stratégiai kérdéseire, valamint a nyílt tudománynak a nemzetközi együttműködésekben meghatározóbbá váló szerepére.” Az állásfoglalás a nyílt tudomány nyolc pillére mentén fogalmaz meg ajánlásokat: nyílt hozzáférés, kutatási adatok kezelése, tudományos autonómia, új generációs tudományértékelési rendszer, új típusú ösztönzők, nemzetközi együttműködési hálózatok, közösségi tudomány, oktatás és képzés.

Nemzetközi szempontból ebben az időszakban az UNESCO nyílt tudományos irányelveit bemutató, 2021. november 23-án publikált dokumentum a legmeghatározóbb.<sup>2</sup> A kétéves egyeztetéseket követően létrehozott ajánlás részletesen tárgyalja a nyílt tudomány eszköztárát, a nyílt kutatási infrastruktúrát, a nyílt oktatási eszközöket, a társadalmi szereplők bevonását a kutatásba, illetve a kutatási rendszerek közti párbeszéd fontosságát. A dokumentum által megfogalmazott főbb értékek a minőség és integritás, a közösségi hasznosulás, az egyenlőség és igazságosság, illetve a diverzitás és befogadás.

Az Európai Bizottság 2021. november végén publikálta *Towards a reform of the research assessment system* című jelentését.<sup>3</sup> A dokumentum kiemeli a kutatás etikai kérdései és a kutatói szabadság fontosságát. Emellett hangsúlyozza az ellenőrizhetőség, átláthatóság feltételeit is, különös tekintettel a kutatási adatokra. A kutatói értékelés reformjánál fontos szempont a nyílt tudományos kritériumok érvényesítése, illetve a mennyiségi helyett a minőségi feltételek előtérbe helyezése.

A nyílt tudomány egyik leghangsúlyosabb eleme a kutatási adatok kezelése. Emiatt mindenképpen meg kell említeni, hogy Franciaországban 2021 decembere óta rendeleti szinten szabályozzák a kérdést.<sup>4</sup> Az elfogadott jogszabály kimondja, hogy a kutatóintézetek feladata megőrizni a nyers adatokat és hozzáférést biztosítani azokhoz, illetve kötelesek az általuk kezelt adatvagyokra vonatkozóan adatkezelési tervet készíteni. A rendelet támogatja a nyílt hozzáféréssel történő publikálást, illetve bátorítja a negatív eredmények közzétételét.

A döntéshozói szinttel párhuzamosan 2021 decemberében hat jelentős európai kutatóintézet-hálózat jelentetett meg közös nyilatkozatot a nyílt tudományról.<sup>5</sup> A Helmholtz-Gemeinschaft Deutscher Forschungszentren, a Leibniz-Gemeinschaft, a Max-Planck-Gesellschaft, a Consiglio Nazionale delle Ricerche, a Centre National de la Recherche Scientifique és a Consejo Superior de Investigaciones Científicas közös nyilatkozata kiáll

1 NKFIH, „Nemzeti Állásfoglalás a Nyílt Tudományról”.

2 „UNESCO Recommendation on Open Science – UNESCO Digital Library”.

3 „Towards a reform of the research assessment system - Publications Office of the EU”.

4 „Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique – Légifrance”.

5 „G6 Statement on Open Science | CNRS”.

a kutatási gyakorlatok és eredmények szabad megosztása mellett. A dokumentum a nyílt hozzáférés, a FAIR-adatkezelés, a kutatási szoftverek, a kutatásértékelési reform, illetve a képzések és készségek témáiban fogalmaz meg iránymutatásokat.

A döntéshozói és kutatói oldal mellett a társadalom szélesebb rétegei is egyre jobban odafigyelnek a kutatással kapcsolatos elvárásokra. Ebben nagy szerepe lehet annak, hogy a koronavírus-járvány kapcsán több esetben előtérbe került és a napi sajtóban is tárgyalt témává vált az adatokhoz való hozzáférés, az adatok kezelése, a kutatás szervezése, a publikációk nyílt hozzáférése. A 2021-ben publikált *European citizens' knowledge and attitudes towards science and technology* című felmérés szerint az Európai Unió lakosságának 79%-a támogatja azt, hogy az állami forrásból megvalósuló tudományos eredményeknek szabadon elérhetőnek kell lenniük.<sup>6</sup>

Látható tehát, hogy az elmúlt időszakban kiemelten nagy figyelem irányult a kutatási munkafolyamatok átalakítására és a nyílt tudományos irányelvek meghatározására. A számos nyilatkozat, állásfoglalás és iránymutatás azonban műfajukból adódóan csak korlátozottan alkalmasak arra, hogy a gyakorlati teendőket is részletesen bemutassák. Az egyik fontos gyakorlati kérdés a nyílt tudomány pénzügyi vonatkozása. Mindenki számára egyértelmű, hogy a nyílt tudományos követelmények a „hagyományos” kutatói életciklushoz képest további erőforrásokat igényelnek. Ide érthetjük a humán erőforrást (pl. adatkezelési szakemberek képzése, új készségek elsajátítása), az infrastrukturális beruházásokat (pl. repozitóriumok fenntartása, metaadat-silók kialakítása) és további költségeket (pl. open access kiadások).

A nyílt tudomány egyes elemei kapcsán már korábban is napvilágot láttak költségelemzések. Sok esetben azonban a kutatók számára nem egyértelmű, hogyan lehetne a nyílt tudományos kiadásokat megfogalmazni és alátámasztani.

## Nyílt hozzáférés

A nyílt tudomány pénzügyi szempontból leginkább ismert és legtöbbet kutatott területe a nyílt hozzáféréssel kapcsolatos költségek elemzése. Ennek egyik legfőbb oka, hogy a téma átmenetet képez a „hagyományos” és a nyílt tudományos gondolkodás között: a cikkhez kötődő kutatási tevékenység összhangban van a hagyományos kutatási paradigmával. Másik ok, hogy a publikálással járó pénzügyi feladatok közvetlen megjelennek a kutatók tevékenységében: sok esetben a kutató feladata, hogy előteremtse a közzétételi díjat (article processing charge, APC). Harmadik szempontként pedig meg kell említeni, hogy a vizsgálatokat egyszerűsíti, hogy ebben az esetben könnyen meghatározható és ellenőrizhető követelményrendszerrel van szó. Ennek az egyértelműségnek köszönhető az, hogy a nyílt hozzáférésű publikálással kapcsolatos követelmények hangsúlyosan és ellenőrizhetően jelennek meg a pályázati és finanszírozási rendszerben.

A nyílt hozzáféréssel kapcsolatban már egy 2015-es elemzés megállapította, hogy a jelenlegi, folyóirat-előfizetésekre építő rendszer pénzügyi szempontból jelentős többletköltséggel jár a tudományos világ számára.<sup>7</sup> Ha globálisan valamennyi megjelenő tanulmány esetén kifizetnék a kiadók által meghatározott közzétételi díjat, akkor lényegesen kisebb költség jelentkezne, mint az előfizetési díjak esetén. A pénzügyi megtakarítás mellett további előnyt jelentene a kutatások számára, hogy valamennyi tanulmány szabadon elérhetővé válna. Az átállást azonban nehezíti, hogy a globális pénzügyi megtakarításhoz globálisan

---

6 „European citizens' knowledge and attitudes towards science and technology – szeptember 2021 – Eurobarometer survey”.

7 Schimmer, Geschuhn, és Vogler, „Disrupting the Subscription Journals' Business Model for the Necessary Large-Scale Transformation to Open Access”.



összehangolt lépésekre lenne szükség. A Max Planck Digital Library munkatársai interaktív segédletet készítettek a transzformatív open access szerződések pénzügyi előrejelzésére.<sup>8</sup> Ez a segédlet intézményi és kiadói szinten, több forgatókönyv alapján szolgáltat adatokat a döntéshozók számára, hogy milyen kimenettel járhat egy-egy kiadói szerződés átalakítása. Az alkalmazás egyben felhívja a figyelmet arra is, hogy számolni kell a nyílt hozzáféréssel kapcsolatos kiadásokkal; az esetleges, ad hoc megoldások nem oldják meg a rendszerszintű problémákat.

### Kutatási adatok kezelése

A kutatásiadat-kezeléssel járó költségek már kevésbé egyértelműek. A kutatók számára gyakran nem egyénileg, hanem intézményi vagy konzorciumi szinten jelentkeznek a kiadások. A költségek jelentős részét teszi ki a munkaerő-ráfordítás, a hozzáértő szakemberek biztosítása. Több esetben pályázati kötelezettségként is megjelenik az adatkezelési terv elkészítése, illetve a FAIR-adatkezelés. Ezeknek az alapelveknek az ellenőrzése azonban gyakran nem egyértelmű. A kérdés összetettségét mutatja, hogy az OpenAIRE által összeállított *What will it cost to manage and share my data?* című segédlet tucatnyi költségelemet sorol fel az adatkezeléssel kapcsolatban.<sup>9</sup> Az infrastrukturális elemek között megtaláljuk a tárolás, az adatvédelem, a repozitálás, a digitalizálás; a személyi elemek között pedig az adattisztítás, a metaadat-előállítás, az anonimizálás költségeit.

Az adatkezelés egyik ágaként foghatjuk fel az egyedi azonosítók alkalmazását (persistent identifier, PID). Ilyen azonosító lehet a dokumentumok szintjén a DOI, kutatói szinten az ORCID vagy az MTMT-azonosító, intézményi szinten a ROR. Segítségükkel könnyebbé válhat az adathalmazok összekapcsolása, a metaadatok cseréje. A brit JISC 2021 elején vizsgálta 6 egyedi azonosító megtérülési mutatóit (DOI, ORCID, ROR, RAiD, Crossref Grants), és megállapította, hogy ötéves időtávban 5,76 millió GBP (kb. 2,7 milliárd forint) értékű munkaerő-megtakarítás várható csak a DOI és az ORCID rendszerszintű használatától.<sup>10</sup> A lényegi megtakarítás azonban nem innen származik, hanem a megfelelő metaadatok használatának köszönhető hatékonyabb kutatás-fejlesztési döntéshozatali mechanizmusokból. Ennek gazdasági értékét éves szinten 420 millió GBP-re (kb. 195 milliárd forintra) becsülik. Ha a pályázati elbírálások csak 2%-kal válnak hatékonyabbá az egyedi azonosítóknak köszönhetően, az ekkora összeget jelentene az országos rendszer számára. Az állandó azonosítók használatát tehát rendszerszintű fejlesztésnek és befektetésnek fogják fel.

### Kutatásértékelés

A fentiekből is érezhető, hogy pénzügyi szempontból várhatóan a kutatásértékelés jelenti a nyílt tudomány legnagyobb hatású részét. Az eddigi, publikációkra és hivatkozásokra építő rendszer könnyen számszerűsíthető, ellenőrizhető, nyomon követhető. A nyílt tudományos szempontok megfogalmazása a kutatásértékelésben azonban nehezen uniformizálható, nehezen ellenőrizhető. Ezek nélkül viszont csak korlátozottan lehet motiválni a kutatókat a nyílt tudomány felé történő elmozdulásra. Ha az értékelés, a pályázati elbírálás továbbra is elsősorban a publikációk és hivatkozások számát veszi figyelembe, akkor a kutatói erőforrások ezekre összpontosítanak, és csak kisebb hangsúlyt fektetnek a kutatási adatok kezelésére, a nyílt hozzáférésre, a társadalmi hasznosulásra.

<sup>8</sup> Schimmer, Dér, és Campbell, „The DEAL Cost Modeling Tool”.

<sup>9</sup> O’Connor, Delipalta, és Jones, „What will it cost to manage and share my data?”

<sup>10</sup> Brown és mtsai., „UK PID Consortium”.

Ebben az esetben is globális jelenséggel állunk szemben. Ha egy-egy intézmény megváltoztatja a kutatásértékelési módszereit (és például kevesebb publikációt vár el egyéb nyílt tudományos szempontok érvényesítése mellett), akkor az ott dolgozó kutatók rosszabbul teljesítenek a kizárólag publikációra összpontosító országos vagy nemzetközi összehasonlításokban és a pályázatok elnyerésében.

A számtalan kutatásértékelési reformjavaslat közül az egyik szempont a lektorálási tevékenység integrálása az értékelésbe. Aczél Balázs és munkatársai megbecsülték a kutatók által ingyenesen végzett lektorálási tevékenység értékét.<sup>11</sup> Évente globálisan 100-130 millió munkaóráról van szó, ami csak az amerikai, brit és kínai kutatók esetén 2,5 milliárd USD (közel 900 milliárd forint) összeget jelent. Ez a tevékenység azonban semmilyen módon nem épül be a kutatói értékelésbe. Jó irányt jelenthet, ha a lektorálás nyílttá válik. A *Nature* folyóirat 2021-ben lehetőséget biztosított, hogy a cikkek bírálói nyílt módon végezzék a lektorálási folyamatot. Az első év tapasztalatai szerint a kutatók jelentős része támogatja a lektorálás nyitottabbá válását, ami segíthet ezt az elemet beépíteni a kutatásértékelési rendszerekbe.<sup>12</sup>

A fenti példák jól érzékeltetik, hogy a tudományos világban már hozzáférhető a nyílt tudományos átállás egyes elemeivel kapcsolatos költségelemzések. Ezek figyelembe vétele segíthet abban, hogy a megjelenő állásfoglalások és iránymutatások mellett napvilágot lássanak azok a stratégiai dokumentumok is, amelyek már a konkrét költségekre, az elvárásokra és az ezekkel járó gyakorlati lépésekre is hangsúlyt fektetnek. Ez segítheti annak a hangsúlyozását is, hogy a nyílt tudomány nem kizárólag a szabadságról és hozzáférésekről szól. Az adatok kezelése, a publikációs folyamatok átalakítása, a munkafolyamatok gazdagodása a kutatók számára újabb kötelezettségeket és többletfeladatokat jelentenek. Ezek a többletfeladatok kompenzálhatók, ha egyértelművé válik a pénzügyi háttér biztosítása, illetve az új szempontok és kötelezettségek hangsúlyosan beépülnek a kutatásértékelési rendszerbe.

## Irodalomjegyzék

Aczel, Balázs, Barnabas Szaszi, és Alex O. Holcombe. „A Billion-Dollar Donation: Estimating the Cost of Researchers’ Time Spent on Peer Review”. *Research Integrity and Peer Review* 6, sz. 1 (2021. december): 14.

<https://doi.org/10.1186/s41073-021-00118-2>.

Brown, Josh, Phill Jones, Alice Meadows, Fiona Murphy, és Paul Clayton. „UK PID Consortium: Cost-Benefit Analysis”. Zenodo, 2021.

<https://doi.org/10.5281/zenodo.4772627>.

„Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l’intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d’utilité publique ayant pour activité principale la recherche publique - Légifrance”. Elérés 2022. május 27.

<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>.

„European citizens’ knowledge and attitudes towards science and technology – szeptember 2021 – Eurobarometer survey”. Elérés 2022. május 27.

<https://europa.eu/eurobarometer/surveys/detail/2237>.

„G6 Statement on Open Science | CNRS”. Elérés 2022. május 27.

<https://www.cnrs.fr/en/cnrsinfo/g6-statement-open-science>.

---

<sup>11</sup> Aczel, Szaszi, és Holcombe, „A Billion-Dollar Donation”.

<sup>12</sup> „Nature Is Trialling Transparent Peer Review – the Early Results Are Encouraging”.





- „Nature Is Trialling Transparent Peer Review – the Early Results Are Encouraging”. *Nature* 603, sz. 7899 (2022. március 1.): 8–8.  
<https://doi.org/10.1038/d41586-022-00493-w>.
- NKFIH. „Nemzeti Állásfoglalás a Nyílt Tudományról”. Text, 2021. október 15.  
<https://nkfi.gov.hu/hivatalrol/strategia-alkotas/open-science>.
- O’Connor, Ryan, Alexandra Delipalta, és Sarah Jones. „What will it cost to manage and share my data?” 2020. <https://doi.org/10.5281/zenodo.4548344>.
- Schimmer, Ralf, Ádám Dér, és Colleen Campbell. „The DEAL Cost Modeling Tool”, 2021. július 27. <https://doi.org/10.17617/2.3331716>.
- Schimmer, Ralf, Kai Karin Geschuhn, és Andreas Vogler. „Disrupting the Subscription Journals’ Business Model for the Necessary Large-Scale Transformation to Open Access”, 2015. április 28. <https://doi.org/10.17617/1.3>.
- „Towards a reform of the research assessment system - Publications Office of the EU”. Elérés 2022. május 27. <https://op.europa.eu/en/publication-detail/-/publication/36ebb96c-50c5-11ec-91ac-01aa75ed71a1/language-en>.
- „UNESCO Recommendation on Open Science - UNESCO Digital Library”. Elérés 2022. május 27. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.



## Centenáriumi média-adattár és virtuális kiállítás létrehozásának tanulságai az SZTE Klebelsberg Könyvtárban<sup>1</sup>

Farkas Katalin

*Szegedi Tudományegyetem Klebelsberg Kuno Könyvtár (Szeged)*

[Katalin.Farkas@ek.szte.hu](mailto:Katalin.Farkas@ek.szte.hu)

SZTE Klebelsberg Library has been using the OMEKA Classic software and its extensions, which is the basis of the SZTE Klebelsberg Library's Picture Library and Media Library, since 2019. On the occasion of the centenary of the University of Szeged and its library, a centenary selection was created to present the ownership marks, entries and objects in the documents and collections of the library. The repository is hosted on the OMEKA platform, which will also be the subject of a virtual exhibition to be created using this software to promote it to users. The study will follow the workflow and present the issues, opportunities and solutions that have arisen along the way.

**Keywords:** history of the university, history of the library, ownership marks and recommendations, OMEKA Classic software, centenary database, virtual exhibition,

### Bevezetés

Az SZTE Klebelsberg Könyvtár Képtár és Médiatéka alapjául szolgáló OMEKA Classic szoftvert és bővítményeit<sup>2</sup> 2019 óta használja az SZTE Klebelsberg Könyvtár. A Szegedi Tudományegyetem és könyvtárának száz éves fennállása alkalmából a bibliotéka dokumentumaiban található tulajdonjegyek, bejegyzések, illetve a gyűjteményben található tárgyak bemutatására centenáriumi válogatás készült. Az adattár az OMEKA felületén található, melyet szintén e szoftver segítségével létrehozandó virtuális kiállítás igyekszik népszerűsíteni a felhasználók körében. A tanulmány a munkafolyamatot követi végig és a közben felmerülő kérdéseket, lehetőségeket és megoldásokat mutatja be.

### 1. Egyetemtörténet – könyvtártörténet

Az 1872-ben Kolozsváron alapított 1881-től Kolozsvári Magyar Királyi „Ferencz József” Tudományegyetem a trianoni békeszerződés következtében 1921-ben Szegeden telepedett le.<sup>3</sup> Az újonnan épült egyetemi épületek és teljes felszerelésük a menekülés következtében Kolozsváron maradt, ezért Szegeden az újrakezdés a teljes egyetemi infrastruktúra újraépítését és felszerelését jelentette. A kényszerű változás Korb Flóris

---

1 Az adattár és virtuális kiállítás létrehozását dr. Kámán-Gausz Ildikó és Farkas Katalin az SZTE Klebelsberg Könyvtár Különgyűjteményi Osztály könyvtárosai végezték. Dr. Kámán-Gausz Ildikó a tanulmány megjelenése idején az Szegedi Egyetemi Levéltár munkatársa. Munkájukat Nagy Dóra az SZTE Klebelsberg Könyvtár Tartalomszolgáltatási Osztályának csoportvezetője segítette.

2 A szoftverről lásd *Omeke*, hozzáférés: 2022.06.02., <https://omeka.org/>

3 A kolozsvári egyetem elődeiről és az egyetem 1872-1921 közötti történetéről Szögi László-Varga Júlia, *A Szegedi Tudományegyetem és elődei története I. rész A Báthory-egyetemtől a Kolozsvári Tudományegyetemig 1581-1872*. Szeged: Szegedi Egyetemi Kiadó, 2011. és Marjanucz László főszerk., *A Szegedi Tudományegyetem és elődei története II. rész A kolozsvári egyetem története az alapítástól a szegedi újrakezdésig 1872-1921*. Szeged: Szegedi Egyetemi Kiadó, 2022.



tervei alapján Kolozsváron épített<sup>4</sup> és 1909-ben átadott egyetemi könyvtárra<sup>5</sup> is igaz volt, mely teljes dokumentumállományával 1919. május 12-én szintén román állami irányítás alá került.<sup>6</sup> Az 1921. október 9-én ünnepélyesen megnyitott első szegedi tanév egyben a könyvtár működésének szegedi indulását is jelentette, melynek célja már a kezdet kezdetén a felsőfokú oktatás és kutatás szolgálata volt.

## 2. Centenáriumi az SZTE Klebelsberg Könyvtárban

A Szegedi Tudományegyetem a „400 éve az oktatás és tudomány szolgálatában, 100 éve Szeged és a régió fejlődéséért” című 2021/2022-es centenáriumi tanévben számos konferenciával, kiállítással és kiadvánnyal, valamint elektronikus tartalom készítésével emlékezett meg az egyetem szegedi működésének száz éves évfordulójáról, mely keretében a központi ünnepélyek mellett a karok önálló programjai is helyet kaptak. A száz éves egyetemi könyvtár szintén sokszínű, többféle érdeklődőt, felhasználói csoportot megszólító tartalommal készült az évfordulóra. 2021. szeptember elején került sor a *Cimélia* című kötet megjelenésére<sup>7</sup>, 2021. december 9-én a bibliotéka megrendezte a „Száz éves a szegedi egyetemi könyvtár – A szegedi könyvkultúra múltjából” elnevezésű könyvtártörténeti konferenciát<sup>8</sup>, 2022. májusában kiadta az *Egyetem Szegeden, az első 100 év képes krónikája* című visszatekintő albumot, illetve *BiblioTÉKA kiállítótér* elnevezéssel kiállítások otthonául szolgáló emlékfalat hozott létre. A fentiekén kívül szándék mutatkozott arra is, hogy a Klebelsberg Könyvtár dokumentumaiban szereplő különleges, egyetem- és könyvtártörténeti szempontból jelentős ajánlásokra és tulajdonosi bejegyzésekre is ráirányítsa a figyelmet. Az elképzelés megvalósítása – a tartalomra utalva – a *Centenáriumi adattár és virtuális kiállítás* elnevezéssel a bibliotéka tizenegy repozitóriuma közül a *SZTE Klebelsberg Könyvtár Képtár és Médiatékába* került, mely 2019-től az egyetem történeti szempontból fontos fényképek gyűjteménye, 2020-tól a virtuális kiállítások szintén e felületen tekinthetők meg.<sup>9</sup>

## 3. SZTE Klebelsberg Könyvtár Képtár és Médiatéka – adattár és virtuális kiállítás

Az egyetemi könyvtár Contenta adatbázisai közül a *Képtár és Médiatéka* nem EPrints, hanem Omeka szoftver alapú. A szabad felhasználású platform igen népszerű egyetemi és

4 Gheorghe Vais, *Biblioteca Centrală Universitară Cluj 1906–1909*. Cluj Napoca: Editura Alma Mater, 2006.

5 Sipos Gábor, „A kolozsvári egyetemi könyvtár története” in Marjanucz László főszerk. *A Szegedi Tudományegyetem és elődei története II. rész A kolozsvári egyetem története az alapítástól a szegedi újrakezdésig 1872–1921*. Szeged, Szegedi Egyetemi Kiadó, 2022, 587–609., újsághír a könyvtár átadásáról A kolozsvári új egyetemi könyvtár in *Vasárnapi Újság* 56, 22. sz. (1909), 459–460, hozzáférés: 2022.06.02., [http://www.epa.oszk.hu/00000/00030/02894/pdf/VU\\_EPA00030\\_1909\\_22.pdf](http://www.epa.oszk.hu/00000/00030/02894/pdf/VU_EPA00030_1909_22.pdf)

6 Gyalui Farkas, *Emlékirataim 1914–1921*. Kolozsvár: Művelődés Egyesület, 2013.

7 Farkas Katalin–Varga András, *Cimélia, a Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára kincsei = Treasures of the Klebelsberg Kuno Library, University of Szeged*. Szeged: Szegedi Tudományegyetem Klebelsberg Könyvtár, 2021, hozzáférés: 2022.05.30., <http://univ.bibl.u-szeged.hu/id/eprint/59952>

8 A konferencia előadásai 2021. december 21. óta megtekinthetők az SZTE Klebelsberg Könyvtár YouTube csatornáján, hozzáférés: 2022.05.30., <https://www.youtube.com/playlist?list=PLiAvwbbmRQxCzqA5r30VlgFwL1LJXucmc>

9 SZTE Klebelsberg Könyvtár Képtár és Médiatéka, hozzáférés: 2022.05.30., <https://mediateka.ek.szte.hu/>

közgyűjtemények között<sup>10</sup>, a Klebelsberg Könyvtár 2019-től alkalmazza<sup>11</sup>. Az elmúlt három évben 23 389 rekord (minimum 24 000 kép) került betöltésre, valamint nyolc tematikus virtuális kiállítás megtekintése közül választhat az érdeklődő. A kiállítások összes oldalát 2021-ben 23754, 2022.01.01–2022.04.19 között 4305 alkalommal tekintették meg. Az adatok jól mutatják a repozitórium népszerűségét és gyakori használatát.

#### 4.1. Munkafolyamat – feladatmeghatározás és adatgyűjtés

Az egyetem és könyvtára 100 éves fennállása alkalmából az adattár létrehozásával 100 dedikáció, kézjegy, marginália, tulajdonosi jegy, tárgy repozitálása történt meg. Az Omeka szoftver lehetőségeit figyelembe véve olyan környezet kialakítása volt a cél, melyben ezeknek a speciális tartalmak adatai rögzíthetők és megjeleníthetők, továbbá esetleges bővítés alkalmával (például tárgyak, relikviák) szintén használható, valamint koherens a már meglévő adatstruktúrával. Előzetesen meghatározott kritériumok alapján – saját dokumentumokon alapuló, egyetemtörténeti, tudományterületi és időbeli harmonikus eloszlás, a befogadó város, Szeged szerepének bemutatása stb. – történt válogatás eredményeként létrejött névlistát alapul véve a bejegyzést jegyző, illetve tulajdonosi jegy esetén a tulajdonos kiemelése volt a cél.

Az adatgyűjtés kétirányú volt. Egyrészt a már meglévő, könyvtári katalógusban található információ alapján, mely a MARC 21 szabvány esetén az 500-as „Megjegyzés”, valamint az 561-es „Gyűjtemény” mező tartalmán alapult. Katalogizálás alkalmával az 500-as mezőben előre meghatározott forma szerint „X. Y. [például Radnóti Miklós] – (szerző) ajánlásával” felvett adatok találata az ajánlások jegyzékét gazdagította.

The screenshot shows a MARC record viewer window titled '1. megtekintő ablak'. It displays a MARC record with the following fields:

Hosszú	MARC
000	00905nam 2200229 i 4500
001	bibJAT00086240
005	20220318150952.8
008	771130s1933 hu 000 1 hun d
035	\$a(OCoLC)908778369
040	\$aSZ1\$bhun
080	\$a894.511-1Radnóti M.
100	1 \$aRadnóti Miklós
245	10 \$aLábadozó szél :\$bRadnóti Miklós új versei
260	\$aSzeged :\$bDélmagyarország Hírlap- és Ny.,\$c1933
300	\$a56 p. ;\$c21 cm
490	1 \$a(Szegedi Fiatalok Művészeti Kollégiuma ;\$v6.)
500	\$aRadnóti Miklós (szerző) ajánlásával\$S\$Z1 J/21 G 341
561	\$aGáspár Zoltán Gyűjtemény\$S\$Z1 J/21 G 341
583	\$xkr, sbzs, ba, rjg, ob, ak, SZT, kfk
650	7 \$aMagyar irodalom\$xsverses\$2szte
830	0 \$aSzegedi Fiatalok Művészeti Kollégiuma kiadványa
856	41 \$3Full text J/21 VI. a 1364/1\$uhttp://univ.bibl.u-szeged.hu/47669/\$za Szegedi Tudományegyetem hálózatán
949	\$aB 80017\$c64572\$hsz1/R2\$1sz1/R2\$nb 80017\$r380.0\$sNEMK\$tkh\$zJ000132466\$B402\$0011000011
949	\$aXA 33216\$c1089952\$hsz1/R1\$1sz1/R1\$nj/21 VI. a 1364/1\$sNEMK\$tkh\$zJ000549485\$0011000011
949	\$aXA 60701\$c1179395\$hsz1/R1\$1sz1/R1\$nj/21 G 341\$sNEMK\$tkh\$zJ000626849\$B500,501\$0011000011
949	\$aXA 215502\$c238841\$hsz1/R1\$1sz1/R1\$nmagy 6795\$sNEMK\$tkh\$zJ001215568\$0011000011
949	\$aXA 234679\$c532109\$hsz1/R1\$1sz1/R1\$nj/21 VI. a 1364/2\$sNEMK\$tkh\$zJ001229552\$0011000011

1. ábra Radnóti Miklós: Lábadozó szél című kötetének katalógusrekordja

10 Az OMEKA-t alkalmazó intézmények listája, hozzáférés: 2022.05.30., <https://omeka.org/s/directory/>

11 Az OMEKA szoftverről Sirhán Bálint, „Repozitóriumépítés: válasszuk az Omeka open source rendszert!” in *Tudományos és Műszaki Tájékoztatás* 64, 12. sz. (2017), 619–622, hozzáférés: 2022.05.30., <https://tmt.omikk.bme.hu/tmt/article/view/1598/10361>  
szegedi alkalmazásról Nagy Dóra–Nagy Gyula, „Haladó Funkciók és innovatív fejlesztések az EPrints és Omeka szoftverek körében,” in *Workshop 2019*, Budapest: Hungarnet, 2019, hozzáférés: 2022.06.02., DOI: <https://doi.org/10.31915/NWS.2019.16>; Nagy Dóra–Nagy Gyula, „Digitális gyűjteményépítés Omeka alapokon: el lehetséges válasz a könyvtárak (helytörténeti) digitalizált anyagainak szolgáltatására” in *Könyvtári Figyelő* 30, 3. sz. (2020), 456–468, hozzáférés: 2022.06.02., <http://publicatio.bibl.u-szeged.hu/20063/>



Az 561-es mező tartalma a „Gyűjtemény”-re leadott kereséssel érhető el. Például ifj. Horváth István ex libriseit tartalmazó könyvek listáját a tulajdonos nevét megadva lehet megkapni. A korábban még nem katalogizált, könyvtári munkafolyamatok mentén előkerült érdekes ajánlások, bejegyzések Excel táblázata szintén a gyűjtés forrásaként értelmeződött.

#### 4.2. Munkafolyamat – a kiválasztott dokumentumok digitalizálása

A 100 személy nevének véglegesítése után a hozzájuk kötődő dokumentumok autopszia alapján történő kiválasztása is megtörtént, kiválasztott oldalaik digitalizálása Zeuschel OS 12002 szkenneren zajlott, mely eredménye dokumentumonként 3-5 (különleges esetben több), mappába rendezett felvételt jelentett. A digitalizálás feltétele a könyvek (ritkább esetben kéziratok) fizikai állapotának előzetes felmérése volt. A felvételek elkészítése a különleges tartalom miatt állományvédelmi szempontok maximális figyelembevételével történt.

#### 4.3. Munkafolyamat – a gyűjteményfa kialakítása és a metaadatok meghatározása – DC

A felhasználóbarát és áttekinthető keresés segítése érdekében a *Képtár és Médiatéka* felületén gyűjteményfa kialakítására volt szükség. A *Tulajdonosi bejegyzések, dedikációk, egyetemi tárgygyűjtemény* a következő típusokat jeleníti meg: *Ajánlás, Aláírás, Címke, Egyetemi tárgygyűjtemény*. Az utolsó kategória a *Tulajdonosi jegy*, mely alegységei: *Ex libris, Kézírás, Pecsét*.

Ezzel párhuzamosan a két jegyzék közös táblázatba rendezése következett: feladat volt a kigyűjtött adatok összeolvasztása és a táblázat fejlécébe írt metaadatok meghatározása. Mivel az Omeka platform a Dublin Core adatelemeit használja<sup>12</sup>, ezért ezen készlet definícióira kellett megfeleltetni a bejegyzések és tulajdonosi jegyek ismérveit. A tizennyolc ismérvre (Pontos szöveg/Tárgy megnevezése; Kiegészítés a pontos szöveghez; Szerző/Kibocsájtó/Tárgy tulajdonosa; Egyéb név; Testületi név; Létrehozás/bejegyzés dátuma; Bejegyzés típusa; Bejegyzés jellege; Tárgy anyaga; Tárgy mérete; Digitalizálásra került; Cím (befogadó dokumentum; Dokumentum típusa; Dokumentum leltári száma; Link a rekordhoz (Qulto); Kapcsolódó URL; Dokumentumra vonatkozó megjegyzés; Bejegyző neve; Megjegyzés)

A	B	C	F	G	J	L	O	P	Q	R	S
Dublin Core Identifier	Record Type	Item	Dublin Core Title	Dublin Core Language	Dublin Core Country	Dublin Core Rights	Dublin Core Rights	Dublin Core Rights	Dublin Core Rights	Dublin Core Rights	File
2	a_115767	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
3	a_115767_1001	File	a_115767								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
4	a_115767_1002	File	a_115767								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
5	a_27038	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
6	a_27038_1001	File	a_27038								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
7	a_27038_1002	File	a_27038								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
8	a_28992	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
9	a_28992_1001	File	a_28992								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
10	a_28992_1002	File	a_28992								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
11	a_28992_1003	File	a_28992								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
12	a_45565	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
13	a_45565_1001	File	a_45565								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
14	a_45565_1002	File	a_45565								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
15	a_84707	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
16	a_84707_1001	File	a_84707								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
17	a_84707_1002	File	a_84707								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
18	a_84707_1003	File	a_84707								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
19	a_89506	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
20	a_89506_1001	File	a_89506								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
21	a_89506_1002	File	a_89506								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
22	a_90480	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
23	a_90480_1001	File	a_90480								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
24	a_90480_1002	File	a_90480								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
25	b_117214	Item	[feldolgozásra vár]		papír	magyar	A kép a Szegedi Tudományegyetem Szegedi Tudományegyetem Szeged			Tulajdonosi bejegyzések, dedikációk, relikviák	
26	b_117214_1001	File	b_117214								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
27	b_117214_1002	File	b_117214								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
28	b_117214_1003	File	b_117214								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris
29	b_117214_1004	File	b_117214								http://digit.bibl.u-szeged.hu/00400/00499/omeka/kincsek_peccetek_ex_libris

2. ábra CSV importra előkészített táblázat megjelenítése LibreOffice társítással

12 *Dublin Core Metadata Innovation*, hozzáférés: 2022.05.31., <https://www.dublincore.org/specifications/dublin-core/dces/>, az alapelemek magyar fordítása Berke Barnabásné, „A szabványosítással kapcsolatos előzetes munkálatok a NAVA projektben,” in *Könyvtári Figyelő* 49, 1. sz. (2003), 75–88, hozzáférés: 2022.05.31., <https://epa.oszk.hu/00100/00143/00044/berke.html>



A Dublin Core tizenöt adateleme volt megfeleltethető: 1. Cím; 2. Tartalmi leírás; 3. Létrehozó; 4. Közreműködő; 5. Térbeli vonatkozás; 6. Nyelv; 7. Típus; 8. Formátum; 9. Adathordozó; 10. Méret; 11. Forrás; 12. Részegység; 13. Azonosító; 14. Jogok; 15. Jogtulajdonos

#### 4.4. Munkafolyamat – a metaadatok betöltése

Az Omeka szoftver alkalmas nagy tömegű adat CSV importálásra<sup>13</sup>. Ebben az esetben ez úgy könnyítette meg a munkafolyamatot, hogy az elkészült táblázat Omeka-ba való betöltésekor egy-egy rekordhoz több fájl is fel lehetett tölteni, a könyvtár szerverén korábban elhelyezett képek egy link megadásával elérhetővé és betölthetővé váltak. Továbbá, mivel több olyan mező is van, melyek tartalma helyenként (*Adathordozó, Nyelv, Térbeli vonatkozás, illetve Részegység*) vagy teljes egészében azonos (*Jogok, Jogtulajdonos*) az Omeka-ba importálás után a kitöltött és a még kitöltetlen adatelemek „rekordcsontváz”-ként viselkedtek, lerövidítve így segítve a további manuális kiegészítést.

#### 4.5. Munkafolyamat – az adatok kiegészítése

A sikeres CSV import után a még nem meglévő adatelemek kitöltése, a képek sorrendezése történt meg. A könyvtári katalógus és az adattár közötti kapcsolat hivatkozással volt megoldható, így a bejegyzést, tulajdonosi jegyet „hordozó” dokumentumnak csak alapvető könyvészeti adatai kerültek a *Forrás* mezőbe.

The screenshot shows the Omeka admin interface for editing a record. The 'Forrás' field is highlighted with a blue box, containing the following text: "Radnóti Miklós: Lábadozó szél / Szeged, Délmagyarország Hírlap- és Ny., 1933 / leltári szám: J/21 G 341". The interface includes a sidebar with navigation options, a main content area with metadata fields, and a right sidebar with actions like 'Szerkesztés', 'Nyilvános oldal megtekintése', and 'Törlés'.

3. ábra Kész rekord a szerkesztőfelületen „Forrás” mezőben a rekordkapcsolattal

Különleges megoldások is születtek. Például ifj. Horváth István két ex libris birtokosa is volt, erre megoldást a Dublin Core adatelem duplikálási lehetősége nyújt, így mind a két míves tulajdoni jegyet és jellemzőit tartalmazó adatelemet hozzá lehetett rendelni egy rekordban egy személyhez. További érdekességként, megoldandó bibliográfiai feladatként jelentkezett

<sup>13</sup> Nagy Gyula–Nagy Dóra-Sándor Ákos, „Tömeges adatkonverzió és rugalmas export-import lehetőségek az EPrints, OJS és Omeka szoftverek körében,” in *NETWORKSHOP 2019*. Budapest: Hungarnet, 2019. 131–140, hozzáférés: 2022.05.31. DOI: <https://doi.org/10.31915/NWS.2019.17>

amikor egy dokumentum több bejegyzést, tulajdonosi jegyet és ajánlást rejtett. Ebben az esetben a kiemelt személy, vagy testület szempontjából fontosnak talált bejegyzés, tulajdonosi jegy vált hangsúlyossá, de úgy, hogy a leírásban a többi érdekesség is megjelenítésre került.

## 5. Virtuális kiállítás – Könyvtári relikviák – possessor bejegyzések

Az adattár lehetőséget teremtett a bibliográfiai elemek megjelenítéséhez, azonban a bejegyzések, tulajdonosi jegyek és tárgyak a hozzájuk köthető személyek történeteik okán válnak igazán értékké. A professzorok, művészek életútjának, az egyesületek keletkezésének megjelenítésére az Omeka platform virtuális kiállításokat létrehozó bővítménye kínált megoldást. A *Könyvtári relikviák – possessor bejegyzések* című tárlaton<sup>14</sup> az adattár elemeiből válogatva és ahhoz kapcsolódva negyven személyről és egyesületről olvashatók érdekes, egyetemtörténeti szempontból is különleges népszerűsítő, ismeretterjesztő összefoglalások. A kiállítás címdoldalán elhelyezkedő menüben tudományáganként csoportosítva kereshet a felhasználó, külön menüpont szolgál művészek, testületek, Szeged kultúrtörténeti érdekességei és neves személyek relikviái, bejegyzései megjelenítéséhez. Bár az Omeka platform kiállításrendező bővítménye lehetőséget nyújt sokféle grafikai elem alkalmazására és egyéb (hanganyag stb.) beágyazására, ebben az esetben a bejegyzésekről, tulajdonjegyekről, tárgyakról készített képek és a hozzájuk írt egyetem- és könyvtártörténeti érdekességek díszítetlenségükkel kerülnek a megjelenítés középpontjába.



4. ábra A Könyvtári relikviák – possessor bejegyzések című virtuális kiállítás nyitó oldalának részlete

<sup>14</sup> Könyvtári relikviák – possessor bejegyzések, SZTE Klebelsberg Könyvtár Képtár és Médiatéka, hozzáférés: 2022. 06. 01., [https://mediateka.ek.szte.hu/exhibits/show/relikviak\\_possessorok/koszonto](https://mediateka.ek.szte.hu/exhibits/show/relikviak_possessorok/koszonto)

## 6. Összegzés

A más feladatok mellett végzett több hónapon át tartó munkafolyamat több tanulsággal is szolgált. Egyrészt lehetőséget adott az Omeka alaposabb megismerésére és tanulmányozására. Minden kreatív megoldás, új feladat egyúttal a meglévő rendszer tökéletesítésére és javítására is szolgál – ez esetben a Dublin Core mezők sorrendjének megváltoztatására, vagy az adatgyűjtés folyamán a könyvtár Corvina katalógusában található esetleges hibás rekordok javítására. A munkában résztvevő könyvtárosok a felmerülő informatikai kérdések megoldásán kívül egyetem- és könyvtártörténeti ismereteik gyarapításával és tudásmegosztással fejlesztették a kompetenciáikat.

## A nyílt archívumi információs rendszer (OAIS) szabványának honosítása

Bódog András

Országos Széchényi Könyvtár, Könyvtári Intézet, Könyvtári Szabványosítási Iroda  
[bodog.andras@oszk.hu](mailto:bodog.andras@oszk.hu)

### Absztrakt

One of the most important contemporary challenges facing the content industry and culture sector is long-term preservation. In recent years, there has been virtually no conference in this sector where this topic has not been discussed in some form. A frequent recurring feature of these presentations is the reference to the OAIS standard. This paper briefly introduce this standard, explaining the process and importance of its Hungarian implementation/adoption.

The reference model for an Open Archival Information System (OAIS) was originally developed for space research purposes at the turn of the millennium in an international cooperation (CCSDS) led by NASA, but within a few years it was also adopted by institutions concerned with the long-term preservation of cultural heritage, such as archives and libraries. Because of this spread, the International Organization for Standardization (ISO) has adopted CCSDS 650.0-M-2 standard without change, which now conforms to ISO 14721:2012, and the Hungarian version of which (with the support of the National Széchényi Library) was published in spring 2022 as MSZ ISO 14721:2022.

As a reference model, OAIS addresses theoretical issues of long-term preservation. While it does not lay down specific implementation guidelines, it is the model on which most software and institutional practices in long-term preservation are based, including those of the Hungarian National Archives and the National Széchényi Library's Digitization Centre and Web Archive. By implementing/adopting the international standard, we hope to take a first step towards standardizing the terminology in the field in Hungary, and to offer an opportunity for institutions not yet using OAIS to learn about the reference model and implement it in the future.

### 1. Bevezető

A tartalom- és kulturális ipar egyik legfontosabb kortárs kihívása a hosszú távú megőrzés. Az elmúlt években gyakorlatilag nem beszélhetünk a kulturális szférában a kulturális javak megőrzése területén olyan konferenciáról vagy szakmai eseményről, amely során ne került volna elő a címben szereplő szabvány. A 2022. évi Networkshop konferencián tartott előadásom jelen összefoglalója kitér a hosszú távú megőrzés fontosságára, röviden bemutatja az OAIS-szabványt, illetve ismerteti a szabvány honosításának hátterét. Az Országos Széchényi Könyvtár (OSZK) Könyvtári Intézetben belül működő Könyvtári Szabványosítási Iroda (KÖSZI) missziója többek között a hosszú távú megőrzést célzó szabályozási/szabványosítási tevékenységek támogatása, elősegítése.<sup>1</sup>

1 Könyvtári Intézet honlapja. Könyvtári Szabványosítási Iroda. <https://ki.oszk.hu/szabvanyositas/konyvtari-szabvanyositasi-iroda> (2022.06.02.)



## 2. A hosszú távú megőrzés relevanciája

A hosszú távú megőrzést az OAIS-szabvány a következőképp definiálja: „Az információnak a célközönség által függetlenül is értelmezhető formában való fenntartása hosszú távon, a hitelességet alátámasztó bizonyítékkal.”<sup>2</sup> A hosszú távú megőrzés az archiválandó hiteles és eredeti információ, kultúrkinccs – vagy bármi egyéb, amit meg kívánunk őrizni az utókor számára – meg nem határozott időtartamú időtávra történő megőrzését és szolgáltatását jelenti. Ez a feladatkör nem új keletű, pontosan ez a mindenkori könyvtár, levéltár, irattár, múzeum és egyéb emlékezetmegőrző intézmény egyik elsődleges feladata.

A magyar nemzeti könyvtár esetében ennek jogszabályi háttere többszintű: egyrészt európai uniós viszonylatban megemlíthető *A Bizottság ajánlása a kulturális anyagok digitalizálásáról és online hozzáférhetőségéről, valamint a digitális megőrzésről (2011/711/EU)* hosszú távú megőrzésre irányuló 8. pontja,<sup>3</sup> másrészt az 1997. évi CXL. törvény a muzeális intézményekről, a nyilvános könyvtári ellátásról és a közművelődésről rendelkezik arról, hogy a nemzeti könyvtár alapfeladata „gyűjteményének archiválása és védelme”.<sup>4</sup> A kötelempéldányokra érvényes archiválási köteleességet előírja a 717/2020. (XII. 30.) kormányrendelet,<sup>5</sup> a 22/2005. (VII. 18.) NKÖM rendelet,<sup>6</sup> illetve a 2010. évi CLXXXV. törvény értelmében a „megőrzési célú kötelempéldányt a közgyűjtemény nyilvántartásából csak akkor lehet törölni, ha megsemmisült vagy javíthatatlanul megrongálódott.”<sup>7</sup> A 30/2014. (IV. 10.) EMMI rendelet előírja a nemzeti könyvtár számára a hozzáférhetőség széles körű biztosítását, az állomány, a hosszú távú megőrzés érdekében, digitalizálási terv mentén végzett digitalizálását, illetve mindezek infrastrukturális-műszaki hátterének biztosítását.<sup>89</sup>

Egészen a közelmúltig a hosszú távú megőrzés adathordozója a papír volt és marad is még nagyon sokáig. Az ókori könyvtárak kora óta eltelt több mint ezer esztendő, illetve a mai könyvtárak, levéltárak évszázadokkal ezelőtt alapított képviselői és jogelődjei okán tudjuk jól, hogy a papír – a megfelelő tárolási körülmények biztosításával – hosszú életű adathordozónak bizonyul. Mi a helyzet azonban az azóta feltalált adathordozókkal? Jelentős részük 20. századi találmány, gondoljunk csak az orsós magnószalagra, a mikrofilmre, az audio- és videokazettákra, hajlékonylemezekre, a CD-re, a DVD-re, majd a Blue Ray optikai lemezekre, sőt manapság a szerverparkokra és egyéb háttértárakra. Ezen adathordozók közül nem egynek megértük néhány évtized alatt a bevezetését, felfutását, elterjedését,

2 MSZ ISO 14721:2022 Úradat- és információközvetítő rendszerek. Nyílt archívumi információs rendszer (OAIS). Referenciamodell. Budapest : MSZT, 2022. p. 26.

3 *A Bizottság ajánlása a kulturális anyagok digitalizálásáról és online hozzáférhetőségéről, valamint a digitális megőrzésről*  
<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:283:0039:0045:HU:PDF>  
(2022.06.02.)

4 1997. évi CXL. törvény a muzeális intézményekről, a nyilvános könyvtári ellátásról és a közművelődésről. 61.§(4)e <https://net.jogtar.hu/jogszabaly?docid=99700140.tv> (2022.06.02.)

5 717/2020. (XII. 30.) Korm. rendelet a kiadványok kötelempéldányainak szolgáltatásáról, megőrzéséről és használatáról <https://net.jogtar.hu/jogszabaly?docid=a2000717.kor> (2022.06.02.)

6 22/2005. (VII. 18.) NKÖM rendelet a muzeális könyvtári dokumentumok kezelésével és nyilvántartásával kapcsolatos szabályokról. 1.§ (2) f), 3.§ (1), 6.§ (1) <https://net.jogtar.hu/jogszabaly?docid=a0500022.nkm>  
(2022.06.02.)

7 2010. évi CLXXXV. törvény a médiaszolgáltatásokról és a tömegkommunikációról. 46.§ (14) <https://net.jogtar.hu/jogszabaly?docid=a1000185.tv> (2022.06.02.)

8 30/2014. (IV. 10.) EMMI rendelet az országos múzeum, az országos szakmúzeum, a nemzeti könyvtár, az országos szakkönyvtár és az állami egyetem könyvtárának kiemelt feladatairól. 8.§ (1), (2) a) <https://net.jogtar.hu/jogszabaly?docid=a1400030.emm> (2022.06.02.)

9 A nemzeti könyvtárra vonatkozó hosszú távú megőrzés jogszabályi hátterének eredeti összefoglalójának forrása: Dancs Szabolcs: A hosszú távú megőrzés szabványos útja (és a Rosetta rendszer példája). = *Tudományos és Műszaki Tájékoztatás*. 63. évf. (2016.) 1. sz. p. 27-28.



majd elavulását és partvonalra kerülését. A fizikai adathordozók azonban csak a mérleg egyik fele, jelentős kihívás még a hosszú távú megőrzés terén az azokat működtető rendszerek meglétének esetleges hiánya. Hiába tárol egy archívum például floppylemezekon évtizedek óta gondosan valamilyen információt, problematikus, ha nincs már olyan számítógépe, amely rendelkezne hajlékonylemez-meghajtóval, sőt már az optikai meghajtó is egyre ritkább. Minél régebbi típusú adathordozón van az információ, annál nehezebb működő (karbantartható) lejátszó- vagy olvasóeszközt találni hozzá. És akkor még nem is beszéltünk a fájlformátumok és az operációs rendszerekkel történő szoftverkompatibilitás kérdésköréről, nem is említve az úgynevezett *born digital*, tehát az eleve digitális környezetben létrejövő dokumentumok megőrzésének, sőt a webarchiválásnak a kérdéseit. Ezek a formátumok és dokumentumtípusok képezik a papírformátum melletti új dimenziót a hosszú távú megőrzés feladatában. Ezek utókornak átörökítésével ugyanúgy foglalkozniuk kell az emlékezetmegőrző szervezeteknek, mint a papírformátumú dokumentumokéval. A kiválasztott, többnyire elektronikus dokumentumok hosszú távú megőrzése nem merül ki a fájlok háttértárra pakolásában. Az információt bitenként kell megőrizni, illetve valamilyen formában fent kell tartani azt az eredeti működési környezetet, amelyben anno az az információ olvasható, értelmezhető volt. Hogy ez többek között emulációval, a régi gépek üzemeltetésével, vagy folyamatos formátumról-formátumra tartó migrációval valósítható meg, az már a gyakorlati megvalósítás kérdéskörébe esik, és nagyban függ az adott archívum preferenciáitól. Érdekes szempont még a megőrizendő dokumentum értelmezési kontextusának megőrzése: adva lehet egy már ma is történelmi jelentőségű dokumentum, amit megőrizhetünk akár digitális formában, akár valamilyen igen tartós fizikai adathordozóra rögzített formában is. Igen ám, viszont a műszaki jellegű megőrzés/tárolás mellett érdemes feltüntetni olyan információkat, mint például: milyen nyelvű az adott dokumentum (ki tudja, hátha holt nyelv lesz az a bizonyos nyelv egy vagy több évszázad múlva), mi célból készült, kik készítették, milyen technológiával és utóbbi hogyan használható.<sup>10</sup> A fizikai megőrzésen túlmutatóan tehát az eredeti működési állapot és a kontextus megőrzése tekinthető a hosszú távú megőrzés legfőbb kihívásának. A ma általánosan e cél megoldására szolgáló koncepció azonban nem a kulturális szférában fogant meg, hanem az űrkutatásban.

### 3. Az OAIS-referenciamodell

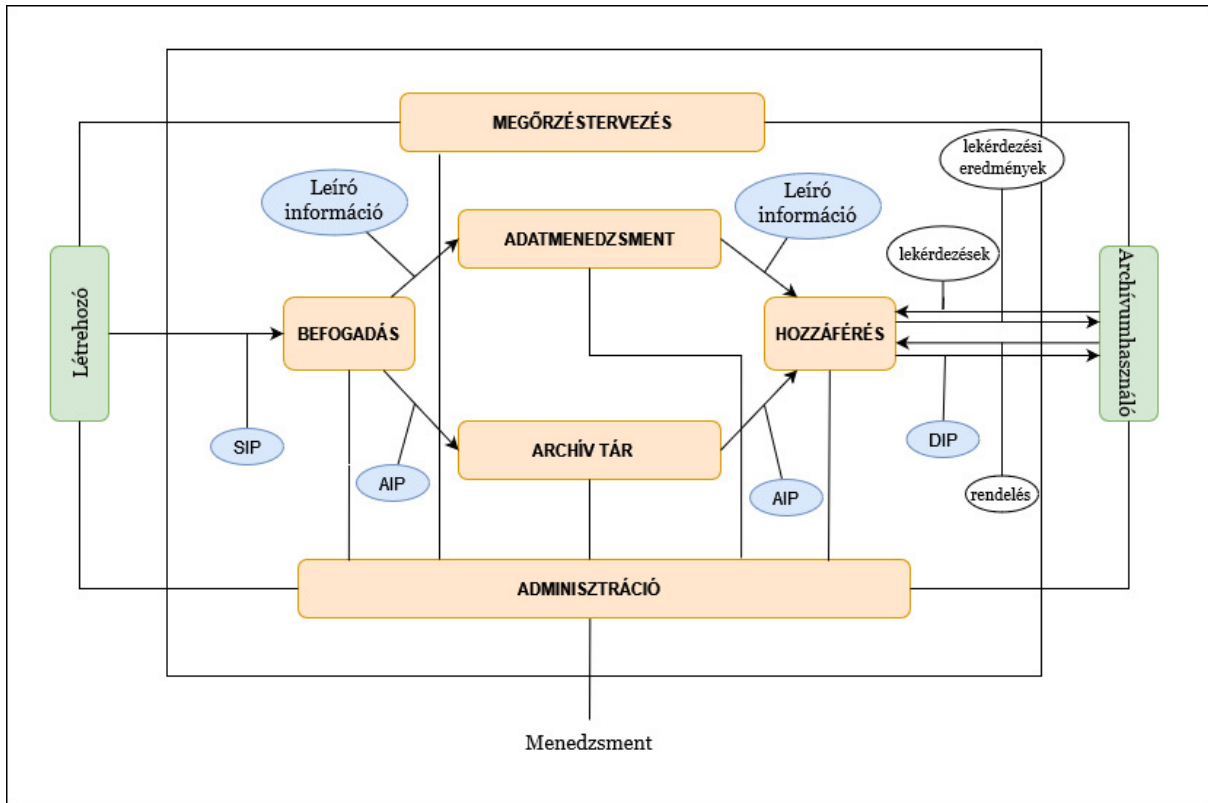
Amint véget ért a hidegháborús űrverseny, a hangsúly fokozatosan áttelődött az egész bolygó tudományos fejlődését szolgáló űrkutatásra, és bizony igény mutatkozott arra, hogy az űrmissziók során begyűjtött fizikai és elektronikus adatokat valamilyen szabványos formában őrizzék meg a tudomány és az utókor számára. Szükség volt egy időtálló fogalmi modell létrehozására, amely az aktuális technológia korlátozta implementációktól függetlenül ad útmutatást a hosszú távú megőrzés elméletéhez. A NASA vezetésével egy 1997-es fehér könyvben dolgozták ki a nyílt archívumi információs rendszer referenciamodelljének első változatát,<sup>11</sup> majd az űrkutatás nemzetközi szabványosítási kérdéseivel foglalkozó Űradatrendszerek Tanácsadói Bizottsága (CCSDS) égisze alatt jelent meg az első OAIS-standard 2002-ben.<sup>12</sup> Ekkorra ez a konceptuális modell már messze túlnőtt az űrkutatás területén, és már az ezredforduló környékén számos nemzeti könyvtár (köztük például a holland nemzeti könyvtár digitális könyvtára, az USA nemzeti levéltára) foglalkozott a

10 MSZ ISO 14721:2022 p. 85-95.

11 Reference Model for an Open Archival Information System – CCSDS 650.0-W-1 White Book, April 10, 1997. <https://ntrs.nasa.gov/api/citations/19980211383/downloads/19980211383.pdf>

12 Reference Model for an Open Archival Information System – CCSDS 650.0-B-1 January, 2002. <https://public.ccsds.org/Pubs/650x0b1s.pdf>

modellel, mint a hosszú távú megőrzés problémájának potenciális megoldásával.<sup>13</sup> A széles körben mutatkozó igényekre reflektálva a Nemzetközi Szabványügyi Szervezet (ISO) ISO 14721 hivatkozási szám alatt változtatás nélkül, nemzetközi szabványként bevezette a CCSDS standardját. Amikor ők 2012-ben felülvizsgálták a modellt, az ISO is követte, így jutottunk el a jelenleg is hatályos ISO 14721:2012 szabványig,<sup>14</sup> amelynek magyar változata idén márciusban jelent meg.



1. ábra: OAIS funkcionális entitások  
(Ilácsa Szabina és Bódog András (KöSZI) ábrája az  
MSZ ISO 14721:2022 szabvány (p. 42.) alapján)

Az OAIS-ban az információkat több fájlból álló úgynevezett információs csomagokban tárolják. Ezekben a megőrizendő tartalmat a reprezentációs információ egészíti ki, biztosítva mindazon plusz (például a dokumentum kontextusával kapcsolatos) információt, amelyet a hosszú távú megőrzés követelményeinél már említettem. Például, ha egy CD-ROM tartalmát archiválják, nemcsak azt mellékelik reprezentációs információként, hogy tulajdonképpen mi is az a CD-ROM, hanem a szabványos műszaki specifikációt is megnevezik, jelen esetben az azt leíró szabványt. Hasonló reprezentációs információ a megőrzendő tartalom nyelve is, amely mellé adott esetben akár szótárat is csatolhatnak. Az OAIS három különböző típusú információs csomagban őrzi meg az információt: az átadási információs csomagot (SIP) a létrehozó adja át az archívumnak, amely azt leíró információkkal gazdagítja, majd egy vagy több megőrzési információs csomag (AIP) formájában kerül hosszú távú tárolásra. Az AIP-okból lehet előállítani a felhasználói igény szerint szolgáltatott disszeminációs információs

<sup>13</sup> Alan Wood, Don Sawyer. Reference Model for an Open Archival Information Systems (OAIS): Overview and Current Status – NASA (2001). 16-June-2001. American Library Association Annual Conference <https://www.slideshare.net/fafllrt/reference-model-for-an-open-archival-information-systems-oais-overview-and-current-status-nasa-2001-presentation-573443>

<sup>14</sup> ISO weboldala. ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference model. Abstract <https://www.iso.org/standard/57284.html> (2022.06.02.)



csomagokat (DIP).<sup>15</sup> Tömören ez tekinthető az OAIS-modell konceptuális vázának, a szabvány ennek az elképzelésnek, keretrendszernek az információs és funkcionális modelljét részletezi.

#### 4. Az OAIS-szabvány honosítása, a magyar nyelvű változat jellemzői

Mivel az Országos Széchényi Könyvtár egyik missziója a hosszú távú megőrzés, és a Könyvtári Szabványosítási Iroda feladata az OSZK és a hazai könyvtárszakma szabványügyi támogatása, ezért az Országos Könyvtári Szabványosítási Bizottság feladatául tűzte ki a nemzeti könyvtár küldetésének támogatása okán e fontos szabvány magyar nyelvre történő átültetését, különösen, hogy a V4 országok közül Csehország és Szlovákia is saját ez irányú projektjüket az ISO 14721 szabvány honosításával kezdték.<sup>16</sup> A másik indok, hogy a magyar könyvtári szabványügy már régóta adós ezzel, hiszen az OAIS már évek óta széles körben használt szabvány – hazánkban elsőként a Magyar Nemzeti Levéltár valósította meg a teljes körű implementációját, amikor e modellre alapozta az elektronikus levéltár<sup>17</sup> hátterét.<sup>18</sup> Az OAIS-t egyre többen használják tehát szabványként Magyarországon, azonban nem rendelkezik a terület egységes magyar szakszókincssel – jobb híján a kollégák saját igényeikhez szabványosítva magyarították –, ezért a honosítással lehetőség kínálkozik egy szakterületről és szakmától független egységes terminológia megalapozására. Sokszor hallottuk már a „SIP-csomag”, „DIP-csomag” kifejezéseket a különböző szakmai rendezvényeken, így a honosítási folyamat részét képező szabványtárgyalás (a szabvány fordítását és fordításellenőrzését követően a szabványjavaslat szakértők általi lektorálása<sup>19</sup>) folyamán úgy véltük, hogy ezt a szóhasználatot tükrözve célszerű lenne ezeket a betűszavakat magyaros kiejtéssel honosítani. Így lett a modell nevének kiejtése is, az ISBN és az ISSN példáját követve, OAIS (ejtsd: fonetikusán oais). Reményeink szerint a magyar nyelvű szabvány egyfajta hivatkozási alapként is szolgál majd bárki számára, aki saját hosszú távú megőrzési projektbe kezdene, vagy érdeklődne a téma iránt.

A 2022. március 1-én megjelent MSZ ISO 14721:2022 szabványt a Magyar Szabványügyi Testület *MB 508 Információ és dokumentáció* nemzeti szabványosító bizottsága<sup>20</sup> honosította.<sup>21</sup> A szabványtárgyaláson a KöSZI munkatársai mellett az OSZK, az (akkor még a Petőfi Irodalmi Múzeumhoz tartozó) Digitális Bölcsészeti Központ, illetve a Magyar Nemzeti Levéltár szakemberei vettek részt. A magyar nyelvű kiadás elkészítését a Széchényi Könyvtár támogatta anyagilag. Mivel az angol nyelvű eredetit változtatás nélkül vette át az ISO, azt forgatva bizony kitűnik az eredeti úrkutatási célú felhasználás. Talán ezért nem épp a legkönnyebb a szabvány szövege, ráadásul a szabványtárgyalás során – ahol gyakorlatilag szóról-szóra végig vitattuk a nyers fordítást az eredeti tükrében – számos kisebb ellentmondást is felfedeztünk. A CCSDS eredeti kiadásában többek között az alkalmazott terminológia sem volt teljesen egységes, ezért a magyar nyelvű szabvány előszavában ezen ellentmondások felsorolása

15 MSZ ISO 14721:2022 p. 41-43.

16 DANCS Szabolcs: A V4 országok könyvtárosainak válaszai a digitális kor kihívásaira = *Könyvtári Figyelő* 62. évf. (2016) 4. sz. p. 528.

17 Elektronikus Levéltári Portál <https://www.eleveltar.hu/kereses> (2022.06.02.)

18 Szatucsek Zoltán: Az elektronikus iratok kezelése. Budapest : Nemzeti Közszerzői Egyetem, 2017. p. 25-26. <https://kti.uni-nke.hu/document/vtkk-uni-nke-hu/Az%20elektronikus%20iratok%20kezel%C3%A9se.pdf>

19 Magyar Szabványügyi Testület honlapja. A szabványkidolgozás folyamata <https://prod.mszt.hu/hu-hu/szabvanyositas> (2022.06.02.)

20 Magyar Szabványügyi Testület honlapja. *MSZT/MB 508 Információ és dokumentáció* <https://ugyintezes.mszt.hu/MBs/Details/294> (2022.06.02.)

21 Magyar Szabványügyi Testület honlapja. *MSZ ISO 14721:2022* <https://ugyintezes.mszt.hu/Publications/Details/178190> (2022.06.02.)

mellett az MSZT értelmezési segédlettel szolgál a szabványalkalmazás megkönnyítésére. A szövegtestben a referenciamodell „építőköveit” konzisztensen eltérő formázással jelöltük, így a **szakkifejezések** mindig félkövérek, a FUNKCIONÁLIS ENTITÁSOK pedig csupa nagybetűsek. További segédlet az OAIS által meghatározott szakkifejezések magyar betűrendes mutatója, ugyanis a szabványhonosításkor a definícióknál meg kellett tartani az angol betűrendet. A szabvány – az összes többi MSZ-szabványhoz hasonlóan – elérhető az MSZT szabványboltjában,<sup>22</sup> illetve a mintapéldány kölcsönözhető is az OSZK-n belül működő Könyvtártudományi Szakkönyvtár gyűjteményéből.<sup>23</sup>

## 5. A Könyvtári Szabványosítási Iroda tervei e területen

A szabvány megjelenését követően a KöSZI részéről az egyik fő tevékenység e területen a nemzeti könyvtáron belül a hosszú távú megőrzéssel foglalkozó szervezeti egységekkel (Gyűjteménymegőrzési Főosztály és Digitális Bölcsészeti Központ) történő szabványügyi együttműködés. A közelmúltban például egy implementációs kérdés szabványosságát vizsgáló szabványértelmezéssel segítettük a kollégák munkáját. Tágabb értelemben a hazai könyvtárszakma támogatása is feladatunk a könyvtári munkát érintő szabványokkal (köztük az OAIS-sal is) kapcsolatos ajánlások és segédletek készítésével és közreadásával, OSZK-s keretekben (általában online, például honlapon), illetve szakmai folyóiratokban megjelenő publikációk formájában.

### Felhasznált források:

Alan Wood, Don Sawyer. Reference Model for an Open Archival Information Systems (OAIS): Overview and Current Status – NASA (2001). 16-June-2001. American Library Association Annual Conference <https://www.slideshare.net/faflrt/reference-model-for-an-open-archival-information-systems-oais-overview-and-current-status-nasa-2001-presentation-573443>

Dancs Szabolcs: A hosszú távú megőrzés szabványos útja (és a Rosetta rendszer példája). = Tudományos és Műszaki Tájékoztatás. 63. évf. (2016.) 1. sz. p. 27-37.

Dancs Szabolcs: A V4 országok könyvtárosainak válaszai a digitális kor kihívásaira = Könyvtári Figyelő 62. évf. (2016) 4. sz. p. 523-535.

Magyar Szabványügyi Testület honlapja. A szabványkidolgozás folyamata <https://prod.mszt.hu/hu-hu/szabvanyositas> (2022.06.02.)

Magyar Szabványügyi Testület honlapja. MSZT/MB 508 Információ és dokumentáció <https://ugyintezes.mszt.hu/MBs/Details/294> (2022.06.02.)

MSZ ISO 14721:2022 Úradat- és információközvetítő rendszerek. Nyílt archívumi információs rendszer (OAIS). Referenciamodell <https://link.oszk.hu/libriurl.php?LN=hu&DB=KSZK&SRV=an&SRE=000003860360>

Reference Model for an Open Archival Information System – CCSDS 650.0-W-1 White Book, April 10, 1997. <https://ntrs.nasa.gov/api/citations/19980211383/downloads/19980211383.pdf>

Reference Model for an Open Archival Information System – CCSDS 650.0-B-1 January, 2002. <https://public.ccsds.org/Pubs/650x0b1s.pdf>

---

<sup>22</sup> <http://www.mszt.hu/web/guest/webaruhaz-start>

<sup>23</sup> OSZK Nektár katalógusrekord: <https://link.oszk.hu/libriurl.php?LN=hu&DB=KSZK&SRV=an&SRE=000003860360>.



Reference Model for an Open Archival Information System – Recommended Practice  
CCSDS 650.0-M-2 Magenta Book June, 2012. <https://public.ccsds.org/pubs/650x0m2.pdf>

Szatucsek Zoltán: Az elektronikus iratok kezelése. Budapest : Nemzeti Közszolgálati Egyetem, 2017. 30 p. <https://kti.uni-nke.hu/document/vtkk-uni-nke-hu/Az%20elektronikus%20iratok%20kezel%C3%A9se.pdf>



## Oktatást segítő gamifikációs alkalmazások, mint szakdolgozati témák

Perlaki Attila

Tokaj-Hegyalja Egyetem, Comenius Intézet, Természettudományi tanszék  
[perlaki.attila@unithe.hu](mailto:perlaki.attila@unithe.hu)

### Abstract

Gamification applications for helping the education as topics of diploma work  
There are disagreements regarding the expectations and possibilities of gamification in education. The main issue is the lack of qualified human resource familiar with the topic, the other hurdles include non-ideal software tools, as well as skepticism regarding the merits of this method. As a consultant of students in information technology, I have attempted to purpose subjects in development of educational applications. Knowing the limitations of inexperienced (student) programmers, we are searching for novel ways to utilise easy-to-use software tools in a successful manner. One of the best choices is a game engine: students can focus on the desired outcome, without getting bogged down by low-level programming details. Once the method is successfully demonstrated with students, it can also be expanded to include others who are not professional programmers, for example educators, who can apply the same principles in teaching their own subjects.

**Keywords:** gamification, education, software development, game engines

Ellentmondásosak a gamifikációval kapcsolatos elvárások és lehetőségek az oktatásban. Leginkább a területen jártas szakemberek hiánya a probléma, valamint a nem ideális szoftver-eszközök és a téma iránti bizalmatlanság is hátráltató tényező. Informatikai szakdolgozatok témakiírójaként és konzulenseként szándékomban áll e területen is előrelépni. A hallgatók gyakorlatlanságát is figyelembe véve olyan eszközöket keresünk, amelyekkel a sikeres diplomamunka elérhető. Az egyik ilyen eszköz a game engine alkalmazása: a hallgató a kitűzött oktatási célra összpontosíthat, anélkül, hogy az alacsony szintű, rutin jellegű programozási részletekkel ideje és energiája pazarlódná. Ha ez sikerül, mások, a hallgatókhoz hasonlóan nem hivatásos programozók, például tanárok is a saját tárgyaikhoz alkalmazhatják ezt a megközelítést.

**Kulcsszavak:** gamifikáció, oktatás, szoftverfejlesztés, game engine-ek

### 1. Gamifikáció és oktatás

Kézenfekvő és veszélyes, talán ez a két ellentmondó érzése lehet a legtöbb oktatással foglalkozónak, ha a *gamification*, azaz a játékoság szóba kerül. Előítéletek és valós problémák, idő-, energia- és eszközhiány egyaránt tartózkodóvá teheti a területen tevékenykedőket. Ezekből talán az eszközhiány a legkönnyebben megoldható, de itt is nagy a veszélye a rosszul választott, alacsony hatékonysággal, vagy egyáltalán célszerűen az oktatásban alkalmazni nem lehetséges eszközök beszerzésének.

Komolyvisszatartóerő lehet a *hype* is, azaz a túlzott elvárások támasztása, a túlzott "divatosság". Felmerülhet, hogy "az oktatás nem játék", de már rég tudjuk, hogy a poroszos módszereken



túllépett az idő. A világjárvány hirtelen alkalmazkodásra kényszerített az oktatásban is... és néhány területen talán megmaradhat a lendület. [1]

A gamifikáció közel sem bejáratott eszköztár: módszerei, eszközei nagyon gyorsan változnak, sok időt, kísérletezést kíván. Tudható, hogy akadnak nehézségek az egyébként is igen feszes tantervi elvárások mellett/felett ezzel foglalkozni, ezt beilleszteni – ha egyáltalán lehetséges. A siker pedig nem feltétlenül garantált, legalább is a mérhetőség oldaláról. A fiatal korosztály ugyanis természetéből adódóan nyitott a játékosagra, de ennek a legkönnyebben mérhető vektora a motiváltság. Hogy az egyéb mérőszámok miként javíthatóak, az még sokaknak sok munkáját igényli. [2, 5]

Ebben a cikkben egyetlen kis részletről esik szó: a felsőoktatásban *programtervező informatikus* szakon szakdolgozatot készítő hallgatók bekapcsolhatóságát olyan gamifikációs oktatási alkalmazások – pontosabban ezek első, vagy demó verzióinak – elkészítésébe, amelyek legalább a kísérletezés szintjén tanulságokkal szolgálhatnak. A kihívás kellően komplex, nem csak öncélú kódgyártás: az informatikusnak a *megrendelő*, a *felhasználó* igényeit is jól fel kell mérnie, azaz kellő jártasságot kell szereznie ahhoz, hogy a párbeszéd eredményes legyen a felek közt.

A közoktatásban dolgozók közt nehezen találni programozásban jártasakat (hiszen a közismert informatikushiány miatt az ehhez értőket jobb anyagi feltételek várják másutt), viszont egy egyetem élhet azzal az “erőforrással”, amit a szakdolgozatot készítő hallgatók jelentenek – nem feledve azt, hogy a szakdolgozat készítésének mind időbeli, mind tapasztalati korlátai vannak, piacképes terméket elvárni illuzórikus. Kísérletezni, ígéretes utakat találni viszont lehet. [3]

## 2. Adottságok

Természetesen léteznek “kulcsrakész” rendszerek is, ezek azonban “mindentudó” (túl általános) célt kitűzve drágák és bonyolultak, hasznosításuk és a helyi elvárásokhoz igazításuk olyan tudást igényelhet, amely a programozóival vetekszik, de ahhoz képest túl specializált. A másik véglet a “partizánakcióban”, helyben végzett fejlesztések, amelyek magukon viselik a nem professzionális fejlesztői munka jellegzetességeit (frissítés, hiba- és verziókövetés, illetve dokumentáció hiányai, stb.), viszont a helyi igényekhez természetszerűen jobban alkalmazkodnak. Több azonos, vagy hasonló célra készülő szoftver közt azonban nincs átjárás, egymással nem kompatibilisek, még adatcsere szintjén sem. Itt tetten érhető egyfajta erőforráspazarlás is, ugyanakkor nincs kézenfekvő mód a hasonló fejlesztések koordinálására. Léteznek ugyan közösségi platformok, fejlesztésre is, de (nálunk) még nem tekinthetők bejáratott utaknak.

## 3. Játékoság (és tanulás) a játékprogramok oldaláról

A hetvenes évek végétől robbanásszerűen terjedő otthoni számítógépek egyik legfontosabb alkalmazása a játék volt. Ez adta el ezeket a tömegeknek. Az akkori gépek igen szerény képességei mellett is kellő találékonysággal készültek nagyon sikeres programok. Ezeknél azonban azt, hogy az adott játékot hogyan kell játszani, mik a szabályai és mi a játék elérendő célja, arról mindig valamilyen kísérő dokumentációból, vagy szóbeszédéből tájékozódhatott a felhasználó. Az akkori gépek memóriájába még ez a cikk sem férne bele.

Ahogy a hardver-feltételek javultak, vált megoldhatóvá a játék tanulásának eszközeit magába a játékba foglalni. Ez eleinte csak a kézikönyv digitális megfelelőjének játékba integrálását jelentette, s a *help* funkció is csak annyi előrelépést jelentett, hogy könnyebben kereshetővé vált egy-egy részlet. Kicsiny, de fontos lépés volt a *help* funkció *alkalmazkodó* változatának



megjelenése, ahol a program az adott szituációhoz közeli pontokat rendezte előre – ahogy az irodai programokban is. Nem is csoda, hiszen ezek a módszerek a programozók közt területtől függetlenül közkézen forogtak.

Más irányt jelentettek a *demók*. Itt a játékos passzívan, egy-egy példán át, egyfajta videoklipként tekinthette meg azt a szituációt, részletet, amelyet teljesítenie kellett. Ez is sok segítséget adhatott, támaszkodik arra a képességünkre, amellyel utánpótlás útján tanulunk. Ezek eleinte valóban rövid videókként kerültek rögzítésre (az ún. átvezető videókkal együtt), ez már a CD-k korszaka volt.

A két fenti *tanulást segítő funkció* az elmúlt évekre már jelentősen összemosódott. A mai játékok egyfajta virtuális világot képesek megjeleníteni. Ebben kaphatunk segítő karaktert, vagy karaktereket, amelyek bőven ellátnak tudnivalóval (írásban és szóban egyaránt), éppen akkor, amikor erre szükség van. A játék, vagy fogalmazzunk itt így, az adott *világ* szabályait pedig lépésről lépésre, szituációról szituációra (egyre nehezedő, komplexebb esetek) *tapasztalás* és *kísérletezés* útján (try and error) sajátíthatjuk el.

Ha a fiatalok örömmel és időt, energiát nem kímélve képesek egy fantáziavilág működését megtanulni, miért ne próbálnánk ezt a módszert a hétköznapi világ *szimulációjával* alkalmazni? [2]

#### 4. Összetett szoftver-eszközök használata

Egy háromdimenziós *real-time* videójáték, vagy *virtuális világ* mögött olyan komplex munka, idő- és energiaráfordítás rejlik, amit egy, vagy néhány *independent* (divatos rövidítéssel *indie*) programozó képtelen lenne belátható időn belül nagyon alacsony szinten is megismételni. De erre nincs is szükség, ugyanis ugyanebbe a problémába maguk a nagy játékfejlesztő cégek is "belebotlottak" korábban, üzletileg is működésképtelen volt mindent az elejéről megírni. Az újrahasznosítás és a modularitás egyébként is hívószavak ezen a területen. Az évek során előbb zárt, majd nyílt, vagy nyíltan használható ún. *game engine*-ek készültek és versenyeztek egymással. A programozónak ezeket használva nem kellett többé azon gondolkodnia, hogy mondjuk egy eldobott kő íve milyen módon kódolható le és hogyan jelenik meg kockáról kockára a képernyőn. Ez az *engine* dolga, a programozó csupán a megfelelő paraméterekkel (dobás ereje, iránya, szöge, dobott tárgy jellemzői) *meghívja* az engine megfelelő szoftver-szolgáltatását. [6]

A programozók eddig sem csináltak mindent egyedül. Ha egy problémát meg kellett oldani, feltételeztük, hogy ehhez már "alattunk" ott ül egy operációs rendszer, egy programnyelv a maga könyvtári rutinjaival, valamint egy adatbázis és hasonlók. Az engine sem más, mint egy ilyen eszköz – ha annak meglehetősen komplex is. Használata során meglehet, sokkal több készséget és időt kell forgatókönyv-íróként (ha-akkor szituációk), vagy építészként és lakberendezőként (környezet) felhasználnunk a kreativitásunkból és viszonylag keveset kódírással. Ennek megfelelően az ún. fejlesztői környezet is inkább egy CAD/CAM rendszerre, vagy egy videóvágóra hasonlít.

Az engine korlátozódhat csak a világ passzív szimulációjára – csak az mozog, amit mi mozgatunk –, ez is komoly erőforrás-igénnyel jár, ha a fizika minél valóságosabb (és nem szeretnénk falon át vagy húszméteres ugrásokkal közlekedő szereplőket a történetünkbe); de ma már elvárás, hogy legyenek ún. NPC karakterek, amiket a gép mozgat. Lehetőleg minél célszerűbben, intelligensebben. A mai engine-ek *mesterséges intelligenciát* is tartalmaznak – természetesen változatos, milyen minőségűt. Csodát ne várjunk, de a fejlődés gyors.

## 5. Egy szakdolgozat példája

Röviden egy példán át ismertetném, hogyan valósulhat meg egy ilyen alkalmazás. A hallgató vállalása egy KRESZ oktatóprogram elkészítése volt. Ehhez önállóan választott engine-t (az egyetlen kikötés az volt, hogy szabadon felhasználható legyen), majd közlekedési szituációkat kezdett megvalósítani úgy, hogy a felhasználó a "városban", az "utakon" más közlekedőkkel is találkozhat – ezeket az engine mesterséges intelligenciája mozgatta. Ki kellett dolgoznia magát a környezetet, a szabályoknak megfelelő módon megírnia a "forgatókönyvet" (mik a lehetséges interakciók, mik a hibásak, azaz szabálytalanok, és mik a tiltottak, mint pl. az átrepülés egy másik autó felett), egyben a megfelelő közlekedési táblákat és útburkolati jeleket is (grafikailag és interakció oldalról is). [4]



1. ábra A szakdolgozat[4] fejlesztői környezetének képernyőfotója egy példaszituációról

Az ilyen komplex feladatoknál az egyik veszélyforrás (déliabáb) az, hogy az elején látványosan sok dolog működni *látszik*, hiszen az engine csak arra vár, hogy megjelenítsen és megmozdítsa bármit, amit kódban adunk neki. Sokkal nehezebb, időigényesebb a *tesztelés*, hiszen összetettebb és szerteágzóbb a feladat. Rég magunk mögött hagytuk már a "lefordult, lefutott" módit...

A hallgató végül határidőre, sikeresen elkészítette feladatát a szakdolgozat összes követelményével, tehát a megfelelő szöveganyaggal együtt. Azaz az egyébként szűkös erőforrás-keretek (egy kevés tapasztalattal rendelkező készítő viszonylag rövid fejlesztési idővel) betarthatóak voltak. Ez azért fontos, mert azt mutatja, szűkös keretek közt is van remény oktatási célú alkalmazást létrehozni.

A külső bíráló véleménye a döntő a szakdolgozatok minősítésénél. Ebben az esetben a bíráló kiemelte, hogy működő, továbbfejleszhető(!) szoftver készült el, a hallgató ennek létrehozása kapcsán megfelelő fejlesztési tapasztalatra tett szert (valahol ez is a célja a szakdolgozatnak). A feladat "szokatlansága" miatt hangsúlyt kapott a megfelelő dokumentáltság is, ez követhető módon mutatta be magát a fejlesztést. Az, hogy szabadon felhasználható (ingyenes) szoftverekre épül, elvárás volt.

A konzulens felelőssége annak a többfordulós játszmának a megfelelő végigvitele, még időben, hogy a hallgató a képességeivel, érdeklődésével jó összhangot mutató feladatot válasszon. Neki, mint idősebbnek, tapasztaltabbnak, éreznie kell, mikor vállalná túl magát a hallgató. Az viszont nem érheti meglepetésként, ha a hallgató kreativitása meghaladja az övét. [5, 6]

## 6. Milyen oktatási területeken lehet bevetni az engine-eket?

Ez az a pont, ahol biztosat írni biztosan tévedés lenne. A kedves olvasó *saját* ötleteit ki nem találhatom. Adhatok néhány tippet, gondolati csírákat.

Többféle *game engine* létezik, ezeknek vannak erősségeik, gyengeségeik, közös vonásuk, hogy származásukat tekintve akciójátékokból jönnek, feladatuk az adott világ szimulációja. Hogy a fizika mennyire valóságos, az egy mércéje egy engine-nek. Vannak ugyanis területek, ahol nem annyira fontos a fizika, másban erős az az engine. Vannak látványban erős engine-ek és vannak a Minecraft kockavilágát tudók is.

Az oktatás szempontjából a szimulációs képességük a lényeges. Ahol a fizikai, természettudományos ismeretekben várunk segítséget, ott a lehető legkorrektebb fizikát megvalósító engine szükséges – és a szimuláció még ekkor sem egyenlő a valósággal! Máshol, mondjuk történelmi események feldolgozásakor inkább a látvány és áttekinthetőség fontos – lássuk például, hogy egy-egy csata min fordult meg, min dőlt el. A “mi lett volna, ha” kényes kérdés, de szimulációval megoldható. [6, 7]

Összefoglalva: A szoftverrendszerek fejlődése egyre közelebb hoz olyan lehetőségeket is, amelyekről korábban, bár akadtak volna jó ötletek, elképzelések, azok megvalósíthatósága irreálisnak tűnt. Érdekes lehet ezt most újra átgondolni. Olyan témákban, ahol a tanulók előismeretei nem elegendőek a megfelelő elképzelések kialakítására, vagy éppen a meglévő hétköznapi fogalmaik, elképzeléseik *félrevezetőek* lehetnek, különösen fontos lehet akár a korlátozott képességű, vagy minőségű szimuláció. A vezetett kísérletek és az oktatófilmek sem aktív, hanem passzív befogadásra készültek, míg az aktívan kipróbálható – és lehetőleg korrekt eredményt adó – szimulációk a *megértésen* keresztül (ún. “aha!” élmény) segíthetik a tanulást.

### Felhasznált irodalom

- [1] Nagy György: Digitális kihívások a környezetismeret oktatása terén. Képzés és gyakorlat: Training and Practice 16 : 4 pp. 73–80. 2018. DOI: [10.17165/TP.2018.4.7](https://doi.org/10.17165/TP.2018.4.7)
- [2] Nagy György: A kísérletezés szerepe a világ megismerésében. Új kihívások a pedagógia gyakorlatban konferencia; Pázmány Péter Katolikus Egyetem Vitéz János Kar, 2016.
- [3] Godó Zoltán Attila, Kocsis Dénes, Kiss Gábor, Stóka György: Age Independent Examination of Algorithm Creating Abilities: Computers in Education 39<sup>th</sup> International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2016) pp. 1125–1129
- [4] Smuczer Krisztián: KRESZ szituációs oktatóprogram. (szakdolgozat) Eszterházy Károly Egyetem, Comenius Campus, 2021.
- [5] Fromann Richárd, Damsai Andrei: A gamifikáció (játékosítás) motivációs eszköztára az oktatásban. Új pedagógiai szemle, 2016/3–4. ISSN 1788-2400
- [6] Jeff Ward: What is a Game Engine? [https://www.gamecareerguide.com/features/529/what\\_is\\_a\\_game\\_.php](https://www.gamecareerguide.com/features/529/what_is_a_game_.php) (hozzáférés: 2022. május)
- [7] Nicola Varcasara: Unreal Engine Game Development Blueprints. Packt Publishing, 2015, ISBN 978-1-78439-777-7

APPropó fejlődés  
A Bács-Kiskun Megyei Katona József Könyvtár mobilapplikációja

Csapó Noémi  
*Bács-Kiskun Megyei Katona József Könyvtár (Kecskemét)*  
[csapo.noemi@kjmek.hu](mailto:csapo.noemi@kjmek.hu)

Dani Erzsébet  
*Debreceni Egyetem, Bölcsészettudományi Kar, Média- és Könyvtártudományi Intézet, Könyvtár-  
és Információtudományi Tanszék (Debrecen)*  
[dani.erszebet@arts.unideb.hu](mailto:dani.erszebet@arts.unideb.hu)  
ORCID [0000-0002-8771-8335](https://orcid.org/0000-0002-8771-8335)

### Abstract

Noémi Csapó – Erzsébet Dani: Appetite for something new: The mobile application of the Katona József Library of Bács-Kiskun County

Libraries have always been a flagship in keeping up with technological changes. Adapting to the latest challenges of the digital world, they are experimenting with several of the ever-expanding range of virtual platforms.

The Katona József Library of Bács-Kiskun County in Kecskemét is also trying to seize every opportunity to keep up with the latest trends. We are on Facebook, Instagram, have a YouTube channel and a moly.hu account.

Though, a mobile app was still needed. Thanks to a few dedicated and enthusiastic librarians, we will be able to download our own custom-designed and developed app from the Google Play Store since the Autumn of 2021. The app was built using the Thunkable API, which was used to create the framework and structure of the app. And to create the content available in the app, we used a variety of online platforms and web applications specialised in creating interactive content/games.

The publication will give a comprehensive overview of the steps of the implementation, the applications and software used and our experiences during the work.

**Keywords:** public library, application, mobile app, library app, innovation

## Bevezetés

A XXI. század harmadik évtizedében kis túlzással már azt is mondhatjuk, hogy szinte mindenre van mobilapplikáció – vagy ha nincs, akkor hamarosan lesz. De vajon mi a helyzet a könyvtárakkal? Egyelőre azt láthatjuk, hogy 2022 tavaszán mindössze öt intézmény<sup>1</sup> rendelkezik saját appal. Közülük a Bács-Kiskun Megyei Katona József Könyvtár az, aminek legutóbb – 2021 őszén – felkerült az alkalmazása Google Play Áruházba<sup>2</sup>.

Könyvtárunk mindig is kiemelten fontosnak tartotta, hogy minden korosztály megszólítson. Célunk eléréséhez pedig elengedhetetlen a digitális korhoz (és annak kihívásaihoz) történő folyamatos és gyors alkalmazkodás. Éppen ezért gondolta úgy az intézményünkben működő *Elektronikus Információszolgáltatás (EISZ) munkacsoport* néhány tagja, hogy egy mobilapplikáció fontos állomás lehetne könyvtárunk életében. A közel háromnegyed évig tartó munka nem volt eredménytelen: a hat lelkes, elkötelezett kollégának köszönhetően elkészült saját tervezésű és fejlesztésű applikációnk Android operációs rendszerű okostelefonokra.

## Melyiket a kilenc sok közül?

A konkrét fejlesztési munkafolyamatot egy előkészületi szakasz előzte meg. Ez idő alatt kellett eldöntenünk többek között azt is, hogy milyen szerkesztővel kívánjuk elkészíteni az alkalmazást. Abban biztosak voltunk, hogy egy komolyabb szaktudást nem igénylőt kell majd választanunk, ugyanis a munkacsoport tagjai közül egyikünk sem jártas a programozásban. Szerencsére manapság már a WYSIWYG-jellegű<sup>3</sup> honlapkészítőkhöz hasonlóan a mobilapplikációk vonatkozásában is léteznek *no-code* szerkesztők. Sőt, már olyannyira elterjedtek ezek az – általában magukat *app creator/builder* néven hirdető – alkalmazások, hogy a kínálat áttekintése az előzetes feltételezéseinkhez képest hosszabb időt vett igénybe.

A válogatásban több fontos tényező is szerepet játszott. Az elsők között említhető a költséghatékonyság: szerettünk volna olyan szerkesztőt választani, amely ingyenes tagság mellett is számos – számunkra is releváns – funkciót kínál. Abban az ötletelés legelső perceiben is biztosak voltunk, hogy lesz néhány olyan menüpontunk, amelyek szeretnénk, ha már létező weboldalaink valamelyikére vezetnének (pl. a Katalógus menüpont a webOPAC-unkra). Éppen ezért azok az app builderek, amelyeknél nem volt feltüntetve, hogy egy adott honlapnak az applikáción belüli megnyitására szolgáló komponenssel is rendelkeznek, szóba sem jöhettek.

Annak, hogy miért Android operációs rendszerű eszközökre fejlesztettünk alkalmazást, több oka is van. Egyrészt a *Google Playre*<sup>4</sup> történő publikálás nemcsak, hogy egyszerűbbnek tűnt, hanem kétségkívül olcsóbbnak is – mivel a Google Playnél csupán egyszeri regisztrációs díjat

---

1 A Google Play Áruházban 2022 májusában a Bács-Kiskun Megyei Katona József Könyvtár mellett a Németh László Városi Könyvtár (Hódmezővásárhely), a Somogyi Károly Városi és Megyei Könyvtár (Szeged), a Vörösmarty Mihály Könyvtár (Székesfehérvár) és a József Attila Megyei és Városi Könyvtár (Tatabánya) applikációja szerepelt. Utóbbi négyet az *INFOtec Informatikai és Információtechnikai Kft.* készítette.

2 Google Play. A *Katona József Könyvtár alkalmazása*. Hozzáférés: 2022.05.27.

<https://play.google.com/store/apps/details?id=com.gmail.bacstudastar.kjkapp>

3 Wikipédia. WYSIWYG szócikk. Hozzáférés: 2022.05.27. <https://en.wikipedia.org/wiki/WYSIWYG>

4 Google Play kezdőoldala. Hozzáférés: 2022.05.27. <https://play.google.com/store>





kell fizetni<sup>5</sup> a fejlesztői fiók regisztrálásakor. Emellett azt feltételeztük, hogy felhasználóink is nagyobb arányban rendelkeznek a Google operációs rendszerét futtató eszközökkel. Szintén nem mellékes szempont, hogy a munkacsoport tagjai mind androidos okostelefonokkal rendelkeznek. Mindezekből következik, hogy csakis olyan app creatorokat kerestünk, amelyek biztosítják a teszteléshez az alkalmazás APK (*Android Package Kit*) fájlként történő letöltését – lehetőleg ingyenesen. Mivel nem volt még tapasztaltunk semmilyen app builderrel, így döntésünk során azt is figyelembe vettük, hogy az adott szerkesztőhöz tartozó dokumentáció mennyire részletes, informatív.

## Thunkable

A felsorolt szempontokat figyelembe véve és mérlegelve végül úgy gondoltuk, hogy a *Thunkable*<sup>6</sup> segítségével próbáljuk meg elkészíteni applikációnkat. Ez a platform volt az, amelyről az előzetes „piackutatásunk” alapján úgy gondoltuk, hogy a legtöbb – számunkra is lényeges – funkcióval rendelkezik, ráadásul ezek mind elérhetőek a *Free Plan* esetében is.

A Thunkable egy mobilalkalmazások fejlesztésére szolgáló, nem nyílt forráskódú, ún. no-code platform. Valamennyi, a segítségével készült alkotásból generálható egyaránt Android és iOS operációs rendszerű eszközökön futó app is.

A Thunkable lehetővé teszi a kész(ülőfélben levő) appunk APK fájlként történő letöltését. Az APK fájlnak nagyon nagy szerep jutott a tervezési fázisban. A jelentősebb módosításokat követően ugyanis minden csapattag letöltötte telefonjára az aktuális próbaverzió APK-ját, így a kollégák valamennyi eszköze esetében tesztelhattük, hogy elégedettek vagyunk-e kijelzőn látható eredménnyel. A Thunkable melletti döntésünket erősítette, hogy egyszerre összesen akár tíz ún. projektet is kezelhetünk – az ingyenes fiók esetében is. Így adott a lehetőség az alkalmazás egyes verzióinak biztonsági elmentésére. Szintén a Thunkable pozitívumai közé sorolható, hogy ez a szerkesztő igen részletes dokumentációval<sup>7</sup> rendelkezik. Emellett a Thunkable-t használó fejlesztők közössége meglehetősen aktív: a fórumukat<sup>8</sup> böngészve is számos hasznos információhoz juthatunk, illetve a YouTube-on is lelkesen osztják meg tudásukat a világgal.

Apró negatívum, hogy a Thunkable szerkesztőfelülete sajnos nem érhető el magyarul. Viszont használatához tapasztalatunk szerint nincs is szükség komoly (szak)nyelvi ismeretre. A különféle funkciók feltérképezése – szükséges mennyiségű próbálkozással és egy nagy adag kitartással kiegészítve – „alap” angoltudással sem okozott gondot.

A Thunkable-ben alapvetően két felületen dolgozhatunk: a *Design* és a *Blocks* „fülön”. Előbbi – mint neve is mutatja – a megjelenítésért felel. Ami viszont talán még a „külcsinnél” is fontosabb: a Design felületén építhetjük fel az alkalmazás vázát, gerincét (1. ábra). Itt hozhatjuk létre az úgynevezett *screen*eket, és itt mondhatjuk meg, hogy szeretnénk-e rájuk valamilyen komponenst – pl. gomb(ok)at. A *Blocks*nak sem kevésbé lényegtelen a szerepe: ő az „életre keltésért” felel (2. ábra). Itt adhatjuk meg ugyanis az eseményeket (pl. mi történjen abban az esetben, ha rábökünk egy gombra) egy vizuális szerkesztő segítségével. A különféle komponensek és a hozzájuk tartozó „parancsok” is – kis képzelőerővel – építőkövekre

5 Play Console Súgó. A *Play Console* használata – Google Play fejlesztői fiók regisztrálása. Hozzáférés: 2022.05.27.

<https://support.google.com/googleplay/android-developer/answer/6112435?hl=hu#zippy=%2Clépcs-a-regisztrációs-díj-megfizetése>

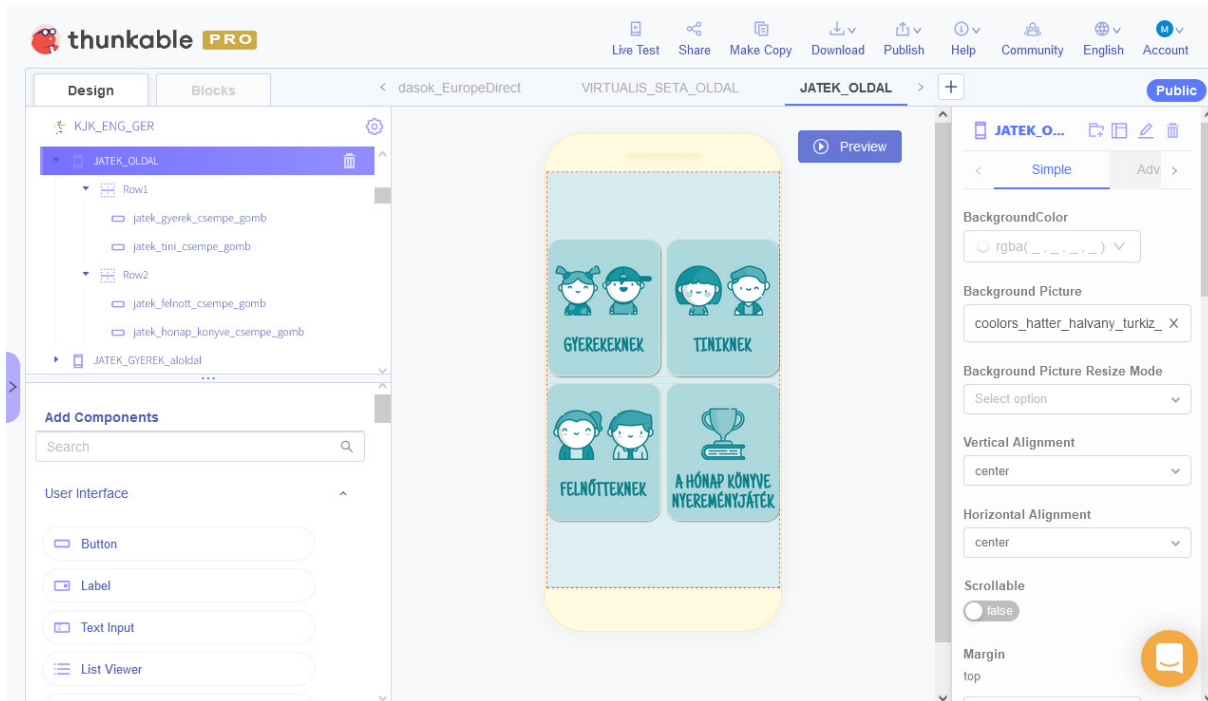
6 Thunkable kezdőoldala. Hozzáférés: 2022.05.27. [www.thunkable.com](http://www.thunkable.com)

7 Thunkable Docs kezdőoldala. Hozzáférés: 2022.05.27. <https://docs.thunkable.com>

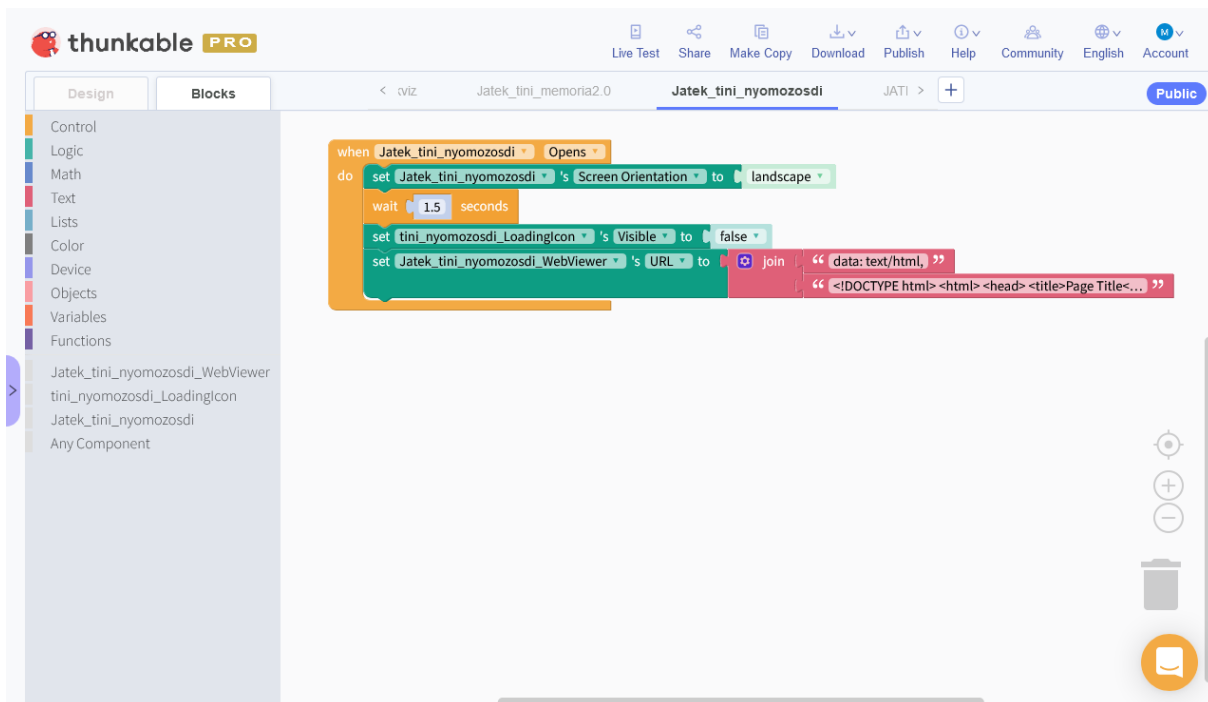
8 Thunkable Community kezdőoldala. Hozzáférés: 2022.05.27. <https://community.thunkable.com>

hasonlító ún. blokkok formájában jelennek meg. Ezeket a – funkciójuktól/típusuktól függően különféle színben pompázó – blokkokat kell egymáshoz illeszteni, ha egy új eseményt szeretnénk létrehozni.

Azt mindenképpen szeretnénk kiemelni, hogy bár a Thinkable számos hasznos funkcióval, komponenssel rendelkezik, mi ezeknek csak töredék részét használtuk az applikációban.



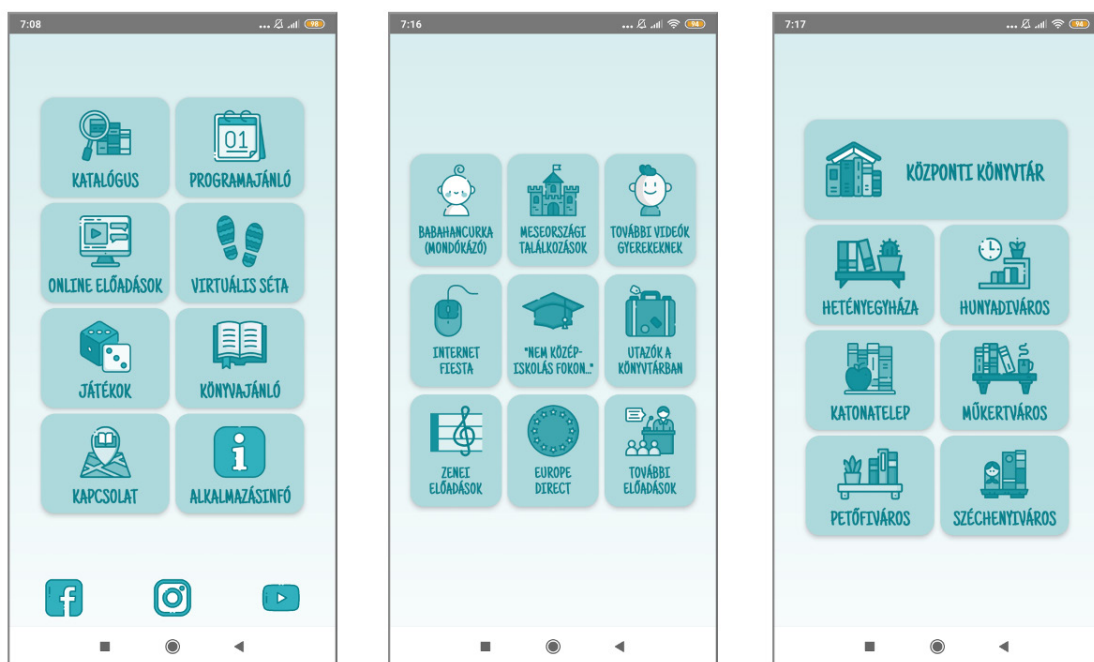
1. ábra: A Thinkable Design felülete



## Megjelenés...

Alkalmazásunk nem jelenlegi weboldalunk ([www.bacstudastar.hu](http://www.bacstudastar.hu)) helyettesítésére készült, és nem is annak okostelefonra optimalizált változata. Így nem is ugyanazok a menüpontok szerepelnek az appban, mint honlapunkon. Végül nyolc menüpontban állapodtunk meg. Vannak köztük olyanok, melyek egy könyvtári mobilapp esetén szinte „kötelezőnek” tekinthetők (*Katalógus*, *Könyvajánló*, *Programajánló*), valamint olyanok is, melyek általánosak, számos mobilapp esetén előfordulnak (*Kapcsolat*, *Alkalmazásinfó*). A további három menüpont (*Virtuális séta*, *Játékok*, *Online előadások*) segítségével pedig – reményeink szerint – egy kis egyediséget, pluszt csempészhetünk alkalmazásunkba. Honlapunkhoz viszonyítva az alkalmazásban természetesen kevesebb tartalom található. Ráadásul a szöveges helyett inkább a képi tartalmak, videók képviseltetik magukat. A kevesebb tartalomnak köszönhetően viszont applikációnk könnyedén áttekinthető, használata pedig valamennyi korosztály számára egyszerűen elsajátítható.

A megjelenés tekintetében egységességre törekedtünk: hasonló stílusú ikonokat használtunk a főoldal és valamennyi aloldal esetében is (3. ábra). A színvilág vonatkozásában pedig megmaradtunk a – már védjegyünknek számító – türkiz szín mellett. Ez a szín – illetve árnyalatai – egyébként a(z) (al)menüpontok mellett a könyvajánlóban és a programajánlóban is visszaköszön.



3. ábra: Példák az egységes megjelenésre  
(az applikáció főoldala, valamint két aloldal: az Online előadások és a Kapcsolat)

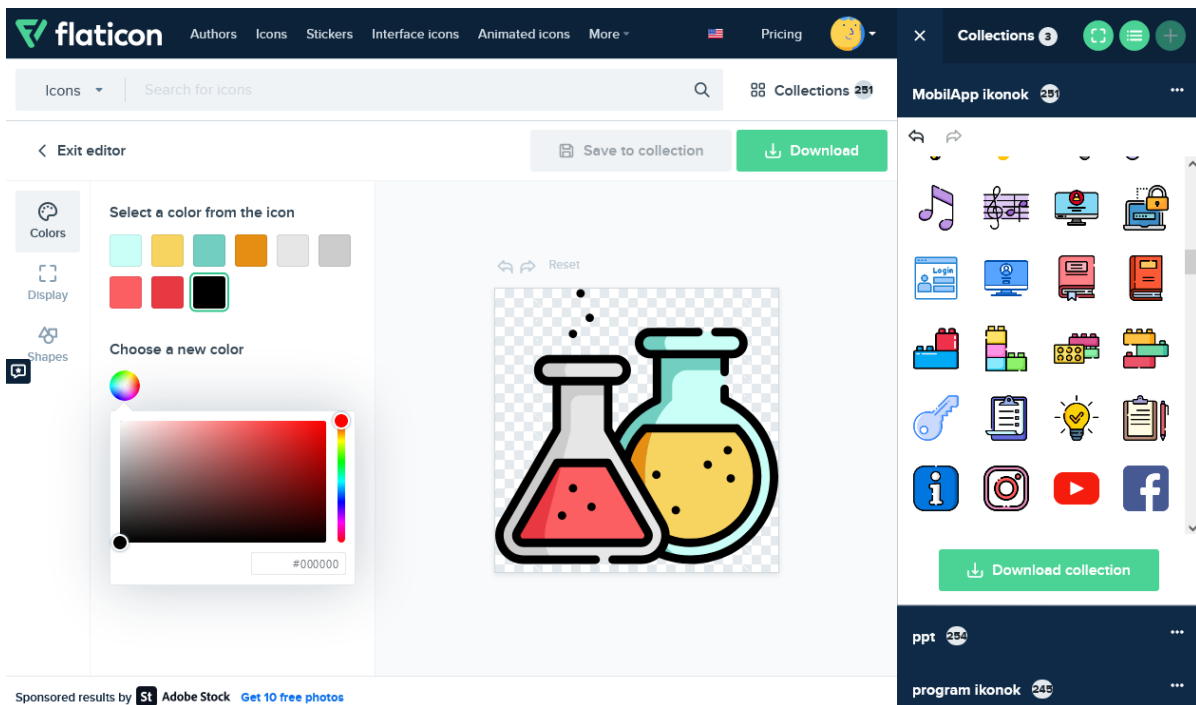
A gombokon látható valamennyi ikon a *Freepik Company*<sup>9</sup> *Flaticon*<sup>10</sup> elnevezésű felületéről származik. Amellett, hogy hasonló formavilágúakat igyekeztünk összegyűjteni, arra is figyeltünk, hogy a kiszemelt képek ingyenesen letölthetők legyenek. Az alkotók nevének feltüntetése a szabad felhasználás ellenében sem maradhatott el: az Alkalmazásinfó menüpontban valamennyiük neve szerepel. Az ikonok az applikációban látható türkiz árnyalatukat csak igen aprólékos munka árán nyerhették el: az eredeti színek külön-külön, egyesével történő átszínezésével. Erre szerencsére a Flaticon oldalán is lehetőség volt (4. ábra).

<sup>9</sup> Freepik Company kezdőoldala. Hozzáférés: 2022.05.27. <https://www.freepikcompany.com/>

<sup>10</sup> Flaticon kezdőoldala. Hozzáférés: 2022.05.27. <https://www.flaticon.com>



Az egyes gombokon látható ikon, illetve szöveg (vagyis az adott (al)menüpont neve) nem független egymástól. Tulajdonképpen egy PNG formátumú képfájlról van szó, amely a Thunkable-ben *Background Picture*-ként került feltöltésre az adott gombhoz. Ezek a képek a *GIMP*<sup>11</sup> nevezetű ingyenes képszerkesztő programmal készültek. A gombok felirataiban esetében igyekeztünk olyan fontot választani, amely – amellett, hogy betűkészlete a „speciálisabbnak” számító ö-ő és ü-ű karaktereket is tartalmazza – kicsit „fiatalosabb”, ugyanakkor könnyen olvasható. Választásunk végül a *Just Another Hand*<sup>12</sup> elnevezésű betűtípusra esett.



4. ábra: Ikonok átszínezése a Flaticon oldalán

## ...és tartalom

Az app menüpontjai – a Kapcsolat és Alkalmazásinfó kivételével – valamilyen weboldalra vagy internetkapcsolattal elérhető interaktív tartalomra vezetnek. Ezt a Thunkable *Web Viewer*<sup>13</sup> nevezetű komponensével tudtuk megvalósítani. Segítségével a megjeleníteni kívánt honlapok és webes tartalmak az alkalmazás keretein belül töltődnek be, így nincs szükség arra, hogy a telefonra telepített böngésző(k valamelyike) is megnyíljon a könyvtári mobilapp mellett.

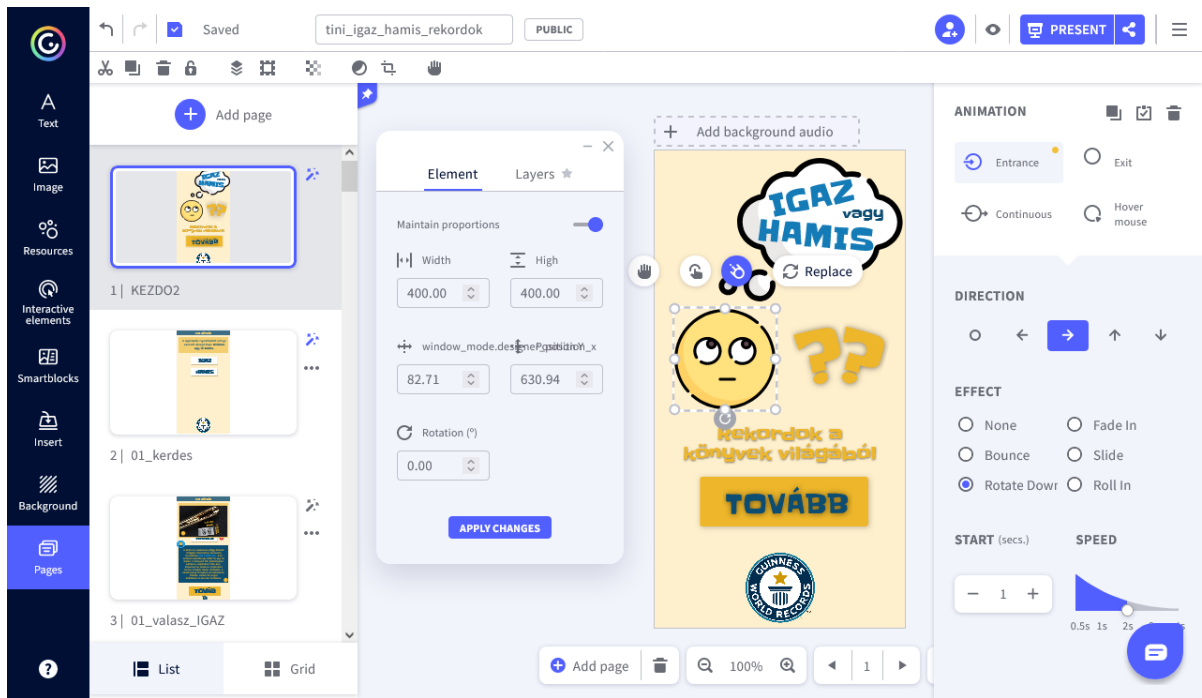
A Web Viewernek azonban van egy apró negatívuma: az általa betöltésre kerülő tartalmakon belül már „túl jól” működik az androidos készülékek alapértelmezett – a navigációs sávban található – vissza gombja. Lenyomásával ugyanis nem visszalépünk egyet, hanem teljes mértékben kilépünk a Web Vieweren keresztül megjelenített tartalomról. Ez a mi alkalmazásunk esetében a webOPAC-ot és a YouTube lejátszási listákat érinti. Éppen ezért a Katalógus és az Online előadások almenüpontjait megnyitva látható, hogy a kijelző alsó sávjában megjelenik egy *Vissza* és egy *Előre* gomb. Segítségükkel már teljes értékűen tudunk navigálni – ugyanúgy, ahogy a böngészőt használva tennénk.

11 GIMP kezdőoldala. Hozzáférés: 2022.05.27. <https://www.gimp.org>

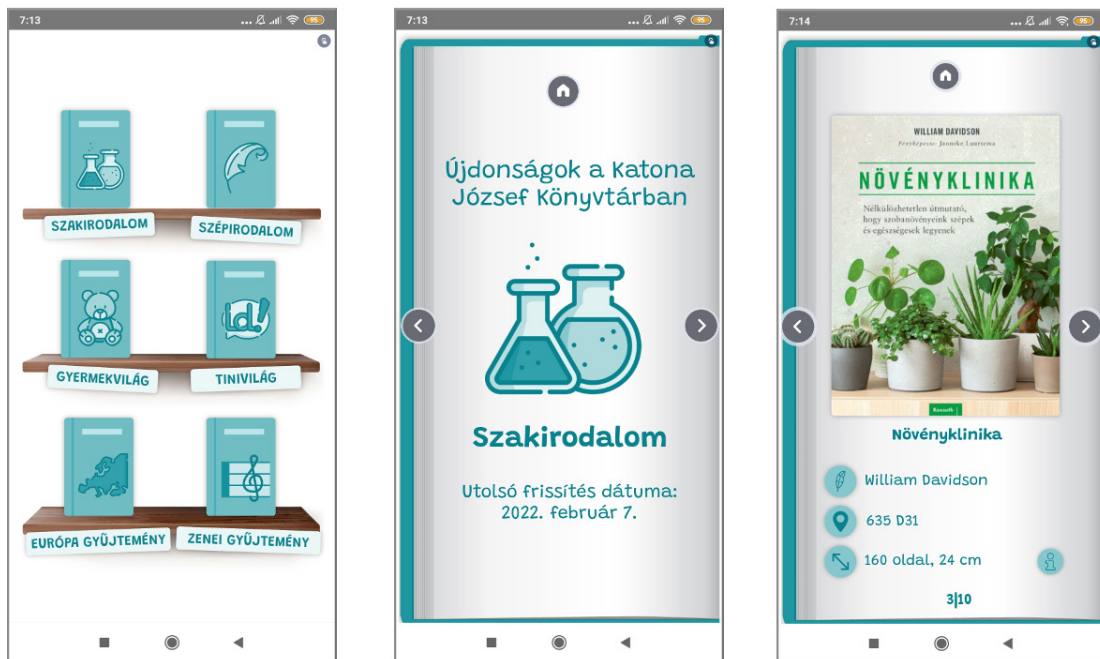
12 1001 Fonts honlapja. *Just Another Hand* betűtípus. Hozzáférés: 2022.05.27. <https://www.1001fonts.com/just-another-hand-font.html>

13 Thunkable Docs. *Web Viewer*. Hozzáférés: 2022.05.27. <https://docs.thunkable.com/web-viewer>

A saját készítésű tartalmak zöme az interaktív, animált tartalmak létrehozására szolgáló Genially<sup>14</sup> segítségével született meg (5. ábra). Ezen az online platformon készültek el – egy kivételtől eltekintve – a tiniknek és a felnőtteknek szánt játékok, A hónap könyve nyereményjáték, valamint a könyvajánló (6. ábra) és a programajánló is. Utóbbi kettőt természetesen folyamatosan frissítjük. A Genially esetében pedig tökéletesen megoldott, hogy a felhasználók is azonnal értesüljenek minden újdonságról, újonnan feltöltött tartalomról. Ugyanis amint elmentjük módosításainkat a Genially felületén, és telefonunkon megnyitjuk az adott tartalmat, máris a frissített változatot láthatjuk.



5. ábra: A Genially szerkesztőfelülete



6. ábra: Képernyőképek a Genially segítségével készült könyvajánlóból

14 Genially kezdőoldala. Hozzáférés: 2022.05.27. <https://genial.ly/>

A Genially mellett a játékok készítéséhez még az *Educandy*<sup>15</sup>, az *Interactyt*<sup>16</sup>, a *Jigsaw Planetet*<sup>17</sup> és a *LearningAppst*<sup>18</sup> használtuk.

A Kapcsolat és az Alkalmazásinfó menüpont esetében sajnos a Thunkable-t nem tudtuk rávenni, hogy az elképzelésünknek megfelelően jelenítse meg az adott menüponthoz tartozó információkat tartalmazó HTML dokumentumokat. Így kompromisszumot kötöttünk: a feltüntetni kívánt adatokat képként, PNG formátumban töltöttük fel. Bár ez egy nagyon egyszerű megoldás, így legalább a kijelzőn megjelenő tartalom teljes mértékben illeszkedik az alkalmazás stílusához, kép- és színvilágában egyaránt.

## Google Play

A több hónapos munkát, a számtalan kísérletezést és tesztelést követően 2021 őszén úgy éreztük, elkészült applikációnk első verziója – így már csak a Google Playre történő feltöltés volt hátra. Az alkalmazás tervezési szakaszában még az APK fájl is elegendő lett volna az app áruházba történő feltöltéséhez. Mire azonban eljutottunk a Google Playre történő publikálás kapujába, a Google-nél meggondolták magukat. 2021 augusztusától már csak kizárólag AAB (*Android App Bundle*)<sup>19</sup> formában lehetett „nevezni” az új applikációkat. Szerencsénkre a Thunkable-ról – egy hónapnyi *Pro* tagságért cserébe – ezt a formátumot is letölthettük.

Az AAB fájl mellett szükségünk volt még egy *Google Play Developer Accountra*, melyhez egy egyszeri regisztrációs díjért cserében jutottunk hozzá. A Google részletes dokumentációt<sup>20</sup> biztosít az új alkalmazás létrehozására és az ehhez kötődő beállításokra vonatkozóan is. Ez első olvasásra meglehetősen bonyolult folyamatként tűntette fel a feltöltés metódusát, de szerencsére a gyakorlatban – az előzetes elképzeléseinknél – egyszerűbben tudtuk megtenni a publikáláshoz vezető lépéseket. Azt viszont még szoknunk kell, hogy a Google-nél nem hamarkodják el a dolgokat. Az app első verziójának feltöltésekor több mint egy hetet kellett arra várni, hogy az alkalmazás állapota az „Ellenőrzés alatt”-ról „Közzétételre kész”-re váltsón, és így ténylegesen nyilvánossá (és letölthetővé) váljon a nagyközönség számára. Verziófrissítés esetén is napokba telt az ellenőrzési folyamat. Éppen emiatt van hatalmas jelentősége annak, hogy könyv- és programajánlókat a Google Playre történő bármilyen újabb verzió(k) feltöltése nélkül, bármikor frissíteni tudjuk – egyszerűen a Genially felületén.

Applikációnk publikáció utáni életének nyomon követésére a *Google Play Console*<sup>21</sup> felületén, a *Statisztikák* menüpont alatt van lehetőségünk (7. ábra). Segítségével a legapróbb részletekig tájékozódhatunk az app állapotáról. Az újonnan szerzett felhasználók, a felhasználóvesztések, vagy az aktív telepítések számának nyilvántartása csak a jéghegy csúcsa. A Google számtalan további szempontot elemezve készít számunkra – exportálható – grafikonokat, idősorokat és adattáblázatokat.

---

15 Educandy kezdőoldala. Hozzáférés: 2022.05.27. <https://www.educandy.com>

16 Interacty kezdőoldala. Hozzáférés: 2022.05.27. <https://interacty.me>

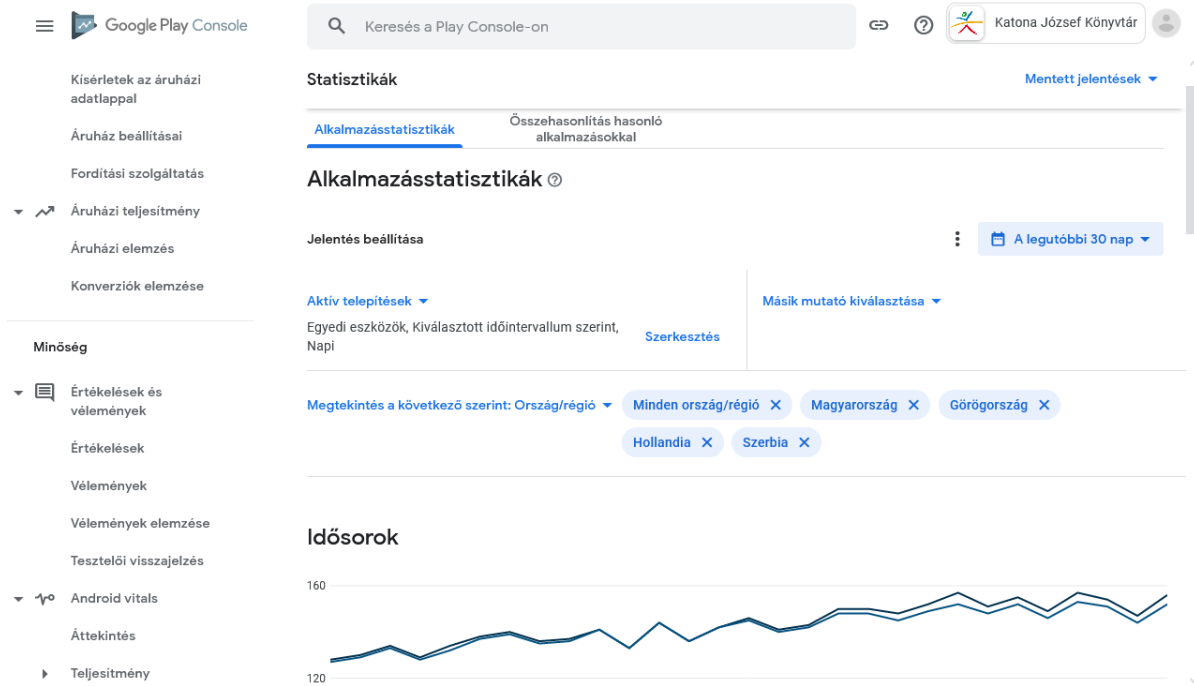
17 Jigsaw Planet kezdőoldala. Hozzáférés: 2022.05.27. <https://www.jigsawplanet.com/?lang=hu>

18 LearningApps kezdőoldala. Hozzáférés: 2022.05.27. <https://learningapps.org>

19 The future of Android App Bundles is here. = Android Developers, 2021. június 29. Hozzáférés: 2022.05.27. <https://android-developers.googleblog.com/2021/06/the-future-of-android-app-bundles-is.html>

20 Google Play Console Súgó honlapja. Az alkalmazás létrehozása és beállítása. Hozzáférés: 2022.05.27. <https://support.google.com/googleplay/android-developer/answer/9859152?hl=hu>

21 Google Play Console kezdőoldala. Hozzáférés: 2022.05.27. <https://play.google.com/console/about/>



7. ábra: Alkalmazásstatisztika a Google Play Console felületén

## Jövőkép

Mivel alkalmazásunk még csak viszonylag rövid ideje szerepel a Google Play Áruház virtuális polcainak egyikén, így komolyabb következtetéseket egyelőre nem szeretnénk levonni eddigi „pályafutásáról”. Azt azonban a Google Play Console statisztikái már most is egyértelműen mutatják, hogy az olvasók kegyeiért meg kell „küzdenünk”. Ha a tartalmak gyakori frissítése, az alkalmazás időnkénti fejlesztése, valamint a rendszeres népszerűsítés, reklámozás hármásából valamelyik tényező kiesik, akkor nem számíthatunk semmi jóra. A felhasználószerezés mellett viszont az applikációt letöltők bizalmának hosszú távra történő megnyerése sem kevésbé fontos feladatunk. Ennek érdekében a jövőben (is) kiemelt figyelmet fordítunk arra, hogy alkalmazásunk naprakész legyen: rendszeresen frissítjük könyv- és programajánlónkat, valamint az Online előadások menüpont alatt elérhető YouTube lejátszási listákat. Természetesen az alkalmazás monitorozása folyamatos, és a jövőben is ehhez szeretnénk tartani magunkat. Rövid távú céljaink között szerepelt még az app angol és német nyelvű felületének létrehozása is. Ezt sikeresen teljesítettük, a Google Playen már elérhető az alkalmazás 3.01-es verziója, amely tartalmazza ezt az újítást. Emellett a játékok frissítése is szerepel az elkövetkező időszakban elvégzendő feladatok listáján.

Reméljük, hogy alkalmazásunk – melyet szívvél-lélekllel készítettünk – a kecskeméti minél szélesebb körének elnyeri majd tetszését. Éppen ezért igyekszünk továbbra is minden tőlünk telhetőt megtenni azért, hogy applikációnk hosszú távon is egy kedvelt, népszerű szolgáltatás maradjon.

## Köszönetnyilvánítás

Köszönet az EISZ munkacsoport tagjainak: Boldizsár Dórának, Katona Tibornak, Molnár Dánielnek, Rédei Árpádnénak és Virág Barnabásnak, amiért mertek „nagyot” gondolni, és a kezdetektől hittek az app megvalósíthatóságában.

Köszönet továbbá Mikulás Mártonnak és Takács Krisztina Ágnesnek: segítségük nélkül az alkalmazás angol, illetve német nyelvű felülete nem jöhetne létre.

## Integrált könyvtári rendszerek tranzakciós rekordjainak vizsgálata, a könyvtári állomány digitalizálásának tervezésekor

Simon András  
ELTE BTK KITDI – PhD hallgató  
Monguz Kft., Szeged, ügyfélmenedzser  
Institute of Library and Information Science, Faculty of Humanities, Eötvös Loránd University,  
Budapest, Hungary; PhD Student  
Monguz Ltd., Szeged, Hungary CS Manager  
ORCID: [0000-0002-7990-3936](https://orcid.org/0000-0002-7990-3936)  
[asimon@monguz.hu](mailto:asimon@monguz.hu)  
[simon.andras.bela@btk.elte.hu](mailto:simon.andras.bela@btk.elte.hu)

### Absztrakt

Mind a digitálisan született, mind az analóg módon keletkezett, és digitalizált tartalmak tömege állandóan növekszik. Tárolásuk, őrzésük és szolgáltatásuk tekintetében megkerülhetetlen a hosszútávú őrzés nagy költségekkel járó megoldása. Természetesen az analóg tartalmak állományvédelme is költséges, de ennek, illetve a megőrzendő objektumok válogatásának, selejtezésének bejáratott módszerei, megoldásai vannak. A digitális tartalmak esetében ráadásul a közgyűjtemények rendelkezésére álló tárolási kapacitás növelésének is jelentős korlátjai vannak. Az IT szektorban lassan állandósuló szakemberhiány mellett a gyártási kapacitások és a nyersanyag készletek is várhatóan nehezen fognak lépést tartani a jövőben a világban tapasztalható, az adathordozók iránt megnyilvánuló keresletnövekedéssel. A tartós őrzés költségeinek mérséklésének lehetséges módja, az őrzendő állomány mennyiségének csökkentése. A digitálisan keletkezett, illetve már digitalizált anyagok esetében célszerű elsősorban a primer és szekunder adatok elválasztása, a még analóg formában létező, digitalizálásra váró dokumentumok esetében pedig a digitalizálásra és tartós őrzésre érdemes tartalmak kiválogatása. A dokumentumszolgáltatás céljából történő digitalizálás során a válogatás szakszerű megalapozásához fontos lehet az integrált könyvtári rendszerekben őrzött tranzakciós rekordok vizsgálatával foglalkozni. A rendszerek kölcsönzési moduljai ugyanis évekre - évtizedekre visszamenőleg őrzik a tranzakciós rekordokat, egyebek mellett a kölcsönzési események tranzakcióinak rekordjait is. Ezeket elemezve megállapíthatjuk, milyen korú és milyen témájú könyvek kerülnek a legnagyobb tömegben kölcsönzésre. A szerző egyik kutatása könyvtárak százainak a kölcsönzési tranzakciós rekordjait vizsgálta a kölcsönzött könyvek átlagos korának tekintetében, ezen kutatás eredményéről számolna be tanulmányában.

**Analysing of the transaction records of the integrated systems to support planning the digitization in the libraries**

### Abstract

The quantum of data both born digital and digitized, stored on electronic devices, grows constantly larger. Storing, safe-keeping and service of these data needs expensive solutions of Long Time Preservation. Hence the preservation of analogue documents is costly too, experts of public collections are traditionally experienced in preservation and in disposal of traditional





documents. However, the digital documents seem to need less space, the possibility of increasing the electronic storage capacity is limited. Beside the almost constant lack of experts in the IT sector, the manufacturing capacity and the raw material supplies are going to be insufficient according to the booming international demand for data carriers. So, it seems to be the only way of cost reduction of preservation is to decrease the quantity of the stocks are to be kept. Nevertheless, it is essential to sort the indispensable primary and the derived secondary contents, and to sort out the analogue documents worth to be digitized and preserved. The information, extracted from the transaction records of databases of integrated library management systems can be an exact point of view of selecting the contents to be digitized for document service. The loan modules of library systems preserve the loan transaction records for years or even for decades. The analysis of these records can help us to realize, what kind and what age of books are most often borrowed. One of the author's research concerns to the transaction data of hundreds of libraries, examining the difference between the year of borrowing and the year of publication of a book, so establishing the average age of the loaned books. The results of this research will be demonstrated in this paper.

#### Tárgyszavak

Integrált könyvtári rendszer

Adatbázis analízis

Digitalizáció

#### Bevezető

Évről évre hatalmas mennyiségben jöttek létre digitális és digitalizált tartalmak. Intézmények sokasága digitalizált formában biztosítja az információforrások szinte teljes körét. Részben az analóg tartalmak retrospektív digitalizálása terén történt óriási előrelépés, részben pedig már digitálisan jönnek létre az újonnan keletkezett anyagok az élet szinte minden területén. Mindez felveti a digitális információ hosszú távú megőrzésének kérdését, elkerülendő egy újabb „Dark Age” beköszöntét.

A korábban – értelmezhetetlenként és felhasználhatatlanként – tartós tárolásra nem került adatokat is már érdemes hosszú távra megőrizni, ezek ugyanis költséghatékonyan újrahasznosíthatóvá váltak a korszerű adatbányászati eszközöknek köszönhetően.<sup>1</sup> A digitális adatvagyon mennyiségének növekedésével annak pénzben kifejezhető értéke is mind nagyobb és nagyobb lesz, ezért a sérülések, adatvesztések következtében előálló potenciális kár mértéke is nő. Az egyre sokoldalúbb kezelő és szolgáltató felületek, a növekvő igények, a mind újabb hardver eszközök a rendszerek fejlesztését, mind gyakoribb frissítését és ezzel együtt járóan, a mind sűrűbb adatmigrációt követelik meg. Mindezek eredményeként a digitális tartalmak hosszú távú őrzése egyre költségesebb lesz.

#### A hosszú távú megőrzés fogalma, és célja

Az elektronikus adatkezelő rendszerekben a digitálisan tárolt információ nagy költséggel jön létre. Az ilyen formában létrejött és tárolt információ egyedi, pótolhatatlan, ezek esetében a sérülés, adatvesztés bármilyen formája megengedhetetlen. Az adatsorok (dokumentált)

1 Wenrui Yan, Jie Yao, Qiang Cao, And Changsheng Xie: A Rack-based Optical Storage System with Inline Accessibility for Long-Term Data Preservation= ACM Transactions on Storage, Vol. 14, No. 3, Article 28. November 2018.



hitelességét, sértetlenségét, integritását és eredetiségét informatikai eszközökkel szükséges biztosítani. Az adatokat sértetlenül meg kell őrizni és az őrzés folyamatosságának tényét dokumentálni kell. Kiemelt gonddal kell kezelni a digitálisan született tartalmakat, illetve azokat az eseteket, amelyeknél nem áll rendelkezésre jó fizikai állapotú eredeti tartalom.<sup>2</sup>

A hosszútávú megőrzés esetében az időtartamon itt azt az időt kell értenünk, amely alatt olyan alapvető, és előre nem látható szociokulturális, és technológiai változások következnek be, mely alatt a keletkező digitális adattartalmak mind formájukban, mind pedig felhasználásuk tekintetében nagy mértékben megváltoznak, és ennek megfelelően a tartalmakat előállító és szolgáltató alkalmazások is döntő változásokon mennek keresztül, így a korábban létrejött tartalmak ezen alkalmazások által kézenfekvően már nem használhatók fel.<sup>3</sup>

## Megőrzés és helyreállítás

Az adatbázisok őrzése, integritásuk biztosítása elsőrangú feladat, a – már minden működési elemében – informatikai megoldásokra építő modern könyvtári gyakorlatban. A digitális adatok megőrzése esetében meg kell különböztetnünk a preventív megőrzést, (preventive preservation) és a helyreállító megőrzést, (curative preservation). Az előbbi esetben az informatikai rendszerünket felkészítjük a lehetséges nemkívánatos változásokra, és az adatokat megpróbáljuk megóvni attól, hogy hardveres, vagy szoftveres okokból kárt szenvedjenek. Az utóbbi esetben a már bekövetkezett káresemény után próbáljuk meg az adatokat eredeti integritásukban és egymás közötti sértetlen összefüggéseikkel helyreállítani. A helyreállítás nemcsak magát az adatfolyamot (bitstream) kell, hogy érintse, hanem rendelkezésre kell állni az alkalmazásnak is, amely az adatfolyamot fel tudja olvasni, és a felhasználó számára, valamilyen periférián, jellemzően egy képernyőn emberi szemmel olvashatóan meg tudja jeleníteni. Ez az idő múlásával egyre több gondot tud okozni, ez teszi a hosszútávú megőrzést a kurrens használat számára való tárolásnál költségesebbé.<sup>4</sup>

Az adatok őrzésének elsőrendű feladata tehát az adatfolyam (bitstream) sértetlenségének biztosítása, valamint a leíró, technikai és adminisztratív metaadatok megőrzése. A metaadatok az adatokat azok keletkezésével összefüggő információkkal együtt őrzik, és emellett a későbbi archiválási folyamatot is dokumentálják. Meghatározó elvnek kell tekinteni, hogy fáradozásaink eredményeképpen nem maguknak a digitális objektumoknak kell feltétlenül sértetlenül fennmaradniuk, hanem mindenekelőtt a bennük tárolt információ megőrzésére kell törekednünk.<sup>5</sup>

Alapvetően tehát nem magát az objektumot őrizzük, hanem annak digitális másolatát vagy szurrogátumát, mely alapján a tárolt információt, illetve az objektumot annak a jövőben szükséges interpretációjához helyre tudjuk állítani. Az adatbázist számos behatás érheti, ezeket két részre oszthatjuk, tervezettekre, és előre nem elgondoltakra. Ez utóbbiak, mivel a

---

2 Fröhlich, Susanne: Der Weg ist das Ziel! Planung und Umsetzung digitaler Archivprojekte. = Mitteilungen der VÖB Vol. 66. Nr. 1 2013 (133–144p.)

3 Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland nestor – Network of Expertise in Long-Term Storage of Digital Resources = Printversion 2.0 beim Verlag Werner Hülsbusch, Boizenburg

4 Tripathi, Sneha: Digital preservation: some underlying issues for long-term preservation. = Library Hi Tech News, Number 2. 2018. 8–12p.

5 Oßwald, Achim: Langzeitsicherung digitaler Informationen durch Bibliotheken. =Zeitschrift für Bibliothekswesen und Bibliographie. 2014, Vol. 61 N. 4-5.



nem tervezett változás egy működő rendszer esetében ritkán jelenthet jót, minden esetben nem kívántaknak, baleset jellegűnek kell tekintenünk.<sup>6</sup>

A tervezett események közé tartozik a frissítés, az adatbázis költöztetése, a rendszer cseréje és az intézményt érintő szervezeti változásokat követő jogosultsági hozzáférési változások.

### A hosszú távú megőrzés költségessége.

A hosszú távú megőrzés hallatlanul költséges, ezt bele kell kalkulálni ma már minden projektbe, ezért a költségek kézben tartása elsőrendű feladat. A költségek felbecsülésének kulcsa az őrzendő állományok méretének, és a formátumok sokféleségének ismerete.<sup>7</sup>

Az adatok hosszú távú tárolását, és még inkább annak költségeit körültekintően kell megtervezni, főként, mert a hosszú távú megőrzés lényegesen költségesebb, mint a rövid távú, a napi használathoz tartozó őrzés. A tervezés nem könnyű feladat, mert a meglévő ismereteink alapján kell a távlati trendeket a közelmúlt változásainak elemzésével készített extrapolációk alapján kiszámítani.<sup>8</sup>

Az adatok tárolásának és őrzésének költségei az adattömeg növekedésével egyre feltűnőbben jelentkeznek. Az adatok maguk, továbbá indexállományokban, és a használatot gyorsító technikai táblákban, illetve a mentésfájlokban jelen lévő másolataik helyet foglalnak a számítógépen. A lemezterület az ezredforduló óta egyre jelentősebb költséget jelent. A rögzített tárolók mind drágábbak és drágábbak, ami az emelkedő kereslettel magyarázható, mely a digitalizált kép és audiovizuális fájlok számának és átlagos méretének állandó növekedésnek köszönhető. A szöveges adatok tárhely igénye, különösen tömörített állományokban, a digitális fotókhoz és főként az audiovizuális anyagokhoz képest elenyésző. Ami rendszercserénél a metaadatok közül az új rendszerbe nem kerül át, annak kimentésére majd őrzésére gondot kell fordítani. Ez idő-, és munkaigényes, de maguk a metaadat rekordok – szöveges tartalmak lévén – nem foglalnak sok helyet. Ha az adatokat annak érdekében, hogy kevesebb helyet foglaljanak, szerkesztjük, módosítjuk, információt veszíthetünk, de a visszakeresés könnyebb, az őrzés olcsóbb, a kezelés, gondozás kevesebb munkát igényel. Ha csak egy adatbázismentést teszünk el az áttöltés előttről, és magát az alkalmazást, melynek keretében az adatok értelmezhetők, magyarázhatók, nem őrizzük meg, akkor később a visszafejtés, értelmezés lesz nehezebb. Ha az alkalmazást magát is megőrizzük, az több helyet foglal el, viszont a rendszer felszámolásával nem veszítünk el információt.

A hosszútávú őrzés költségeinek növekedése várhatóan a jövőben is meghaladja majd az infláció mértékét. Legalább két tekintetben a kereslet még sokáig felül fogja múlni a kínálatot. Az egyik terület az adathordozók, az erre alkalmas hardver eszközök köre. Ebben a vonatkozásban ugyanis a termelésnek objektív korlátjai vannak, mindenekelőtt a szükséges anyagok tekintetében. Másrészt az IT munkaerő iránti igény is növekszik, mert az informatikai tevékenység, és ezen belül a digitális dokumentumok hosszútávú őrzése ugyanis igen munkaerőigényes tevékenység.

A hosszútávú megőrzés tehát igen költséges, ezért, ha úgy is gondoljuk, hogy használati szempontból nézve jó volna minden analóg dokumentumot digitalizálni, fel kell tenni a kérdést, hogy lehetséges és elengedhetetlenül szükséges is? A döntések meghozásánál

6 Tripathi, Sneha: Digital preservation: some underlying issues for long-term preservation. = Library Hi Tech News, Number 2. 2018. 8-12p.

7 MAYER, Adelheid: Workshop „Software-Lösungen zur Langzeitarchivierung und Repositorien-Verwaltung aus Anwendersicht“ = Wien, 21. April 2016.

8 Trevar D. Riley-Reid: The hidden cost of digitization – things to consider= Collection Building Volume 34 Number 3 · 2015 (89–93. p.)

persze különbséget kell tenni a használati és megőrzési célú digitalizáció között. Összegezve elmondható, hogy egy intézmény, vagy intézményi hálózat igen gyakran választásra kényszerül, hogy mit digitalizáljon. Ez a választás elsősorban nem is az állományvédelmi célú digitalizációnál merül fel, (ebben az esetben ugyanis a védelemre leginkább rászoruló tartalmakkal kell természetesen kezdeni a digitalizációt) hanem a használati célú digitalizációnál érdemes válogatni. A kiválasztás elsődleges alapja a használati célú digitalizáció esetében, annak felmérése mellett, hogy lesz-e mód szerzői jogi és személyiségi jogi korlátozás nélkül szolgáltatni az anyagot, az állomány iránti érdeklődés lehet. A dokumentum iránt megnyilvánuló érdeklődés vizsgálatakor persze igen bizonytalan területre tévedünk. Amikor egy dokumentumot valamilyen módon elhelyezünk, az iránta való érdeklődés sokszor már az elhelyezés következménye lesz. Egy hagyományos könyvet is gyakrabban vesznek kézbe az olvasók, ha a szabadpolcon van, mintha a raktárban, netán egy külső raktárban porosodik. Nincs ez másképpen a digitális tartalmakkal sem, itt sem mindegy, mennyire vannak „szem előtt”. Ezért is fontos, hogy a használatát a lehető legegyszerűbb módon és a lehető legátfogóbban mérjük. Fel kell tennünk a kérdést mindenekelőtt, a keresettség szempontjából számít-e a könyv kora, vagyis a kiadása óta eltelt idő? Igaz-e, hogy a legújabb könyvek a könyvtárakban a legkeresettebbek?

### A katalógusrekord mint adat

Az integrált könyvtári rendszerekben az elmúlt fél évszázadban katalógusrekordok milliói jöttek létre. Ezek a rekordok bibliográfiai adatbázisok, számítógépes leltárkönyvek mellett a tartalmi feltárás, illetve a besorolási állományok építése során létrejött tudásszervező rendszerek<sup>9</sup> részeinek is tekinthetők. Az elmúlt két évtizedben az adatbázisokban a katalógusrekordokhoz kapcsolt teljes tartalmak tömege is egyre növekszik. A számítógépes katalógusokat a könyvtárak napi működésük során használják, maga az adatbázis szüntelenül módosul, alakul. A katalógusba került információk tekintélyes része annak pillanatnyi használati értékén túl is feltétlenül érdemes az utókor figyelmére, ezért a könyvtári számítógépes adatbázisokban tárolt metaadat - és teljes tartalom rekordokra kiemelt figyelmet kell fordítanunk a hosszútávú megőrzés szempontjából is.<sup>10</sup>

Az integrált rendszerekben tárolt adatok eddig is fontos forrásai voltak a könyvtárhasználati statisztikáknak. Az adatok kinyerésére és szolgáltatására valamennyi ismert rendszer fel volt és fel van készítve. A legfontosabb statisztikák, melyeket a rendszerekben tárolt adatok alapján el szoktak készíteni:

- Kölcsönzések és egyéb könyvtárhasználati események száma, olvasók beíratása, tagságuk meghosszabbítása,
- Legolvasottabb művek és szerzők,
- Beszerzésre fordított költség

A rendszerekben tárolt metaadatok elemzésére egyre nagyobb az igény a könyvtári világban, mind a vezetők, mind a szakfelügyelet, mind pedig a fenntartók, illetve finanszírozók részéről. Az így kinyert adatokat a beszerzési politika, a gyarapítás és a finanszírozás, de a feldolgozás, a tartalmi és formai feltárás terén is jól lehet tervezéshez, munkaszervezéshez használni. Azokat az adatokat, melyek kiadására a rendszereket azok fejlesztői nem készítették fel, közvetlenül az adatbázisból kell kinyerni. Az adatokat elemzésre fel kell készíteni, ennek

---

9 KOS- Knowledge Organisation System

10 Lengyel Monika, Simon András: „Rendszerváltás” a hazai könyvtárakban – divathullám vagy kényszer? = Tudományos és Műszaki Tájékoztatás 50. köt. 8.sz. (2003.) 313–317 p.

lépései, a megfelelő adatok kiválasztása, adattisztítás, különféle átalakítások módosítások, adatok összevonása, és a statisztikai analízis.<sup>11</sup>

Az adatelemzést, főként annak egységesített elvégzését nehezíti, hogy integrált könyvtári (és múzeumi) rendszerek az ezredforduló előtti évtizedekben nagy heterogenitást mutattak az adattárolás tekintetében. Az adatok lehetnek szabványos, vagy nem szabványos szerkezetű adatfájlokban, "raw device"-okon, tárolhatták azokat adatbáziskezelők is. A könyvtári rendszerek fejlődésével, a szolgáltatási szintek egységesedésével, a szolgáltatási színvonal emelkedésével spontán szabványosodás zajlott le, így az elmúlt fél évszázadban az adattárolás és szolgáltatás tekintetében alábbi trendek érvényesülését állapíthatjuk meg:

- Az adattárolásban uralkodóvá váltak az SQL alapú relációs adatbáziskezelőkre épülő rendszerek,
- Az adatbáziskezelők és az operációs rendszerek fejlődésével az adattárolás biztonságossá vált, a különféle műszaki problémák, vis maior-ok eredményeképpen keletkezett adatvesztések pedig ritkábbak lettek,
- Uralkodóvá vált a szabványos adatszerkezet, a csereformátumok (MARC, LIDO) egyben belső tárolási formátumokká is váltak,
- Az adatbázisok általában már relációs szerkezetűek, a tárolásban a rendszerek egyre inkább elkerülték redundáns adatfolyamok (stringek) tárolását,
- A múzeumi igényeket követve megjelent a változáskövetés, az adatok módosulását a rendszerek egy része már naplózza és felületen is megjeleníti,
- A metaadatokhoz tartozó teljes tartalmakat nem az adatbáziskezelőben magában, hanem külső fájlrendszerben helyezik el, és az alkalmazás által olvasható módon tárolják az adatbázisban a hozzáférési adatokat.

A szabványosodás, és a közös keresőfelületek, aggregátori adatbázisok kiszolgálása az adatok egységesedését is eredményezte. A metaadatokkal való megfelelő ellátás kulcsfontosságú a távolabbi jövőben várható keresőkérések kiszolgálása kapcsán, ezért a dokumentumok archiválásakor erre kiemelt figyelmet kell fordítani.<sup>12</sup>

Az elmúlt fél évszázadban a könyvtári használatban kialakult az integrált könyvtári rendszer standardja az alapvető modulokkal: katalogizálás, kölcsönzés, folyóirat, gyarapítás. Ezek metaadatokkal dolgoztak, a gépi rendszerekben az első időkben csak a metaadatok voltak digitalizálva, illetve keletkeztek digitálisan. A mai integrált könyvtári rendszerek, és az ezek mintájára létrejött integrált gyűjteménykezelő rendszerek a metaadatok – mindinkább szemantikus – hálózatára épülnek, a teljes tartalmak ezek kiegészítései.<sup>13</sup>

Akatalogusokba felvitt metaadatok tehát szemantikus hálóba szerveződtek és az adattartalom struktúrájára vonatkozó információk (technikai adatok, és a rekord létrehozására és tárolására vonatkozó történeti adatok) fontos elemei lettek az adatrekordoknak.<sup>14</sup>

---

11 Trail, Stacie, Patrick, Martin: Core Concepts and Techniques for Library Metadata Analysis= Code4lib. Issue 52, 2021-09-22. <https://journal.code4lib.org/articles/16078>

12 Kopácsi, Sándor and Hudak, Rastislav and Ganguly, Raman: Implementation of a classification server to support metadata organization for long term preservation systems / Implementierung Eines Klassifikationsservers Für Metadatenorganisation Für Langzeitarchivierungssysteme = Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, vol. 70, n. 2, 2017. 225–243 p.

13 Kouzari, Elia–Stamelos, Ioannis: Process mining applied on library information systems: A case study= Library and Information Science Research 40 (2018) 245–254 p.

14 Zong Peng: Cloud-based service for access optimization to textual big data. = School of Informatics, Computing, and Engineering, Indiana University, 2018. 141 p.

Technikailag az integrált rendszerek katalógusaiban az adatféleségek szerint általában négy féle rekordot különböztetnek meg:

- Bibliográfiai,
- Példány,
- Besorolási,
- Média.

Az elterjedt könyvtári rendszerek moduljai emellett az alábbi katalóguson kívüli rekordokat kezelik:

- Gyarapítás – rendelési csomagok, tételek, pénzügyi tranzakciók,
- Kölcsönzés – Olvasók és kölcsönzési műveletek,
- OPAC – keresések és találatok,
- Adatbázis tranzakciók rekordjai.

A tranzakciós rekordról (melyeket kutatásaim során elemzek) az alábbi ismérveket állapíthatjuk meg:

- Az alkalmazás állítja elő működés közben
- Tartósan őrzik a rendszerek
- A különféle integrált rendszerek más jellegű rekordokat hoznak létre
- Az adatbázis működésére vonatkoznak: (keresés, módosítás, létrehozás, törlés, kölcsönzés, előjegyzés stb.).

Jelen tanulmányban a kölcsönzési tranzakciós rekordok elemzésével foglalkozom. Az ebben tárolt főbb információk:

- Könyv (bibliográfiai és példány rekord) azonosítására alkalmas információk,
- Olvasó azonosítására alkalmas információk,
- Könyvtáros azonosítására alkalmas információk,
- Tranzakció dátuma,
- Kölcsönzés lejárat dátuma,
- Visszahozatali dátum,
- Hosszabbítás dátuma,
- Tranzakció fizikai helye.

Az adatok egy élő, napi használatban levő adatbázisban állandóan módosulnak. Amennyiben a változásokat nem követjük, információt veszíthetünk.

A változáskövetés módszerei:

- Az adatbázisok különféle korú változatainak tárolása,
- Az adatváltozások verziókövetése,
- A módosítások ellátása időbéléggel, és a módosítási lépések tárolása.<sup>15</sup>

Az integrált könyvtári rendszerekben ma még a rekord alapú adatkezelési szemlélet uralkodik. A katalógus rekordok a fizikailag létező, vagy a már digitálisan keletkezett dokumentumokat írják le. A katalógus rekordok terén ezért megkülönböztetünk metaadat rekordokat és média rekordokat. A média rekordok, a teljes tartalmak képi, hangzó, videó, háromdimenziós, vagy teljes szöveges dokumentumok lehetnek, és egy vagy több fájlból állhatnak. A múzeumi gyakorlatban leíró és adminisztratív metaadatrekordnak nevezik, a könyvtári gyakorlatban pedig bibliográfiai és példány rekordnak ezeket az egységeket. A katalógus rekordokat technikai metaadat rekordok egészítik ki, melyek a rekord keletkezésére, módosulására, az integrált rendszerben

---

<sup>15</sup> Andreas Rauber, Ari Asmi, Dieter van Uytvanck, and Stefan Pröll: Identification of reproducible subsets for data citation, sharing and re-use. =TCDL Bulletin, year 2016, vol. 12.





betöltött szerepére nézve tartalmaznak információkat. A katalógus rekordok tartalmát a korszerű rendszerekben a lehető legnagyobb mértékben értéklisztákból építik fel. Az értékliszták mögött a rendszer jellegétől és az alkalmazott szabványoktól függően besorolási rekordállományok lehetnek. A katalógusrekord korszerű szerkezete alatt tehát valójában már több rekord alkotta szemantikus hálót kell értenünk. A katalógus rekordok és a kölcsönzési tranzakciós rekordok mellett egyéb rekordok is vannak az integrált könyvtári rendszerekben. A személyes, pénzügyi, gyarapítási adatokat tartalmazó hosszú távú őrzése azonban nem szükséges.<sup>16</sup>

### Primer adat – szekunder adat

A hosszú távú őrzés tervezése során fontos feladat a primer és szekunder adatok meghatározása, az egyes adatbázisokban megkülönböztetésük, szétválasztásuk, és mentés esetén elkülönített kezelésük lehetőségeinek a megteremtése. Csak a primer adatokat érdemes ugyanis tartósan megőrizni. Szekunder adatnak kell tekintenünk mindent, ami primer adatokból származtatott információt tartalmaz. A szekunder adatok esetében csak a forrásul szolgáló primer adatot, illetve a létrehozásuk módját kell megőrizni. Primer adatnak kell tekintenünk mindent, ami nem tartalmaz más adatokból származtatott információt. A szekunder adatokat a rendszerek a használat megkönnyítése, meggyorsítása érdekében állítják elő, rendszerint a keresést, böngészést, megjelenítést támogatják az integrált rendszerek felhasználói felületein.

A szekunder adatok legnagyobb tömegét jelentő csoportját az indexállományok alkotják. Ezeket nem szükséges beemelni a hosszú távú megőrzési tervbe. Mivel általában a keresést és a böngészést támogatják egy adatbázisban, ezért funkciójuk az igények változásával módosulhat. Előállításuk, képzésük módja pedig, jóllehet akár jelentős programozói tudást is igényelhet, alapvetően a rendszer működéséhez tartozik, nem pedig az adatbázisban tárolt adatállományhoz magához, ezért a rendszert alkotó programcsomaggal célszerű együtt tárolni, addig ameddig az alkalmazást magát, mintegy szoftvermúzeumi tárgyként meg akarják őrizni.<sup>17</sup>

A könyvtári adatra, nemcsak metaadatként, hanem szolgáltatott teljes tartalom forrásaként is kell tekintenünk. Ez akár függetlenül is tárolható, elkülönítve a könyvtári katalógustól, ez esetben csak belső hivatkozással mutatunk rá. Ezt nevezi Siegmüller hibrid könyvtárnak (hybride Bibliothek).<sup>18</sup> Viszont abban az esetben is, ha a könyvtári rendszer a tartalmakat nem a saját adatbázisából, hanem külső tárolókról szolgáltatja, tartalmi indexek épülnek rá, melyek a tárolt teljes szövegeket indexelik. Ezen tartalmi indexek tartós megőrzésére nincs szükség. A kapcsolt teljes tartalmak hozzáférését, és a számítógépes katalógus metaadat rekordjaival való kapcsolatát kell hosszú távon megőrizni.

A lemásolt (szekunder) adatoknak tehát, melyeket Riley-Reid nem tart szükségesnek beemelni a hosszú távú megőrzési tervbe, lényeges csoportját képezik az indexállományok. Ezek a keresést és a böngészést támogatják egy adatbázisban. Különösen nagy hozzáadott értéket képviselnek azok a szóindexek, melyek az adatbázisba betöltött teljes szöveges tartalmakra készültek el. Az indexhez hozzáadott szellemi értéket,

16 Simon András: Integrált Könyvtári Rendszerben tárolt tranzakciós rekordok felhasználása a könyvtárhasználat statisztikai elemzésére = Tudományos és Műszaki Tájékoztatás 66. évf. 12. sz. 2019. 683–693 p.

17 D. Riley-Reid, Trevar: The hidden cost of digitization – things to consider = Collection Building, Vol. 34. Number 3. · 2015. 89–93. p.

18 Siegmüller, Renate: Verfahren der automatischen Indexierung in bibliotheksbezogenen Anwendungen = Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 214 Heft. 2007. 7. p.



azok tervezése, szakmai és műszaki kialakítása jelenti. Ez a munka semmiképpen nem automatizálható, mindenképpen egyedi kialakítás eredménye. Az indexek lefuttatására felhasznált gépidőt mint hozzáadott technikai ráfordítást, kell megemlítenünk. Az adatbázis általában nem csak az indexállományt magát tárolja, hanem az index elkészítési technikáját is, függvények, tárolt eljárások, illetve triggerok formájában. Az indexek kialakítását leíró algoritmusok természetesen számos rendszerben az adatbáziskezelőn kívül programkódban, vagy külső konfigurációs állományokba is lehetnek. Ez esetben az adatbázis mentés hosszú távú őrzése nem is elég minden a rendszer integritását érintő működési információ tartós meglétének biztosítására.

A szakirodalom áttekintése során képet alkothattunk arról, mekkora terhet jelent az emlékeztetintézmények számára a lezárt számítógépes katalógusok őrzése. A nem lezárt, folyamatosan működő rendszerek adatbázisainak esetében pedig a változások előtti állapotok helyreállítható formában való megőrzése. Láthatjuk, mekkora terhet jelent a csatolt teljes tartalmak tárolása, és hosszútávú őrzése. Emellett szempontokat kaphatunk egy, az adatbázisok időbeli növekedését, tehát a metaadat-, és médiarekordok számának és méretének növekedési trendjeit számba vevő vizsgálathoz is.

## A vizsgálat bemutatása

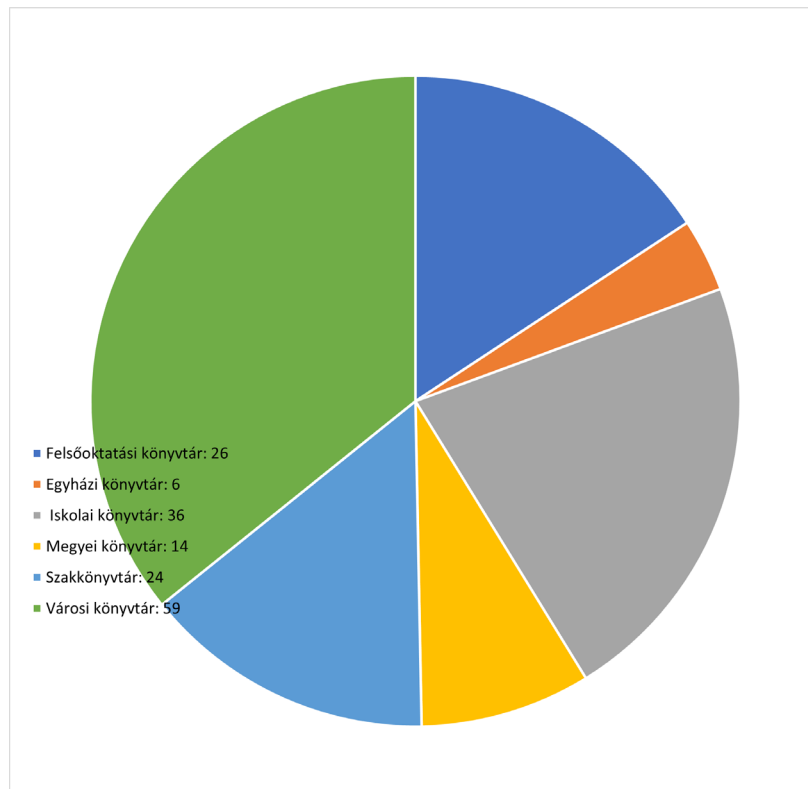
Az ELTE Irodalomtudományi Doktori Iskola PhD hallgatójaként doktori kutatási témám az integrált könyvtári és múzeumi rendszerek műszaki vizsgálata, a közgyűjteményi digitalizálás költségeinek felmérésére szolgáló módszertan kidolgozása céljából. A könyvtári és múzeumi informatikával foglalkozó, mindenekelőtt az integrált gyűjteménykezelő rendszerek magyar nyelvterületen piacvezető szállítójaként ismert Monguz Kft alkalmazottjaként feladataim közé tartozik munkaadóm ügyfeleinek adatbázisaiban karbantartási munkák végzése. Ilyen módon lehetőségem van a – kivétel nélkül SQL alapú - adatbázisokhoz való közvetlen hozzáférésre. Jól ismerem ezek szerkezetét, és az SQL nyelvvel is immár három évtizede dolgozom. Ügyfeinkkel kötött szerződéseink szellemében katalógusaikból kinyerhetünk anonimizált adatokat szakmai, statisztikai célból. Az általunk forgalmazott rendszerekre, a Qulto – Corvina, és a Qulto – Huntéka rendszerre egyaránt igaz, hogy szabványos adatszerkezetű, és valamennyi megszokott könyvtári modullal rendelkezik. Az ezekben a rendszerekben tárolt adatok közül csak az látható, amit a felhasználói felületek az olvasóknak, illetve a rendszert kezelő könyvtárosoknak elérhetővé tesznek. A technikai adatok esetleg nincsenek rejtve a rendszergazdák és rendszerkönyvtárosok előtt, de a rendszergazdák közvetlenül csak a saját adatbázisukat látják, és a rendszerek szerkezete, még ha szabványokra épül is, igencsak bonyolult. Nehéz ezekből hasznos információt kinyerni, én tehát itt egyedülálló lehetőséget ragadhatok meg, és igyekszem a könyvtáros szakma számára ezt a lehetőséget ki is használni.<sup>19</sup>

A kutatás során 360 a rendszereinket használó könyvtár adatbázisát volt módom megvizsgálni, távoli eléréssel. Valamennyi ügyfelünk szerverére nem tudtam eljutni, mert nem minden felhasználónk tárolja a mi szervereinken adatbázisát. Aki pedig saját szerveren használja valamelyik rendszerünket, sokszor nem ad kézenfekvő hozzáférési módot. Tekintettel arra, hogy a vizsgálat a könyvtári adatbázisok egyenként való elérésével amúgy is hallatlanul időigényes volt, sok könyvtár ebből a vizsgálatból időszüke miatt így kimaradt. Főként igaz volt ez a szakkönyvtári adatbázisokra, különösen a múzeumi segédkönyvtárak és a hatósági könyvtárak gépi katalógusaira, ezek ugyanis az átlagosnál nagyobb arányban vannak telepítve számunkra nehezen elérhető szerverekre.

---

<sup>19</sup> Kiszl Péter, Simon András: Aspects of the Long-Term Preservation of Digitized Catalogue Data: Analysis of the Databases of Integrated Collection Management Systems = PDT&C 2021; 50(2): 51–64. <https://doi.org/10.1515/pdct-2021-0007>

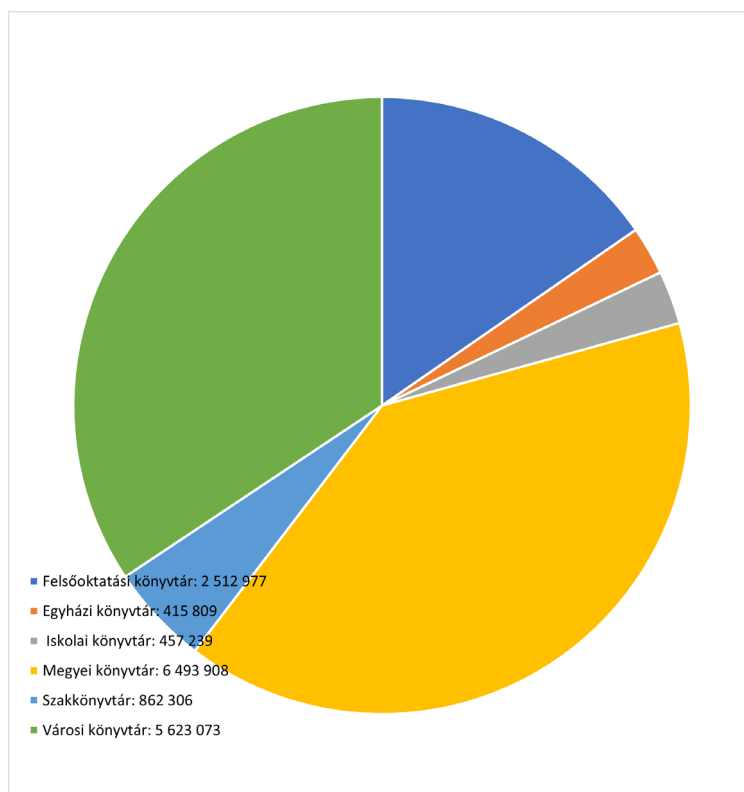
Az 1. ábrán a vizsgálatban végül részt vevő könyvtári adatbázisok számát mutatom be, könyvtártípus szerinti bontásban.



1. ábra: Vizsgálatban végül részt vevő könyvtári adatbázisok

Az ezen tanulmányban ismertetett vizsgálat a kölcsönzési tranzakciós rekordokkal foglalkozott, ezért feltétele volt a vizsgálatba való bevonásnak, hogy a könyvtár egyáltalán kölcsönözzön, és használja az integrált rendszer kölcsönzési modulját. Mivel egy bizonyos időtávlatot szükségesnek tartottam figyelembe venni, ezért szempont volt, hogy legalább 2017 óta üzemszerűen kölcsönözzön a mi integrált rendszerünkkel (lejárt kölcsönzési tranzakciós rekordokat ugyanis nem konvertálunk, a vizsgálat tehát később a mi rendszerünkre átállt könyvtárakban nem folytatható le). A vizsgálati eredmények elemzése 2020 elejével lezárult (a Covid-korlátozások óriási hatással jártak a könyvtárak működésére, ezek elemzése külön előadás és további kutatások feladata<sup>20</sup>). Fontos szempont volt továbbá, hogy a kiadás éve a bibliográfia rekordban legyen jelen, és aritmetikai műveletek számára bizonyuljon hozzáférhetőnek. (Ez az adat ugyanis a könyvtári leírási szabványoknak megfelelően nem feltétlenül pusztán egy évszámból áll.) Hosszabbításokat és a kölcsönzési időtartamot - a kölcsönzési politika helyi különbségei miatt - nem vettem figyelembe. A 2. ábrán a vizsgálatban végül részt vevő könyvtárakban a vizsgált időszakban lefolytatott összes kölcsönzési tranzakciók számát mutatom be, könyvtártípus szerinti bontásban.

<sup>20</sup> A kérdéssel a Bobcatss 2022-es debreceni konferenciáján tartott előadásomban foglalkoztam: „Examination of the effects of the pandemic on the digital transformation of the libraries, for the purpose of analysing the database transaction records of the Integrated Library Management Systems” Bobcatss 2022 Debrecen.



2. ábra: Vizsgált kölcsönzési tranzakciók száma

A kutatás során a megadott adatokat egyforma SQL parancsokkal kérdeztem le az összes vizsgálható adatbázisban. A kutatás során a könyvtárak, vagy azok tagintézményei által önállóan használt integrált könyvtári rendszerek adatbázisait, vagy az adatbázisokban elkülönült leőhelyekhez tartozó állományegységeket vizsgáltam.

A kinyert számítási eredmények egy külön - általam létrehozott - adatbázis munkatáblájába kerülnek, amelyben az egyes intézményeket már csak egy kód azonosítja. A kód információt tartalmaz a könyvtár típusáról, de a könyvtár maga, mint intézmény a kódból már nem visszafejthető. A könyvtárak, tagintézmények, illetve állományegységek és a kódok közötti megfeleltetést egy általam könnyen megjegyezhető, mások számára ellenben megfejthetetlen jelleggel oldottam meg. Minden könyvtár vagy tagintézmény, illetve állományegység esetében az ahhoz tartozó lekérdezéseket, és az eredményeket ellenőrzési célból külön fájlban tároltam el. A fájl nem tartalmaz utalást az intézmény nevére, vagy az adatbázis elérésére, azonosítási célból kizárólag a kódot és a jellegét tartalmazza. Az adatbázisban számítások végezhetőek, illetve CSV outputok nyerhetőek ki az adatbáziskezelő eszközeivel. A vizsgálat tárgyát a könyvtárak és könyvtári részlegek képezték attól függően, hogy önálló számítógépes katalógussal, vagyis önálló adatbázissal és arra épülő könyvtári rendszerrel rendelkeznek-e.

A vizsgálat tárgyát képezte ellenőrzési célból:

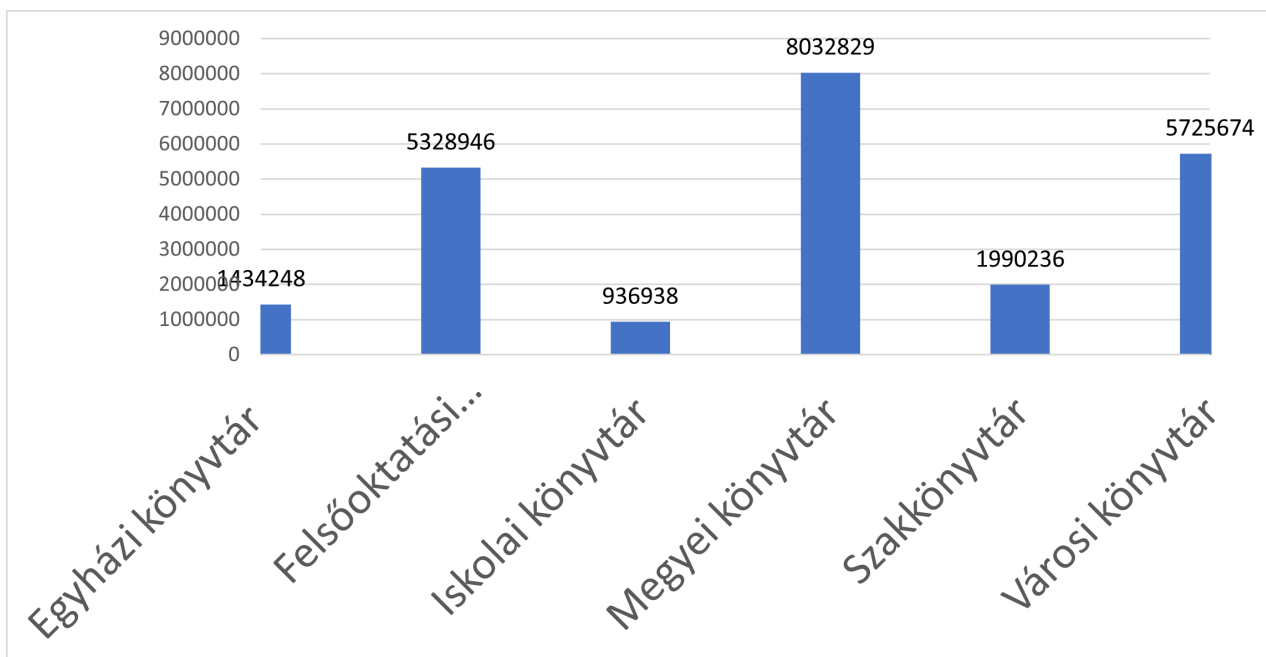
- Összes példányszám: a vizsgált könyvtárban az összes könyvhöz tartozó teljes példány állomány.
- Összes (a vizsgálat pillanatában) kölcsönözhető kategóriájú példány száma.
- Összes kölcsönözhető kategóriájú példány 2018-ban és 2017-ben. Erre az adatra ugyancsak a Magyar Könyvtárak Statisztikai adatait tartalmazó, a Könyvtári Intézet weboldalán található táblázat adataival való összevetéshez volt szükség.
- Összes könyv típusú, és példánnyal rendelkező bibliográfiai rekord száma.
- Összes (a vizsgálat pillanatában) kölcsönözhető példánnyal rendelkező bibliográfiai



rekord száma.

- Teljes kölcsönözhető könyvállomány átlagos életkora, és a teljes könyvállomány átlagos kora.
- Kölcsönzött könyvek átlagos kora kölcsönzés éve szerinti bontásban.
- Összes kölcsönzési tranzakció
- Összes kölcsönzési tranzakció évenkénti bontásban

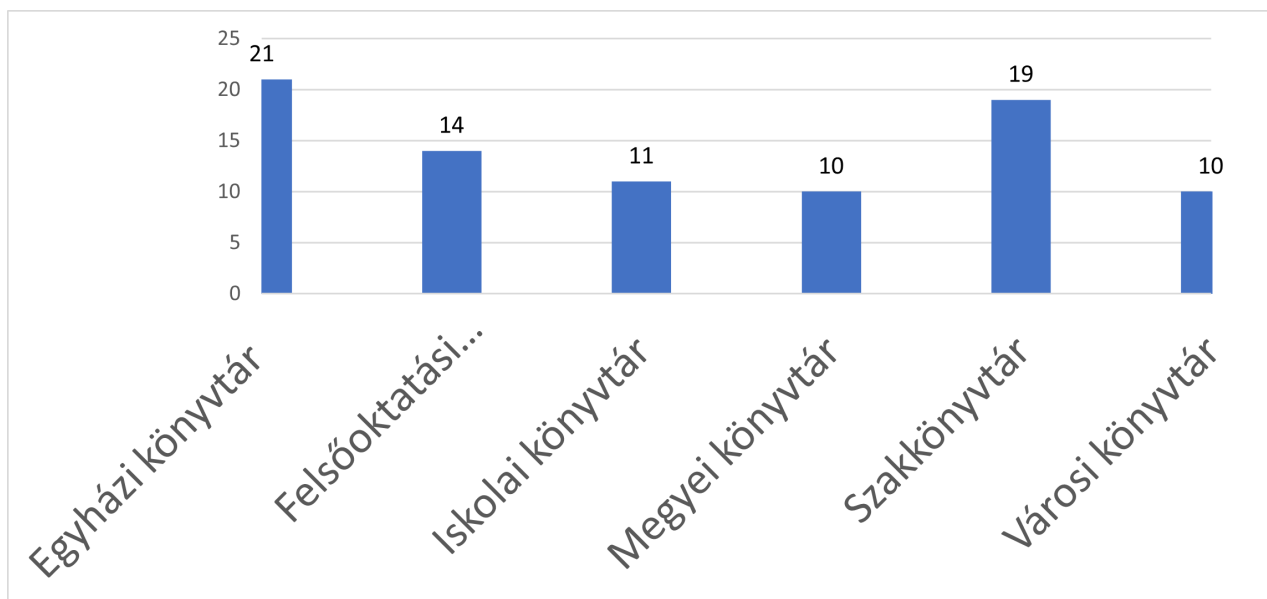
A 3. ábrán a vizsgálatban végül részt vevő könyvtárakban a kölcsönözhető példányok számát mutatom meg, intézménytípusok szerinti csoportosításban



3. ábra: Kölcsönözhető példányok száma intézménytípusok szerinti csoportosításban

Érdekes lett volna a kölcsönzéseket ETO főcsoport szerint vizsgálni, de nem minden könyvtár használ ETO-t, főként nem a teljes állományra, ezért ebből értelmezhető adattömeg kinyerését nem tudtam megvalósítani. Ugyancsak érdekes lett volna a kölcsönzött könyvek mellett a raktári kéréseket is vizsgálni, de elektronikus raktári kérés eddig csak néhány ügyfelünkénél valósult meg.

A vizsgálat fő célja a Kölcsönzött könyvek átlagos kora volt. Ezt a 4. ábrán mutatom be.



4. ábra: Kölcsönzött könyvek átlagos kora

A kölcsönzött könyvek átlagos kora tekintetében a könyvtárak típusonkénti bontása hozott értelmezhető eredményt. Ezzel szemben az alábbi tényezők egyike sem okozott észrevehető különbséget:

- Ország.
- Nemzetiség.
- Országon belül földrajzi elhelyezkedés.
- A kölcsönözhető állomány átlagos kora.
- Könyvtár mérete (állomány nagysága vagy az olvasók száma).
- Kölcsönözhető könyvek átlagos kora,
- Használt integrált rendszer (Huntéka vagy Corvina).

#### További lehetőség, a médiatartalmak használatának vizsgálata

A könyvtári állomány használata, sőt általában a könyvtárhasználati, olvasási, sőt kultúrafogyasztási szokások terén jelentős átalakulás figyelhető meg. A hagyományos, analóg tartalmak olvasása könyvtártípusonként változóan enyhén csökkenő vagy enyhén stagnáló tendenciája mellett megfigyelhető a fokozódó érdeklődés a könyvtárak által szolgáltatott digitális dokumentumok iránt.<sup>21</sup>

Az újabb anyag egy része már eleve digitálisan keletkezik, és ezek őrzésére és gyűjtésére az anyag sérülékenysége miatt a közgyűjteményeknek kiemelt figyelmet kell fordítaniuk. Emellett folyamatos a hagyományos állomány digitalizációja, részben az eredeti anyag védelme, részben a teljes szöveges kereshetőség biztosítása miatt. Emellett igen sok katalógus tovább szolgáltatja, más gyűjtemények katalógusát, tartalmait is.

Természetesen vannak ellentétes irányú hatások. A felhasználók sokszor még vonakodnak a digitális tartalmak használatát igénybe venni, főként, ha azokhoz sem férnek hozzá azonnal,

<sup>21</sup> Simon András: Examination of the effects of the pandemic on the digital transformation of the libraries, for the purpose of analysing the database transaction records of the Integrated Library Management Systems” Bobcatss 2022 Debrecen.



fáradtság, illetve szerzői jogi megfontolásoknak eleget tevő korlátozás nélkül vagy ingyen.<sup>22</sup> Természetesen minél több tartalom érhető el a könyvtárak online katalógusaiból azonnal és közvetlenül, minden fajta jogi és adminisztratív akadály nélkül, akkor a gyorsaság és kényelem a teljes szöveges keresés lehetőségével kiegészítve ezt a szolgáltatástípust egyre népszerűbbé, megszokottabbá teszi majd.<sup>23</sup>

Természetesen az elektronikus könyvtári szolgáltatások útjában a legnagyobb akadály a szerzői jogi védelemből fakadó korlátozás jelenti. Mind a nemzeti kormányok, mind az Európai Unió illetékes hatóságai számára elsőrendű cél, hogy a közgyűjtemények a lehető legtöbb digitalizált, vagy már digitálisan keletkezett dokumentumot korlátozás nélkül szolgáltatathassák az érdeklődő közönségnek.<sup>24</sup> Természetesen a korlátozott hozzáférés (vízjel, snippet-view vagy időbélyeg használata, a letöltést, sőt a képernyőkép másolását megakadályozó speciális formátumok, pl. E- Pub használata, csak meghatározott hálózatból, munkaállomásról, vagy azonosított olvasó számára megengedett megtekintés) egyaránt jobb, mint a fájl olvasásának, megnézésének teljes tiltása. Persze minden ilyen megoldás alkalmazása munkaidő- és ennek megfelelően költségigényes, ráadásul az összes ilyen alkalmazás hamar elavul, ugyanis mind a jogi szabályozás, mind a felhasználói eszközök, mind pedig a tárolási és szolgáltatási módszerek igen gyorsan változnak.

A fennálló nehézségek és akadályok ellenére is egyértelmű, hogy az igény a digitális tartalmak használata tekintetében nagy, és ez az igény a jövőben folyamatosan növekedni fog. Lényeges, hogy ezen tartalmak használatát is mérni lehessen, így adva teljes képet a könyvtárhasználati statisztikákhoz. Ezek az információk a további dokumentumok digitalizálása esetén is hasznos támpontot adnak a digitalizálandó példányok kiválasztásához. A médiatartalmak felhasználók által való megtekintésének, a megtekintési események számbavételének azonban jelentős nehézségei vannak. Amennyire kézenfekvő a fizikai példányokat érintő kölcsönzési események számolása, hiszen az integrált könyvtári rendszerek már négy évtizedre ezen folyamat végzésére és a kölcsönzési esemény nyomon követésére vannak tervezve, az elektronikus tartalmak használatának figyelésére igazából csak közvetett eszközökkel rendelkezünk.

Egyrészt már maga a dokumentum, mint entitás megállapítása sem olyan egyértelmű feladat egy digitális dokumentumnál, mint a hagyományos fizikai példány esetében. A digitális dokumentum, egy szellemi termék (mű – „work”) manifesztációja. Megjelenése, időben változhat, alakulhat, sokszor nem is egy szerző alkotása, hanem keveréke a szellemi tulajdonnak, és a közösségi alkotásnak. Mindannak, ami szerzői jogi védelem alatt áll, annak használatát valamilyen módon korlátozni kell, és ez egy digitális dokumentum esetében, még a könyvtárban is annak veszélyét vetíti előre, hogy kevésbé lesz, éppen a korlátozás, vagy a finanszírozási kényszerűség miatt használva.<sup>25</sup> Ebben az esetben a digitális és analóg könyvtári dokumentumok használtságát nehéz ugyanazon elvek mentén összevetni.

22 Niggemann, Elisabeth; Fernau, Michael; Schwens, Ute: Bibliothek ohne Bücher? = Dialog mit Bibliotheken 29 Bd. 1 Heft 2017.

23 Meyer, E.T.; Eccles, K.: The Impacts of Digital Collections Early English Books Online & House of Commons Parliamentary Papers =London University of Oxford, Oxford University Consulting, Oxford Internet Institute 2016

24 Tóth Máté: A digitalizálás trendjei az Európai Unióban= Tudományos és Műszaki Tájékoztatás 62. évf. 1. sz. 2015.

25 Lior Fink, Jianhua Shao, Yossi Lichtenstein: The ownership of digital infrastructure: Exploring the deployment of software libraries in a digital innovation cluster=Journal of Information Technology. July 23, 2020. <https://doi.org/10.1177/0268396220936705>



A médiatartalmak használatának a mérésének nehézségeit jól szemlélteti, hogy milyen sok megoldás merül fel a használatnak nemcsak a figyelésére, de magának a használati eseménynek az azonosítására is. A használati esemény azonosításának és regisztrálásának lehetséges módjai:

- A weboldalhoz intézett kérések számolása, az ugyanarról az IP címről érkező kérések lehetséges szűrésével,
- Sikeres kérések számolása valamilyen visszacsatoló kód segítségével,
- Weboldal logfile elemzése,
- Oldalak megnyitásának figyelése az oldalhoz illesztett szkript segítségével.

Bármelyik megoldást választjuk is, a zavaró tényezők: (robotok, crawlerek, közös keresők kiszolgálása céljából végzett tömeges letöltések) kiszűrése minden esetben első számú feladat.<sup>26</sup>

A használat mérésénél hasznos javaslatokat olvashatunk William Denton "On Two Proposed Metrics of Electronic Resource Use" című tanulmányában. Ebben két módszert javasol a digitális tartalmak használatának mérésére: A „Users per mille (UPM)” módszer azt vizsgálja, hogy ezer potenciális felhasználóból hány tekintette meg az illető dokumentumot egységnyi idő alatt. Az „Interest factor (IF)” azt méri, mennyire tartották érdekesnek a felhasználók az illető dokumentumot, tehát hányan térnek vissza, hogy a dokumentumot újra megtekintsék. (A szerző feltételezi, hogy elsőre csak átfut az ember egy interneten ingyenes és akadály nélkül elérhető dokumentumot).<sup>27</sup>

Jól látható tehát hogy nagyon fontos lenne a dokumentumok iránti érdeklődés pontos mérése, és a használat esetében mind a hagyományos papír alapú könyvek, mind pedig az elektronikus könyvek használatát mérni kell, illetve kellene. De míg az előbbiek esetében a kölcsönzéseket az integrált könyvtári rendszerek regisztrálják, hiszen a kölcsönzések gépi nyomon követése a kezdetektől fogva az egyik alapvető feladatuk, az elektronikus dokumentumok használatát már nehezebb pontosan mérni. Hacsak itt nem a szerzői jogi és kiadói érdekek védelmében „elektronikus könyv kölcsönzésről” – vagy helyesebben digitális tartalmak ellenőrzött rendelkezésre bocsátásáról – van szó, mely pontosan ugyanúgy mérhető, mint a kölcsönzés. A webes felületen akárcsak belső hálózaton korlátozás nélkül elérhetővé tett dokumentumok használata ezzel szemben éppen olyan nehezen nyomon követhető, mint a könyvtári olvasótermi és raktári példányok helybeni használata.

Szerzői bemutatkozás:

Simon András 1969-ben született Budapesten. 1993-ban végzett az ELTE BTK-n, történelem-könyvtár szakon. 1992 óta dolgozik magyar könyvtárakban a könyvtárgy automatizálás területén. 1999 és 2002 között a MOKKA rendszergazdája. 2002-től alvállalkozóként, majd alkalmazottként működik együtt az MTA SZTAKI ASZI osztályával, majd a szegedi Monguz Kft.-vel a Huntéka integrált könyvtári rendszer tervezése, fejlesztése, ügyféltámogatása és forgalmazása területén. 2019-től az ELTE Irodalomtudományi Doktori Iskola PhD hallgatója.

András Simon was born in 1969 in Budapest. He finished his studies in 1993 in History and Librarianship at the Faculty of Arts and graduated from ELTE. He has been working in Hungarian libraries since 1992, for the scope of library automatization. Between 1999 and 2002 he was the system librarian of the Hungarian Shared Cataloguing System. Since 2002

---

26 COUNTER Code of Practice – Maintenance and development of the Code of Practice=COUNTER committee dir. Lorraine Estelle. <https://media.springernature.com/full/springer-cms/rest/v1/content/10988228/data/v2>

27 William Denton: On Two Proposed Metrics of Electronic Resource Code4lib. Issue 52, 2021-09-22. <https://journal.code4lib.org/articles/16087>



he worked as subcontractor and later as employee for the MTA SZTAKI, then for Monguz Ltd. from Szeged in the field of design, development, customer support and distribution of the Huntéka integrated library management system. Since 2019, he is the PhD student of ELTE Institute of Library and Information Science.

## Bibliográfia

COUNTER Code of Practice - Maintenance and development of the Code of Practice= COUNTER committee dir. Lorraine Estelle. <https://media.springernature.com/full/springer-cms/rest/v1/content/10988228/data/v2>

D. Riley-Reid, Trevar: The hidden cost of digitization – things to consider = Collection Building, Vol. 34. · Number 3. · 2015. 89–93. p. <https://doi.org/10.1108/CB-01-2015-0001>

Fröhlich, Susanne: Der Weg ist das Ziel! Planung und Umsetzung digitaler Archivprojekte. = Mitteilungen der VÖB Vol. 66. Nr. 1 2013 (133–144p.)

Kiszl Péter, Simon András: Aspects of the Long-Term Preservation of Digitized Catalogue Data: Analysis of the Databases of Integrated Collection Management Systems = PDT&C 2021; 50(2): 51–64. <https://doi.org/10.1515/pdct-2021-0007>

Kopácsi, Sándor and Hudak, Rastislav and Ganguly, Raman: Implementation of a classification server to support metadata organization for long term preservation systems / Implementierung Eines Klassifikationservers Für Me-tadatenorganisation Für Langzeitarchivierungssysteme = Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, vol. 70, n. 2, 2017. 225–243 p. <https://doi.org/10.31263/voebm.v70i2.1897>

Kouzari, Elia - Stamelos, Ioannis: Process mining applied on library information systems: A case study= Library and Information Science Research 40 (2018) 245–254 p. <https://doi.org/10.1016/j.lisr.2018.09.006>

Lengyel Monika, Simon András: „Rendszerváltás” a hazai könyvtárakban – divathullám vagy kényszer? = Tudományos és Műszaki Tájékoztatás 50. köt. 8.sz. (2003.) 313–317 p.

Lior Fink, Jianhua Shao, Yossi Lichtenstein: The ownership of digital infrastructure: Exploring the deployment of software libraries in a digital innovation cluster= Journal of Information Technology. July 23, 2020. <https://doi.org/10.1177/0268396220936705>

Mayer, Adelheid: Workshop „Software-Lösungen zur Langzeitarchivierung und Repositorien-Verwaltung aus Anwendersicht“ = Wien, 21. April 2016. <https://doi.org/10.31263/voebm.v69i1.1407>

Meyer, E.T.; Eccles, K.: The Impacts of Digital Collections Early English Books Online & House of Commons Parliamentary Papers =London University of Oxford, Oxford University Consulting, Oxford Internet Institute 2016 <https://doi.org/10.2139/ssrn.2740299>

Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland nestor – Network of Expertise in Long-Term Storage of Digital Resources = Printversion 2.0 beim Verlag Werner Hülsbusch, Boizenburg

Niggemann, Elisabeth; Fernau, Michael; Schwens, Ute: Bibliothek ohne Bücher? = Dialog mit Bibliotheken 29 Bd. 1 Heft 2017.

- Oßwald, Achim: Langzeitsicherung digitaler Informationen durch Bibliotheken. =Zeitschrift für Bibliothekswesen und Bibliographie. 2014, Vol. 61 N. 4-5.  
<https://doi.org/10.3196/18642950146145194>
- Rasmussen, Casper Hvenegaard : Is digitalization the only driver of convergence? Theorizing relations between libraries, archives, and museums Journal of Documentation; Bradford Köt. 75, Kiad. 6, 2019 <https://doi.org/10.1108/JD-02-2019-0025>
- Rauber, Andreas , Ari Asmi, Dieter van Uytvanck, and Stefan Pröll: Identification of reproducible subsets for data citation, sharing and re-use. =TCDL Bulletin, year 2016, vol. 12.
- Siegmüller, Renate: Verfahren der automatischen Indexierung in bibliotheksbezogenen Anwendungen = Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 214 Heft. 2007. 7. p.
- Simon András: Examination of the effects of the pandemic on the digital transformation of the libraries, for the purpose of analysing the database transaction records of the Integrated Library Management Systems” Bobcatsss 2022 Debrecen.
- Simon András: Integrált Könyvtári Rendszerben tárolt tranzakciós rekordok felhasználása a könyvtárhasználat statisztikai elemzésére = Tudományos és Műszaki Tájékoztatás 66. évf. 12. sz. 2019. 683–693 p.
- Simon András: Rendszerváltás 2.0? Előnyök és hátrányok, vélt és valós törvényszerűségek az integrált gyűjteménykezelő rendszerek életciklusában. = Tudományos és Műszaki Tájékoztatás 68. évf. 2021. 3–4. sz.
- Tóth Máté: A digitalizálás trendjei az Európai Unióban= Tudományos és Műszaki Tájékoztatás 62. évf. 1. sz. 2015.
- Trail, Stacie, Patrick, Martin: Core Concepts and Techniques for Library Metadata Analysis= Code4lib. Issue 52, 2021-09-22. <https://journal.code4lib.org/articles/16078>
- Trevar D. Riley-Reid: The hidden cost of digitization – things to consider= Collection Building Volume 34 · Number 3 · 2015 (89–93. p.) <https://doi.org/10.1108/CB-01-2015-0001>
- Tripathi, Sneha: Digital preservation: some underlying issues for long-term preservation. = Library Hi Tech News, Number 2. 2018. 8–12p.  
<https://doi.org/10.1108/LHTN-09-2017-0067>
- Wenrui Yan, Jie Yao, Qiang Cao, And Changsheng Xie: A Rack-based Optical Storage System with Inline Accessibility for Long-Term Data Preservation= ACM Transactions on Storage, Vol. 14, No. 3, Article 28. November 2018.
- William Denton: On Two Proposed Metrics of Electronic Resource Code4lib. Issue 52, 2021-09-22. <https://journal.code4lib.org/articles/16087>
- Zong Peng: Cloud-based service for access optimization to textual big data. = School of Informatics, Computing, and Engineering, Indiana University, 2018. 141 p.

## Az OSZK Webarchívum nemzetközi kapcsolatai

Németh Márton  
Országos Széchényi Könyvtár  
Digitális Bölcsészeti Központ  
Digitális Filológiai és Webarchiválási Csoport  
ORCID: [0000-0003-1864-8107](https://orcid.org/0000-0003-1864-8107)

Az alábbiakban rövid összefoglalást nyújtunk az OSZK webarchívumának külföldi kapcsolatairól, illetve a nemzetközi munkánk során elért eredményekről. A tanulmányban öt különféle témakörrel ejtek szót. Elsőként a webarchiválási projektünk megalapozásában kulcsszerepet játszó, dániai szervezésű online tanfolyamról ejtek szót. Ezt követően bemutatom az OSZK részvételét a webarchiválásért felelős nemzetközi konzorcium (IIPC) szakmai tevékenységeiben, különös tekintettel az oktatási munkacsoportra. A harmadik téma a közép-európai webarchívum együttműködési kezdeményezés (CEWA), melyet a 2021-ben megrendezett „404 Not Found” online workshop első napjának délutánján tartott kerekasztal-beszélgetésén alapoztuk meg. Felvázolom az együttműködés tervezett területeit, s a várható eredményeket. A webarchívumok kutatási célú felhasználásának kérdéseire fókuszál a 2023 nyaráig tartó WARCNET projekt, mely a dániai Aarhusi Egyetem koordinálásában valósul meg. Negyedik témaként röviden összefoglalom a projekt főbb területeit, eddigi eredményeit, s a jelenleg is zajló tevékenységekről is áttekintést nyújtok. A zárótémakör keretei között pedig az amerikai Internet Archive magyar állományára alapozódó teljesszövegű kereső és portál fejlesztéséről adok áttekintést.

A Magyar Internet Archivum pilot projektjének formálódása közben, egyben a saját webarchiválással kapcsolatos kutatások megalapozásaként is elsődleges feladat volt feltérképezni az internet archiválás főbb nemzetközi szereplőit, tanulmányozni a különböző nemzeti webarchiválási modelleket a hasznosítható jó gyakorlatok szempontjából. A projekt előrehaladása közben az OSZK csatlakozni tudott a webarchiválásban érdekelt szereplőket összefogó International Internet Preservation Consortium-hoz, aktívan bekapcsolódtunk több munkacsoport tevékenységébe is. Emellett felvettük a kapcsolatot közép-európai partnerintézményekkel későbbi intenzívebb együttműködési formák megalapozása céljából.

### 1. A NetLab kurzus

A dániai Aarhus egyetemén az országos DIGHUMLAB kutatási infrastruktúra részeként működő, Niels Brügger professzor vezette NetLab kutatócsoport online egyetemi kurzusokat kínál PhD hallgatók, ill. a témában érintett szakemberek, kutatók számára a webarchiválás témakörében. 2017 őszén a webarchiválási projektünk indulásakor első alkalommal – angol nyelven – nemzetközi keretek között is meghirdették a kurzust, ráadásul a részvétel ingyenes volt. A szervezési-moderátori feladatokat Asger Harlaug, a kutatócsoport helyettes vezetője látta el. A szakmai tartalom mellett nagyon fontos volt a részvétel abból a szempontból is, hogy áttekintést kaptunk arról, hogy miként kell egy webarchiválással foglalkozó tanfolyamot felépíteni, megszervezni és gondozni. Szembesülhettünk a távoktatási felületen zajló kollaboratív munka módszertani jellegzetességeivel is. Külön növelte a kurzus értékét, hogy az OSZK kísérleti projektjével nagyjából egyidőben indult belga webarchiválási program munkatársaival is alkalmunk nyílt megismerkedni, a dán szakemberekkel zajló közös munka mellett. Mély benyomást tett a résztvevőkre a konstruktivista pedagógiai háttér, a segédanyagok, a kurzus menetrendje, a közös munkafelület. Emellett azonban alapvetően a résztvevők szakmai irányultságán, lelkesedésén múlt, hogy milyen tartalommal lehetett

megtölteni a közös munkát. Végül soron a tanfolyam sikerének talán kisebbik hányadát adta az előzetes felkészülés, a tananyag kidolgozása. A sikeresség a tartalom megfelelő felvázolásán, a résztvevőkkel zajló osztálytermi, illetve virtuális kommunikáció minőségén állt vagy bukott.

Az online szeminárium egyéni felhasználó név/jelszó párossal védett, Moodle-alapú távoktatási felületen zajlott. Itt publikálták a szervezők a kurzus segédanyagait, a feladatok szövegeit, és itt volt lehetőség a válaszok megfogalmazására és az egymással történő társalgásra is a kurzus-fórum keretében. Ez a felület csak a tanfolyam ideje alatt volt elérhető, de annak végeztével lehetőségünk volt az összegyűlt anyag lementésére.

Öt fő terület került megtárgyalásra. Először annak felmérése zajlott, hogy a résztvevőknek milyen kapcsolatuk van a webarchiválással és milyen elvárásaik fogalmazódnak meg a kurzussal szemben? Már ebből is következtetni lehetett arra, hogy a konstruktivista pedagógiai alapon szerveződő szeminárium minden egyes félévben más és más profillal bír, a tartalmát erőteljesen befolyásolja a résztvevők érdeklődési köre, szakmai irányultsága.

A második téma körül zajló társalgás éppen arra irányult, hogy az általános érdeklődési kör megfogalmazása után mindenki elmondja, hogy miért hasznos számára webarchiválással foglalkozni és felvázoljon valamiféle, a kurzus további teljesítéséhez kapcsolódó konkrét munkatervet vagy kutatási tervet.

A harmadik témakörben keresni kellett három, az egyéni érdeklődési területhez tartozó honlapot, melyek legalább egy éve üzemelnek már, s áttekintést kellett adni ezek archiválásának kihívásairól. Mintaként lehetett használni az Internet Archive vagy bármelyik nemzeti webarchívum nyilvánosan elérhető mentéseit. Itt lehetett igazán először megtapasztalni a webarchiválásra jellemző technikai-szakmai kihívások sokaságát és szinte áttekinthetetlen összetettségét. A kurzus résztvevői a saját munkájuk során szembesülhettek azzal, hogy egyáltalán milyen keretek között lehet webarchiválásról beszélni, hogy az archiválás sikerességének kritikus tényezője a megfelelő szoftveres eszköztár kiválasztása, mely sokszor csupán a tartalom szerkezetének alapos feltérképezése után végezhető el. Persze szembenézhattünk a korlátozott szakmai-emberi erőforrások természetével is, hogy a webarchiválási tevékenység szempontjából kritikus jelentőségű a rendelkezésre álló kapacitások hatékony felhasználásának művészete is.

A negyedik témakörben a kutatási terület (gyűjtőkör) előzetes felmérése, valamint a megadott idő alatt egy adott személy által elvégezhető munka kereteinek operatív meghatározása után meg kellett fogalmazni a kutatási célunkhoz kötődő további honlapok összeválogatásának stratégiáját. Ezt követően ki kellett választani a megfelelő szoftveres háttérrel, kísérleti aratásokat végezni, majd mini riportban összegezni a tapasztalatokat (Németh, 2018c). Számomra ez bizonyult a leghasznosabb pontnak, hiszen a belgák is és mi is elő tudtunk állni konkrét tartalmak kísérleti mentéseinek tapasztalataival, s meg tudtuk beszélni egymással az aratáshoz használt szoftverek előnyeit, hátrányait, illetve értékeltük a végeredményt. A társalgásba több ponton bekapcsolódtak a dán nemzeti webarchívum munkatársai is, akik saját tapasztalataikkal színesítették a kommunikációt. Az volt a fő cél, hogy mindenki választ kapjon a kérdéseire a konkrét kihívások kapcsán, s ezeket hasznosítva tudja tovább építeni a saját projektjét.

Mindezek után pedig az ötödik témakör keretében összesítettük a kurzussal kapcsolatos tapasztalatainkat, s ezzel zárult a szeminárium, melynek elvégzéséről a résztvevők bizonyítványt is kaptak.





## 2. Az IIPC konzorcium és az abban zajló magyar tevékenység

Az egyes nemzeti webarchiválási modellek (könyvtárakban, levéltárakban, egyetemeken, kutatóintézetekben) tulajdonképpen arra válaszul jöttek létre, hogy a web fejlődése egyetlen szervezet számára már követhetlenné vált archiválási szemszögből. Ezek egy része eleve csak korlátozott időre szólt, másokat pedig egy idő után újragondoltak és -terveztek, így már több második-generációs webarchívummal is lehet találkozni. A nemzeti szereplők, illetve az Internet Archive egymást kiegészítő tevékenységei adják meg a webarchiválási szolgáltatások teljességét. Jelenleg mintegy 45 nemzeti webarchívum létezik valamivel több mint harminc országban, ugyanis egyes helyeken az államalkotó nemzeti népcsoportoknak külön archívumuk van. Ezeket a webarchiválási tevékenységeket az IIPC nemzetközi konzorcium fogja közös keretbe, melynek az Országos Széchényi Könyvtár is tagja lett 2018-ban. Ennek révén hatékony tudásmegosztási, kompetenciafejlesztési és szakmai együttműködési csatornák válnak elérhetővé számunkra is. A konzorcium 2003-ban alakult meg a Francia Nemzeti Könyvtár kezdeményezésére 12 alapító taggal. Jelenleg a 45 országból összeálló tagság könyvtárakat, levéltárakat, múzeumokat, illetve a webarchiválás területén érdekelt civil és piaci szereplőket is magában foglal. A konzorcium alapvető küldetése az internetes tartalmak archiválásához és hosszú távú biztonságos megőrzéséhez szükséges tevékenységek koordinálása és támogatása<sup>1</sup>.



1. ábra: Az IIPC honlapja

Az IIPC tevékenységét korábban a British Library-ből koordinálták, manapság már önálló titkárságot működtetnek. Az operatív irányítást a tagintézmények képviselőiből választott Végrehajtó Bizottság látja el. A közös munka nagy része tematikus munkacsoportokban

1 <https://netpreserve.org>



zajlik, melyek jelenleg a tartalomfejlesztéssel, a hosszú távú megőrzéssel, illetve az oktatási kérdésekkel foglalkoznak. Az OSZK képviselőjében ez utóbbi (2017 végén alakult) munkacsoportban, illetve a 2019-ben indult kutatási munkacsoportban veszünk részt. Külön figyelmet érdemel a 2019-2020-ban létrehozott repozitórium<sup>2</sup>, melynek keretében fokozatosan teszik közzé az IIPC szakmai rendezvényein elhangzott előadások prezentációit, a kutatási jelentéseket és cikkeket. Ennek használata szinte azonnal megkerülhetlenné vált a webarchiválás tanulmányozásához. Az aktuális kutatási területek jóval rövidebb átfutási idővel bukkannak itt fel, mielőtt tudományos publikációkban jelennének meg. Képet kaphatunk arról, hogy mely témák állnak a gyűjteményfejlesztéssel, a technikai fejlesztésekkel, illetve a webarchívumok kutatási célú felhasználásával foglalkozó szakemberek érdeklődésének középpontjában.

Alkalmam nyílt részt venni az IFLA 2017. évi kongresszusán a lengyelországi Wrocławban, ahol önálló szekció foglalkozott az IIPC szervezésében a webarchiválással. Itt történt meg 2017 augusztusában az első személyes kapcsolatfelvétel az IIPC tagjaival, vezetésével. Tanácsokat kaptam tőlük a webarchiválási kísérleti projektünk megalapozásához, még mielőtt hivatalosan taggá váltunk volna. Az IIPC tagságunk lényeges eredménye az a kutatás-fejlesztési együttműködés, melyet szerződéses munkatársunk, Vitéz Gábor alakított ki a Dán Nemzeti Könyvtár munkatársaival, elsősorban Thomas Egense-vel, aki a SolrWayback nevű teljes szövegű kereső és megjelenítő eszközt fejleszti. A szoftver tesztelésébe és egyes hibák javításába mi is be tudtunk segíteni. A dán fejlesztők az IIPC 2018-as, mi pedig az OSZK képviselőjében az IIPC 2019-es közgyűlésén számoltunk be e közös munka eredményeiről, mely a konzorcium vezetésének körében is pozitív visszhangot keltett.

Csatlakozásunk évében még nem volt módunk a IIPC konzorcium éves közgyűlésén, illetve az ahhoz kapcsolódó konferencián való részvételre, mivel ezekre Új-Zélandon került sor. Lehetőségünk nyílt viszont egy öt perces videó összefoglalót készíteni a rendezvényen részt vevők számára, bemutatva magunkat és eddigi eredményeinket. 2019-ben már részt vállaltunk a Zágrábban rendezett közgyűlés szervezésében, a programbizottság munkájában is. A kísérleti projektről tartott rövid összefoglaló mellett a dán partnereinkkel közös fejlesztési munkákról adtam elő, a SolrWayback és a saját fejlesztésű SolrMIA keresőkről, és azok találati listáinak relevánsabbá tételéről metaadatok segítségével. Sikeres együttműködést alapoztunk meg a Web Curator Tool webarchiválási munkafolyamatokat segítő keretrendszer új-zélandi és holland fejlesztőivel, a rendezvényen tartott workshop alatt. 2019 őszén ennek tapasztalatai alapján munkatársaimmal egyeztetve eljuttattam kérdéseinket a programcsomag fejlesztése kapcsán a partnereinkhez, így közvetlen befolyással bírnak, informatikus kollégáinkkal együtt, a fejlesztési irányok meghatározására. Megteremtettem az alapjait a közép-európai webarchívumok közötti együttműködésnek is, melynek konkrét szervezeti kereteit közös fejlesztési programokkal, rendezvényekkel a 2021. év során terveztük partnereinkkel együtt meghatározni.

---

2 <https://digital.library.unt.edu/explore/partners/IIPC/>

## SOLRWAYBACK

The screenshot shows the SolrWayback search interface. The search query is 'Múzeumkert'. The results are filtered by domain to 'mnm.hu'. The search results show a single entry for 'MÚZEUMKERT | MAGYAR NEMZETI MÚZEUM' with a score of 121.20181. The entry details include: Type: html, web page @ mnm.hu; Date: 20/05-2020; Url: https://mnm.hu/hu/muzeumtortenet/muzeumkert. The highlighted content is: 'AkadálymentesFacebookInstagramFree WiFi Bejelentkezés LátogatókLátogatók Megközelítés Nyitva tartás A Múzeumkert házirendje'. Below the text, there are four image thumbnails showing various views of the museum building and grounds. The interface also includes navigation buttons for 'Previous 20' and 'Next 20', and a 'View data fields' button.

2. ábra: A Solrwayback kereső találati oldala

Az IIPC konzorciumon belül 2017 végén alakult meg az oktatási és képzési munkacsoport<sup>3</sup>. Első projektjük keretében kérdőíves vizsgálatot végeztek. Arra irányult a kutatás fókusz, hogy az egyes országokban milyen típusú szervezetek, mekkora létszámban foglalkoznak webarchiválással, valamint ezeknek a szakembereknek milyen igényeik vannak az oktatás, a szakmai továbbképzés terén. Az IIPC tréning munkacsoportja a kérdőív tapasztalatait felhasználva először felmérte azokat a fajta oktatási tevékenységeket, melyekkel a leghatékonyabb segítséget tudják majd nyújtani tagintézményeik, illetve a lehető legszélesebb körű szakmai célcsoportok számára. Fokozottan szeretnének építeni az egyes országokban eddig felhalmozott tapasztalatokra. Most először mérték fel azt is, hogy az egyes országokban milyen oktatási és képzési tevékenységek zajlanak, milyen keretek között, s hányan vesznek azokban részt. A munkacsoport online értekezletein lezajlott eszmecsere során kiderült, hogy a legtöbb oktatási tevékenység igen szűk körű, csak az egyes archiválási tevékenységek ellátásához feltétlenül szükséges készségek és képességek átadására fókuszáló oktatási tevékenység nyert teret. Olyan széleskörű, általános tananyagok kidolgozása<sup>4</sup> lett tehát az elsődleges cél, melyek rugalmasan beilleszthetők bármiféle szervezet fizikai jelenlétre épülő, illetve távoktatási programjának a keretei közé, megbízható szakmai tartalmat kínálva fel a webarchiválással kapcsolatos különböző kompetenciák megalapozásához. A későbbiekben kerülhet sor haladóbb szintű webkuratori, illetve informatikai ismeretekhez kötődő témakörökhöz illeszkedő tananyagfejlesztésre is.

3 <http://netpreserve.org/about-us/working-groups/training-working-group/>

4 <http://netpreserve.org/web-archiving/training-materials/>

Aszakmai tartalmak meghatározása mellett a módszertani felépítés is egyenrangú fontossággal bír a tananyagok fejlesztése során. A moduláris konzisztencia megteremtése különféle szakmai csoportokban, szinteken válik ellenőrizhetővé. A csoport általános álláspontjaként lett elfogadva, hogy habár fontos a fejlesztésre fordított időtényező optimalizálása is, de a megfelelő minőségű oktatási anyag rendelkezésre bocsátása élvez elsődlegességet. A tananyagok mögött álló koncepció megvilágítására a zágrábi IIPC konferencia alatt videó interjúk is készültek a webarchiválás oktatásának különféle aspektusairól. A munkacsoportban összeállítottunk egy mintegy száz tételből álló kérdéslisztát. Az interjúalanyok számára a felvétel készítői (előzetes konzultációt követően) ezekből szemezgettek különféle témakörökhöz illeszkedve. Az interjúkat a tananyagok közrebocsátásával egyidőben publikálták<sup>5</sup>.

Az elkészült oktatási anyag az IIPC különféle fórumain keresztül kerül terjesztésre. Egy önálló GitHub csatorna létrehozását<sup>6</sup>, az alapvetően informatikusok által szoftverek közzétételére szolgáló felület használatát több tényező is indokolja. A verziókövetéssel jól nyomon követhetők az oktatási anyagban bekövetkező változások. A kiegészítő vizuális és szoftveres elemek is könnyedén beilleszthetők a tananyag mellé. A felhasználói visszajelzések, korrekciós javaslatok precízen rögzíthetők és dokumentálhatók.

Az első nemzetközi kurzus a Digital Preservation Coalition és az IIPC Training Working Group szervezésében 2020 októberében került megszervezésre, a webarchiválásban még nem jártas IIPC tagintézmények munkatársai részére. Itt kerültek először kipróbálásra a kifejlesztett tananyagok konkrét oktatási tevékenység közben is. A tanfolyam a korábban említett modulok anyagát kínálja, válaszokat keresve arra, hogy miért is kell a webet archiválni, hogyan néz ki egy webarchívum, milyen koncepciók szerint, s milyen technológiai eszközökkel lehet webarchiválást szervezni és lebonyolítani, és hogy milyen kutatói és egyéb felhasználói igények kielégítésére kell megoldásokat találni.

### 3. Közép-európai együttműködés

A webarchiválási tevékenység elindulásától kezdve személyes találkozásokon alapuló kapcsolatokat sikerült kiépíteni a cseh, a szlovák és az osztrák webarchívumok munkatársaival is. A pozsonyi Egyetemi Könyvtárban több előadást is tartottunk, illetve információcserét folytattunk a házigazda intézmény és a Cseh Nemzeti Könyvtár webarchiválásért felelős munkatársaival. Az Osztrák Nemzeti Könyvtárban pedig egynapos szakmai tanulmányúton vettünk részt munkatársaimmal együtt, hogy megismerjük az ott zajló webarchiválási tevékenységet. Sikeres szakmai rendezvényt is szerveztek fennállásuk 10. évfordulóján, ezen is alkalmam nyílt részt venni, ahol elsősorban a német nyelvű országok webarchiválási projektjeiről lehetett széleskörű áttekintést szerezni. A horvát kollégákkal a 2019-ben megrendezett zágrábi IIPC konferencián kerültünk közelebbi kapcsolatba.

Az OSZK által ötödik alkalommal megszervezett „404 Not Found” online konferencia és workshop első napján (2021. november 23.) bemutakoztak a közép-európai országok nemzeti szintű webarchívumai Ausztriából, Csehországból, Horvátországból, Szlovéniából és Szlovákiából, valamint a Lengyel Állami Levéltár is ismertette a terveit (ott 2022 elején indult el a webarchiválási tevékenység előkészítése). Az angol nyelvű prezentációk elérhetőek a rendezvény weboldalan<sup>7</sup>.

A délutáni kerekasztal beszélgetésen ismertettük, majd megvitattuk a résztvevőkkel azt a javaslatunkat, hogy ezek a webarchívumok működjenek szorosabban együtt, egy Central

---

5 <http://netpreserve.org/web-archiving/iipc-training-video-case-studies/>

6 <https://github.com/iipc/training>

7 <https://webarchivum.oszk.hu/404-not-found-workshop-2021-november-23-24/>



European Web Archives (CEWA) nevű projekt keretei között, tekintettel a régió közös kulturális és történelmi örökségére, ami az interneten is leképeződik. A kezdeményezést mindenki pozitívan fogadta, de természetesen ez még nem jelenti az anyaintézmények hivatalos csatlakozását a CEWA projekthez, jelenleg az előkészítés zajlik. Terveink között szerepel egy többnyelvű portál és metaadatkereső megalkotása. A portálon a résztvevő nemzetek nyelvén bemutatkozna mindegyik webarchívum. Erre a célra felhasználnánk az IIPC által támogatott „Dark and Stormy Archives” nevű rendszert<sup>8</sup>, mellyel látványos és a közösségi médiában is megosztható módon hozhatók létre válogatások archivált webtartalmakból. Beépítenénk továbbá a szintén IIPC-s forrásból fejlesztett technológiát, a „Bloom filter” módszer használatát a webarchívumokban való közös kereséshez<sup>9</sup>. Ennek az a lényege, hogy a korábban erre a célra használt Memento protokollnál nagyságrendekkel gyorsabban lehet megállapítani, hogy egy, a felhasználó által keresett URL cím megvan-e valamelyik webarchívumban. (A hozzáférés utána természetesen attól függ, hogy egy nyilvános vagy egy csak helyben használható gyűjteményrészben van-e az adott URL-ről lementett fájl.) Ezt a megoldást a horvát webarchívumban tesztelték tavaly, így az ottani kollégáknak már van ezzel tapasztalatuk. Fontos lenne az egymás közötti tapasztalatcsere és a jó gyakorlatok átvétele a webarchiválással kapcsolatos tevékenységek terén (különös tekintettel az informatikai fejlesztésekre). Közös rendezvények és együttes publikációs tevékenységek szervezése is a tervek között szerepel. Az együttműködő intézmények megoszthatnák egymással saját URL listáikat a regionális jelentőséggel bíró témák, események kapcsán és felmerülhet az archiválási feladatok részleges összehangolása is. Emellett közös projektekkel lehetne pályázni az IIPC forrásaira is.

#### 4. A WARCNET projekt

A projekt eredeti célkitűzése egy olyan kutatási hálózat megalkotásában rejlett, mely segít értelmezni a nemzeti, illetve nemzetközi felsőszintű web doménekből archivált anyag felépítését, rétegzettségét történetét, tartalmi összefüggéseit (ilyenek például a .hu vagy a .com tartományok), illetve a globális hatású társadalmi események európai szintű webes lenyomatát igyekszik feltárni, mint a webarchívumokban tárolt digitális kulturális örökség fontos szeletét. Számos társadalmi krízis, konfliktus, vitatéma tanulmányozásához elengedhetetlenek a webes források. A projekt közös platformot nyújt a webarchívumokban tárolt anyagot kutatóknak, illetve a gyűjtemények működtetésében főszerepet játszó közgyűjteményi szakembereknek és informatikusoknak is. A csapat tagjai egyéni kutatók, a nemzeti könyvtári webarchívumok és az IIPC munkatársai is. A fő cél, hogy az archívumokban tárolt anyagokban rejlő összefüggésekre, a web történetére vonatkozó kutatásokra sok szempontból lehessen tekinteni (pl. kommunikáció- és médiatudományi, történettudományi, digitális bölcsészeti nézőpontokból).<sup>10</sup> Az 1990-es évektől a számítógépes világháló fokozatosan a társadalmi kommunikáció, a tudásátadás, illetve tágabb értelemben a közélet megkerülhetetlen médiumává lett, természetéből adódóan nemzetközi térben. Az itt zajló diskurzusok szinte egy pillanat alatt a múlt részévé válnak, ezek feltárása, elemzése nagyon fontos kutatási tevékenységként jelenhet meg. Az Internet Archive, illetve a nemzeti könyvtárak és levéltárak javarészt már évtizedek óta gyűjtik a webes múlt fontos emlékeit. A webarchívumok kutatása azonban számos esetben speciális kompetenciákat igényel, messze nem olyan triviális, mint a hagyományos levéltári, könyvtári kutatási tevékenységek végzése. Fontosságukat általában alábecsülik tudományos szempontból, s miután a nemzeti

8 <https://oduwsdl.github.io/dsa/>

9 <https://netpreserve.org/projects/bloom-filters/>

10 <https://cc.au.dk/en/warcnet/about>

szintű kutatási háttér mindenütt nagyon korlátozott mértékű Európában, ezért is vetődött fel ötletként egy olyan hálózat létrehozása, melynek segítségével egymásra találhatnak a szakemberek, s közös és magas színvonalú tevékenységeket tervezhetnek az erőforrások hatékony felhasználása révén.<sup>11</sup> Így talán sikerül elérni azt a kritikus tömeget, melynek köszönhetően a webarchívumokra irányuló kutatások széleskörű létjogosultságot nyernek.

Sajnos a járvány következtében alig sikerült személyes részvételen alapuló találkozókat tartani, melyekből legalább évente kettőt terveztek a projekt időtartam alatt (döntően a londoni, a luxemburgi és a dán aarhusi egyetemek bázisán). Az online térben is hatékonyan lehet azonban szervezni a különféle tevékenységeket. Fontos cél lett volna az egymás közgyűjteményi, intézményi, illetve kutatási tevékenységeinek jobb megismerését szolgáló, maximum öt napos tanulmányutak finanszírozása, mely az ismert helyzetben szintén háttérbe szorult. A projekt időtartamát azonban sikerült kitolni az eredetileg tervezett 2020-2021-es periódustól egészen 2023 tavaszáig. A finanszírozási háttérrel a Dán Kutatási Tanács biztosítja, a koordinátori szerepet a webtörténetírás atyjának is tekinthető Niels Brügger, az aarhusi egyetem professzora látja el. A projekt működésének három legfontosabb fóruma az operatív irányítást végző szűk testület, a plenáris munkaértekezletek, illetve a szekciókban folyó tevékenységek. A projektirányító bizottság tagjai a vezető kutatók mellett a hálózat tagjaként szereplő fiatal kutató, egy informatikus, illetve egy webarchívumban dolgozó szakember. Őket a hálózat tagjai választották meg az első értekezleten. A munkaértekezleteken az általános projektmenedzsment teendők elvégzése mellett a szekciók beszámolnak a folyamatban lévő tevékenységekről, eredményekről, és keynote előadások keretében teret adnak átfogóbb, a webarchiválást esetleg csak érintőlegesen érintő témáknak, illetve olyan előadásoknak is, melyek új kutatási irányokat alapozhatnak meg. Minden értekezleten ingyenesen részt vehetnek tanácskozási joggal nem rendelkező érdeklődők is, előzetes regisztrációt követően. Közülük számosan (mint e sorok írója is, aki a második – online – értekezleten kapcsolódott be a munkába) a projekt tagjává is válhattak. A tagok számára a fizikai térben megrendezhető értekezleteken való részvétel ingyenes, a szállásköltség is biztosítva van (sajnos eddig csak egy ilyen, részben személyes részvételen alapuló hibrid értekezlet került lebonyolításra Aarhusban, 2021 őszén). 2022-ben a járvány múltával – egy nyári londoni találkozót követően – a projekt záróértekezletére ősszel Aarhusban kerül majd sor.

A projekt hat szekcióban különféle szakmai témák köré szerveződik,<sup>12</sup> e szekciók munkája szervesen kapcsolódik egymáshoz. Az egyes számú szekció egyrészt nemzeti, illetve nemzetközi doméneket (webtereket) vizsgál (webhelyek száma, mérete, multimédia elemek használata, a hiperlinkekkel összefogott hálózat sajátosságai). Másrészt a felhasználókra összpontosít (milyen jellemzőik vannak és milyen webhelyeket használnak). Harmadrészt pedig vizsgálat alá vonja a globális közösségi hálózatok működését és társadalmi hatásait is. A második munkacsoport a különféle kiemelt események webes megjelenésére, s az ebből adódó kutatási, elemzési lehetőségekre összpontosít. A harmadik szekció a webarchiválással összefüggő módszertani teendőkre fókuszál. A negyedik munkacsoport a webarchiválás során létrejövő és a kutatás során használható adattömeg kezelésére jött létre. Sok szó esett benne a szerzői jogi, illetve adatvédelmi dilemmákról és jó gyakorlatokról a webarchívumokhoz való hozzáférés kapcsán. Az ötödik munkacsoport voltaképpen a negyedikben vázolt kutatási adatkezelés egy speciális összetevőjére összpontosít: a webarchívumokból származó kutatási adatok szabványos megosztására. A hatodik munkacsoportban a kutatási hálózat hosszabb

---

11 <https://cc.au.dk/en/warcnet/the-warcnet-team>

12 <https://cc.au.dk/en/warcnet/working-groups>





távon való fenntarthatóságra koncentrálnak a WARCnet pályázat lejárata utáni időszakra. Felmérik a különböző európai uniós pályázati lehetőségeket, az egyes, már létező nemzeti, illetve nemzetközi kutatási hálózatokba történő bekapcsolódás formáit.

Az első, a harmadik és az ötödik munkacsoporthoz kapcsolódva fontos kérdésként vetődött fel, hogy miként lehetne becsatornázni a kutatói igényeket a webarchívumok szolgáltatási környezetébe, hogyan lehet azokat egyáltalán megfogalmazni, milyen adatformátum és szoftveres eszköztár szükséges a kutatásokhoz. Miként lehetne megteremteni egyfajta átjárható közös alapokkal bíró szolgáltatási és adatformátum környezetet a kutatók számára, mely egységes alapokat nyújt az egyes nemzeti webarchívumokban zajló vizsgálódásokhoz.

Bemutatásra kerültek továbbá az eseményalapú adatelemzési munka első eredményei a COVID gyűjtemény kapcsán. Elkezdtek leválogatni, hogy a felsőszintű domének szerinti eloszlásban honnan és mennyi adatot sikerült begyűjteni, továbbá vizsgálják, hogy az egyes nemzeti webterekről lementett anyagoknak milyen sajátosságai vannak, illetve összetételt, tartalmi jellemzőket tekintve milyen eltéréseket lehet tapasztalni, továbbá melyek voltak azok a főbb kulcsszavak, témakörök, melyek az archivált webtartalomban megjelentek, és milyen összetétele van a webhelyek típusa szerint a begyűjtött anyagnak általában, illetve a nemzeti webterek szintjén. A tartalomelemzésben rejlő lehetőségek bemutatására külön datathont is szerveztek.

Niels Brügger és csapata a Dán Nemzeti Könyvtár webarchívumának 2006-2016 közötti anyagára alapozva egy olyan tartalomelemzési munkát végez, ami során felmérik, hogy melyek voltak a főbb témák a dán webtérben és ezek hogyan változtak az idők során. Itt több milliónyi kulcsszó leválogatásáról és elemzéséről van szó, amihez a megfelelő erejű informatikai infrastruktúrát csak a közelmúltban sikerült felállítani.

A projekt lényeges eredményeként említhető meg a WARCnet Papers sorozat<sup>13</sup>, mely a kutatómunka közepette született eredmények gyors publikálásra ad lehetőséget. A sorozat első darabjai 2020 tavaszán jelentek meg és elsősorban a webarchiválással kapcsolatos kutatások perspektíváinak felvillantásával foglalkoztak. 2021-ben Niels Brügger összefoglalót készített a projekt első évének tevékenységeiről, eredményeiről. A publikációk magvát a COVID járvánnyal kapcsolatos webarchiválási tevékenységet feltáró interjúk jelentik.

Reményeink szerint a projekt lezárultával mérleget lehet majd vonni arról, hogy miképp sikerült nemzetközi szinten intézményesíteni a webarchívumokra irányuló, illetve az archivált webes anyagot forrásként használó kutatásokat, és teljes összkép állítható majd fel a projekt eredményeiről.

## 5. Internet Archive kereső a .hu doménról gyűjtött anyagok feltárására

Az Internet Archive egy magyar nemzeti portál kialakítására tett ajánlatot, mely tartalmazza az amerikai non-profit szervezet globális webarchívumából leválogatott összes .hu végződésű domén archivált anyagát. A portál az OSZK márkanevéhez illeszkedve kerül kialakításra, számos, a keresést segítő innovatív eszköz felhasználásának lehetőségével, melyek teljes képet igyekeznek adni arról, hogy miként jelenik meg a magyar nemzeti web domén, illetve annak egyes elemei az Internet Archive gyűjteményében.

A szolgáltatás keretében kialakítandó portál mintegy 1.2 milliárd, a .hu doménból gyűjtött URL archivált verzióját teszi elérhetővé, 1996-tól 2021 júniusáig.

13 <https://cc.au.dk/en/warcnet/warcnet-papers>



The screenshot shows the Internet Archive search results for the keyword 'kutyafajták'. The search interface includes a search bar with the keyword, a search button, and filters for file types like 'webhelyek', 'képek', 'hang', 'videó', and 'PDF'. The results are listed with their respective URLs and snippets of text. The first result is 'Kutyafajták | tlap.hu', the second is 'Magyar kutyafajták | magyar-kutyafajtak.linkpark.hu | LinkPark', and the third is 'Magyar kutyafajták | tlap.hu'. Each result includes a 'Legkorábbi verzió' link.

3. ábra: Az Internet Archive .hu domain alatti teljesszövegű keresést nyújtó portáljának találati oldala

A tömörített állomány mérete kb. 66 TB. A portál alapvető szolgáltatásai a következők: Részletes keresést lehetővé tevő funkciók, pl. kulcsszó, URL cím, dokumentumtípus (kép, videó, audió, PDF) szerint. Hozzáférés az archivált anyaghoz dedikált API segítségével, centralizált ElasticSearch klaszter felhasználásának lehetősége. Az archivált anyag megjelenítését a továbbfejlesztett Python alapú Wayback Machine eszköz biztosítja. A keresési találatok a webhelyekről és képfájlokról készült bélyegképeket is tartalmazzák. A keresési találati lista hivatkozásokat ad külön a legrégebbi, a legújabb és az összes mentés felé, valamint domén statisztikai funkciókat is kínál aggregált információkkal és MIME-típus szerinti szűréssel. A portál felületének megjelenése illeszkedni tud az OSZK honlapjához, amennyiben az intézmény igényli, és saját egyedi domén névvel is el lehet érni azt. Egy olyan verzió is rendelkezésre állhat, melyet integrálni lehet az intézmény szolgáltatási kínálatának egyéb részei mellé. A gyűjtemény egyes elemeihez történő hozzáférés korlátozására is van lehetőség, de miután a tartalom megjelenítése az Internet Archive szervereiről történik a Wayback Machine eszközzel, ezért alapértelmezésben az állomány szabad hozzáférése a kívánatos. Felhasználói analitikák és keresőlogok is az OSZK rendelkezésére állnak a szolgáltatás részeként.

Az éves fenntartáshoz és frissítéshez az alábbi szolgáltatási elemek kötődnek: a .hu doménről archivált állomány negyedévenként történő rendszeres frissítése. Az újonnan bekerülő anyag URL és kulcsszó alapú indexelése a keresőszolgáltatás számára. Webhely hosting, storage és fenntartási költségek, ideértve az összes hálózattal, sávszélességgel és karbantartással kapcsolatos díjakat is. Hozzáférés az anonimizált felhasználói és kereső analitikát érintő naplófájlokhoz. Évente 5-10 TB-ra becsült új anyag hozzáadása a gyűjteményhez. Az újonnan hozzáadott tartalmak teljesszövegű keresésének biztosítása.

Jelentős hozzáadott értékkel bír ezen felül a .hu domén anyagából összeállított zóna fájl, melynek révén az Internet Archive egyszeri listát állít össze hozzávetőleges teljességgel a .hu



domén alatti URL címekről, melyek az Internet Archive globális archívumában és kapcsolódó erőforrásai révén érhetők el. Ennek köszönhetően az OSZK jelentős mértékben tudja majd bővíteni az általános aratásba bevont webhelyek körét.

## 6. Epilógus

Az Országos Széchényi Könyvtárban a szervezett keretek között zajló webarchiválási tevékenység csupán 2017-ben tudott megindulni, ami nagyon jelentős hátrányt jelenthetett volna számunkra, ha nem találkozunk készséges külföldi partnerekkel, akik minden segítséget megadtak, hogy a már meglévő tapasztalatokat, jó gyakorlatokat felhasználva ki tudjuk alakítani ezen új tevékenységnek a kereteit, s be tudjunk illeszkedni a partnerintézmények által korábban megalkotott szakmai környezetbe. Nem sokkal az indulás után pedig már konstruktív módon elkezdhattünk részt vállalni a közös projektekben. Öt év múltán érdemes volt tehát áttekinteni, hogy miként alakult ki a nemzetközi kapcsolatrendszerünk, milyen együttműködési formákba tudtunk bekapcsolódni, s néhány tervet is fel tudtunk vázolni a jövőre nézve.

## Irodalomjegyzék

- Brügger, Niels., Laursen, Ditte., & Schafer, Valerie (2019, június 6). Opportunities and Challenges in Collecting and Studying National Webs [Presentation]. Letöltés 2020. augusztus 20., helye: <https://digital.library.unt.edu/ark:/67531/metadc1609008/>
- Egense, Thomas., & Klindt Myrvoll, Anders (2018, november 15). SolrWayback: Demo of the SolrWayback search interface, tools and playback engine for WARC's [Presentation]. Letöltés 2020. augusztus 20., helye: <https://digital.library.unt.edu/ark:/67531/metadc1477121/>
- IIPC. (2019). IIPC portal. Letöltés 2021. június 09; helye: <https://netpreserve.org>
- International Internet Preservation Consortium (IIPC) repository. (2020). Letöltés 2021. június 09; helye: <https://digital.library.unt.edu/explore/partners/IIPC/>
- IIPC TWG. (2019). IIPC Training Working Group. Letöltés 2021. június 09; helye: <http://netpreserve.org/about-us/working-groups/training-working-group/>
- IIPC Training Working Group. (2020). Training materials. Letöltés 2021. június 09; helye: <http://netpreserve.org/web-archiving/training-materials/>
- IIPC Training Video Case Studies. (2020). Letöltés 2020. június 09; helye: <http://netpreserve.org/web-archiving/iipc-training-video-case-studies/>
- Németh, Márton Nemzetközi körkép a webarchiválás gyakorlatáról. Könyvtári Figyelő, 63, 4.sz. (2017) 575–582. Letöltés helye: [https://epa.oszk.hu/00100/00143/00349/pdf/EPA00143\\_konyvtari\\_figyelo\\_2017\\_04\\_575-582.pdf](https://epa.oszk.hu/00100/00143/00349/pdf/EPA00143_konyvtari_figyelo_2017_04_575-582.pdf)
- Németh, Márton. Metadata supported full text search in a web archive. Előadás IIPC conference, Zagreb. Letöltés ideje 2020. január 8; helye: [http://mekosztaly.oszk.hu/mia/doc/Metadata\\_supported\\_full-text\\_search\\_in\\_a\\_web\\_archive-IIPC-2019-Zagreb.pptx](http://mekosztaly.oszk.hu/mia/doc/Metadata_supported_full-text_search_in_a_web_archive-IIPC-2019-Zagreb.pptx)
- Németh, Márton. Nemzetközi tanácskozás az internet megőrzéséről. Letöltés 2020. augusztus 17; helye: [https://nemzetikonyvtar.blog.hu/2019/06/18/nemzetkozi\\_tanacskozas\\_az\\_internet\\_megorzeserol](https://nemzetikonyvtar.blog.hu/2019/06/18/nemzetkozi_tanacskozas_az_internet_megorzeserol)
- Németh, Márton. Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary [E-paper]. Letöltés helye: [https://cc.au.dk/fileadmin/user\\_upload/WARCnet/Geeraert\\_et\\_al\\_COVID-19\\_Hungary.pdf](https://cc.au.dk/fileadmin/user_upload/WARCnet/Geeraert_et_al_COVID-19_Hungary.pdf)

- Németh, Márton Webarchívum mint a tudományos kutatások tárgya. Tudományos és Műszaki Tájékoztatás, 67, 12.sz (2019), 757–765. Letöltés helye: <https://tmt.omikk.bme.hu/tmt/article/view/12804/14543>
- Németh, Márton. How to catalogue a web archive? Előadás Information Interactions 2018 workshop, Bratislava. Letöltés ideje 2020. január 8; helye: [http://mekosztaly.oszk.hu/mia/doc/How\\_to\\_catalogue\\_a\\_web\\_archive.pptx](http://mekosztaly.oszk.hu/mia/doc/How_to_catalogue_a_web_archive.pptx)
- Németh, Márton. The education of web archiving. Előadás CDA 2018 conference, Bratislava. Letöltés ideje 2020. január 8; helye: [http://mekosztaly.oszk.hu/mia/doc/The\\_education\\_of\\_web-archiving\\_CDA\\_2018.pptx](http://mekosztaly.oszk.hu/mia/doc/The_education_of_web-archiving_CDA_2018.pptx)
- Németh, Márton. Using semantic microformats for web archiving – an initial project conception. In Katarina Tomková (Szerk.), Nové trendy a východiská pri budovaní LTP archívov: Zborník príspevkov zo 4. Medzinárodnej konferencie o dlhodobej archivácii Bratislava, 5. 11. 2019. (pp. 31–38). Bratislava: Univerzitná knižnica v Bratislave. Letöltés ideje 2020. január 14; helye: <https://cloud.ulib.sk/index.php/s/JITVgF9TmRruFtC>
- Németh, Márton. Web Archives as a Research Subject. In Information and technology transforming lives: Connection, interaction, innovation proceedings (pp. 471–478). Osijek: Faculty of Humanities and Social Sciences, University of Osijek. Letöltés helye: [http://bobcatsss2019.ffos.hr/docs/bobcatsss\\_proceedings.pdf](http://bobcatsss2019.ffos.hr/docs/bobcatsss_proceedings.pdf)
- Németh Márton. Webarchiválás két szakmai rendezvény tükrében. Könyv, könyvtár, könyvtáros, 28 (2019), 6.sz., 26–29.
- Németh, Márton & Drótos, László. A blended learning-based curriculum on Web archiving in the National Széchényi Library. Digital Library Perspectives, 35. 2.sz (2019), pp. 97–114. doi: 10.1108/DLP-03-2019-0012. Letöltés helye: <https://doi.org/10.1108/DLP-03-2019-0012>
- Németh, Márton & Drótos, László. Metadata Management and Future Plans to Generate Linked Open Data in the Hungarian Web Archiving Pilot Project. ITLIB, 2019(2). Letöltés helye: <https://itlib.cvtisr.sk/buxus/docs/38-metadata.pdf>
- Netlab. NETLAB Web Archiving Course Brochure. Letöltés 2019. január 28; helye: <http://www.netlab.dk/wp-content/uploads/2017/04/NetLab-Web-Archiving-Course-Brochure.pdf>
- Netlab. Netlab-courses. Letöltés 2019. január 28; helye: <http://www.netlab.dk/services/courses/>
- Nielsen, J. Using web archives in research – an introduction (1.). Aarhus: Netlab. Letöltés helye: [http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen\\_Using\\_Web\\_Archives\\_in\\_Research.pdf](http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf)
- WARCnet (2020). Letöltés 2021. június 09; helye: <https://cc.au.dk/en/warcnet/>
- The WARCnet team (2020). Letöltés 2021. június 09; helye: <https://cc.au.dk/en/warcnet/the-warcnet-team>
- WARCnet working groups (2021). Letöltés 2021. június 09: <https://cc.au.dk/en/warcnet/working-groups>
- WARCnet presentations (2021). Letöltés 2021. június 09: <https://cc.au.dk/en/warcnet/presentations>
- WARCnet papers (2021). Letöltés 2021. június 09: <https://cc.au.dk/en/warcnet/warcnet-papers>

## A mesterséges intelligencia kihívásai a XXI. század társadalmára

Antal Péter

*Eszterházy Károly Katolikus Egyetem, Digitális Kultúra Tanszék*

[antal.peter@uni-eszterhazy.hu](mailto:antal.peter@uni-eszterhazy.hu)

### Absztrakt

A mesterséges intelligencia jelenlétét, már hétköznapi tényként fogadjuk el. Ennek ellenére még mindig nagyon keveset tudunk azokról a vitákról, melyek a mesterséges intelligencia létének etikai, jogi, erkölcsi, kérdésiról, jelenlegi és várható társadalmi hatásairól szólnak. A fejlődés természetesen töretlen, de még nem tisztáztuk azokat a kérdéseket, melyek a mesterséges intelligencia fejlettségének különböző szintjein várnak ránk. A kutatók jó része még a technológiai szingularitás bekövetkezésével és annak időpontjával szemben is szkeptikusan fogalmaz. Annak ellenére a jövő szempontjából nagyon fontos kérdés, hogy ha bekövetkezik ez a pillanat, mennyire felkészülten tesszük ezt.

Az Általános Mesterséges Intelligencia alapvetően nem jó vagy rossz. De a sebességét, összetettségét és intelligenciáját tekintve az emberiségnek csak egy apró pillanata lesz, mielőtt az ÁMI meghalad minket. Tehát feltétlenül fontos, hogy az ÁMI nagyszabású bevezetése előtt foglalkozzunk a megbízhatóságával és a célok összehangolásával.

Előadásomban ezekre és más hasonló kérdésekre keresem a választ a MI és az emberiség jövője szempontjából.

**Kulcsszavak:** Általános Mesterséges Intelligencia, AI, jövő kutatás

### Artificial Intelligence-Related Challenges onto 21st Century Society

While artificial intelligence has become an integral part of our life, we are still not fully familiar with the respective ethical, legal and moral issues along with its current and prospective social impact. Although the progress appears uninterrupted, certain questions related to the respective developmental levels should still be clarified. Most researchers maintain a sceptical attitude towards the emergence and actual time of the onset of technological singularity despite the importance of being prepared when it takes place.

Artificial General Intelligence cannot be considered either inherently good or bad. However, regarding the speed, complexity and intelligence level it will be a hardly recognizable moment when Artificial General Intelligence will surpass the capabilities of humans. Therefore, it is increasingly important that attention be paid to its reliability and to the reconciliation or synchronisation of the given objectives before its large scale introduction.

My lecture exploring AI's potential impact on the future of humanity searches for answers to these and similar questions.

**Keywords:** Artificial General Intelligence, AI, future studies,

## 1. Bevezetés

A mesterséges intelligencia jelenlétét, már hétköznapi tényként fogadjuk el. Ennek ellenére még mindig nagyon keveset tudunk azokról a vitákról, melyek a mesterséges intelligencia létének etikai, jogi, erkölcsi, kérdésiról, jelenlegi és várható társadalmi hatásairól szólnak. A technológiai fejlődés természetesen töretlen, de még nem tisztáztuk azokat a kérdéseket, melyek a mesterséges intelligencia fejlettségének különböző szintjein várnak ránk. A kutatók jó része, még a technológiai szingularitás bekövetkezésével és annak időpontjával szemben is szkeptikusan fogalmaz. Annak ellenére, a jövő szempontjából nagyon fontos kérdés, hogy ha bekövetkezik ez a pillanat, mennyire felkészülten tesszük ezt.

Az Általános Mesterséges Intelligenciáról (ÁMI) nem jelenthetjük ki kategorikusan, hogy jó vagy rossz, de a fejlődés sebességét, összetettséget és intelligenciáját vizsgálva az emberiségnek csak egy apró pillanata lesz, mielőtt az ÁMI meghaladja az ember tudását. Tehát feltétlenül fontos, hogy az ÁMI nagyszabású bevezetése előtt foglalkozzunk a megbízhatóságával és a célok összehangolásával. Vagyis az egyik legégetőbb kérdés, hogy milyen hatásai lehetnek majd a jövőnkre nézve.

## 2. Az intelligencia és az élet szakaszai

Max Tegmark szerint az intelligenciát egyszerűen definiálhatjuk úgy, mint a „komplex célok elérésének képességét”. [7] Azonban nagyon hosszú utat jártunk be, amíg maga az intelligencia megjelent a Földön, hiszen az élet megjelenésekor, még nem beszélhettünk a mai értelemben vett intelligenciáról.

Mondhatjuk azt is, hogy az intelligencia is egy sajátos evolúciós folyamaton ment keresztül, az élethez hasonlóan.

Tágabb értelemben az életet nevezhetjük egy „önszaporító információ-rendszernek, mely képes a komplexitását fenntartani”. [7] Az élet fejlődését ennek alapján, három szakaszra bonthatjuk:

- Élet 1.0, biológiai evolúció (biológiai)
- Élet 2.0, kulturális evolúció (ember - tanulás)
- Élet 3.0, technológiai evolúció (már nem biológiai intelligencia)

### Az élet 3 szakasza

	Képes túlélni és szaporodni	Képes megtervezni szoftverét	Képes megtervezni hardverét
 Élet 1.0 (egyszerű biológiai)	✓	✗	✗
 Élet 2.0 (kulturális)	✓	✓	✗
 Élet 3.0 (technológiai)	✓	✓	✓

1. ábra: Az élet három szakaszának jellemzői [7]



Az **Élet 1.0**, maga a biológiai evolúció, amely során, az első élőlények a sokszorozódásukhoz a génjeikben meghatározott információt másolták le, oly módon, hogy nem az (atomokból felépülő) anyag sokszorozódott, hanem az a bitekből álló információ, amely meghatározta az atomok elrendeződését. A környezeti hatások változásaira reagálva, ebben a genetikai kódban mutációk jöttek létre, melyek a fejlődés mozgatórugóivá váltak. Ennek ellenére ez egy nagyon lassú folyamat, hiszen a viselkedésüket (a szoftvert) és a testfelépítésüket (hardvert) is a DNS-ben tárolt információ határozta meg.

Az **Élet 2.0**, a kulturális evolúció, amire legjobb példa maga az ember. Abban különbözik az elsőtől, hogy a szoftver már nem fejlődés, hanem tervezés eredménye.<sup>1</sup> Óriási előnyt jelent, hogy az emberi hardver (a test) is fejlődésen megy keresztül, nem beszélve arról, hogy a szoftver (tudás) a tanulás folyamata során alakul ki, így a szerzett intelligenciánk nem függ a DNS-ben tárolt információ mennyiségétől. A neuronjainkat összekötő szinaptikus kapcsolatok hozzávetőlegesen százezerszer több információt képesek tárolni, mint a születéskori DNS.

Az **Élet 2.0** sokkal rugalmasabb, hiszen ha megváltozik a környezet, nagyon rövid idő alatt alkalmazkodik a környezet változásaihoz, mégpedig szoftverfrissítéssel. Például, ha valaki megtudja, hogy allergiás a laktózra, azonnal meg tudja változtatni a viselkedését, és kerülni fogja a tejtermékek fogyasztását.

Amikor megszületünk, csak reflexszerű dolgokra vagyunk képesek, aztán a fejlődés során bonyolult új készségeket leszünk képesek elsajátítani. Előbb a szülők, az iskola, végül saját magunk döntjük el, milyen szoftvert táplálunk be a hardverünkbe.

Amíg az emberi DNS-ben tárolt információ, az utóbbi 50 ezer évben, nem ment keresztül drámai változáson, addig az agyunkban, könyvekben és számítógépeken tárolt információ összessége robbanásszerűen megsokszorozódott.

Mindezt agyunk rugalmasságának, a tanulás készségének és a vele kialakuló intelligencia megjelenésének köszönhetjük. Ez a telepített szoftvermodul lehetővé tette olyan eszközök „telepítését” mint a több csatornás kommunikációs készség, aminek segítségével akár egy másik ember agyában tárolt információt le tudunk másolni, így az információ túlélheti az eredeti agy halálát. Az írás-olvasás szoftvermodulja lehetővé tette, az információ külső tárolórendszerekbe történő lementését, kibővívte ezzel az agyi kapacitásunkat. *Marshall McLuhan* szerint, a mai ember számára a tudást már nem az információ birtoklása, hanem az elektronikusan hozzáférhető végtelen információáradatban való eligazodás képessége határozza meg. [5]

Az emberiség természetesen elindult az **Élet 3.0** megvalósításának irányába, hiszen kisebb hardverfrissítéseket magunk is végre tudunk hajtani (művégtagok műszív stb.), de még nem élhetünk több száz évig, vagy nem növelhetjük meg az agyunk méretét.

Az **Élet 3.0** pedig ezen is tovább lép, sorsának teljesen saját ura lenne, hiszen nemcsak szoftverét, hanem a hardverét is képes lenne megszervezni, megszabadulva az evolúciós béklyóktól. Eljön az **Általános Mesterséges Intelligencia (ÁMI)** kora, melynek legfontosabb jellemzői a következők:

- már nem biológiai intelligencia,
- állandó emberi beavatkozás nélkül képes válaszolni környezeti behatásokra (automata eszközök),

1 Szoftver alatt itt azokat az algoritmusokat és tudást értjük, amelyeket az érzékeink nyújtotta információk feldolgozásához használunk, és amelyek alapján eldöntjük, hogy mit tegyünk – az ismerőseink felismerésének képessége, a mozgás, olvasás, írás, számolás, éneklés és viccmesélés mind-mind idetartozik.



- képes hasonlóan viselkedni, mint egy természetes intelligenciával rendelkező élőlény (gépi karakterek).
- képes viselkedését célszerűen és megismételhető módon változtatni (tanulás, önfejlesztés)

### 3. Kétségek és tények

Az emberi intelligencia széles spektrumú a tanulás képessége révén, hiszen bármit képesek vagyunk megtanulni, de agyunk csak bizonyos fiziológiai határok között képes működni. A feltételezett intelligenciारobbanás azon alapszik, hogy létrejön egy biológiai korlátoktól mentes emberi szintű MI, amely önmagát szabadon tudja fejleszteni és a visszacsatolás miatt, véges idő alatt tudna magának lényegében akármilyen nagy számítási teljesítményt fejleszteni.

A legnagyobb hiányosság azonban az, hogy a mai MI-k egyelőre nem univerzálisak. Bár képesek tanulni, de csak egy bizonyos részterületen, bizonyos korlátok között és bizonyos szempontok alapján. Hogy a teljes univerzalitás mennyire csak a számítási kapacitás kérdése, az vita tárgya.

*„Az ultraintelligens gépek hamar maguk mögött hagynák az ember intelligenciáját, emiatt az ultraintelligens gép lenne az utolsó találmány, amit az embernek létre kéne hoznia”. [2] Így fogalmazott Irwing. J. Good már 1965-ben.*

Raymond Kurzweil közismert mesterségesintelligencia-kutató a *Spirituális gépek kora és a Szingularitás küszöbén*, című könyvek szerzője a szingularitás bekövetkezését 2045-re jósolja. A dátum relatíve közelinek tűnik, de Kurzweil szerint ez a fejlődés érzékelésének lineáris ütemű illúziója miatt van, miközben a valóságos fejlődés exponenciális ütemű. Ennek megfelelően a 21. században nem 100 évnyi, hanem a jelenlegi ütemben mérve 20.000 évnyi fejlődést fogunk megtapasztalni. [3]

Kurzweil jóslatának alapja, az exponenciális ütemű fejlődés, mely a technológia számos területén igazolódott. A legismertebb a Moore-törvény, [6] amely szerint az integrált áramkörök összetettsége 18 hónaponként megduplázódik. De hasonló exponenciális ütemű fejlődés érvényes a számítási sebességre vagy a chipék méretére is.

### 4. Lehetőségek és kihívások

Az MI fejlesztése gyakorlatilag minden fontos iparágban, (közlekedés, energiaipar, kommunikáció és médiaipar, hadipar) az egészségügyben, a törvényhozásban, vagy a természetes nyelvi feldolgozásban, mindenhol nagyon fontos stratégiai szerepet tölt be.

Azonban bevezetésével kapcsolatban mindenhol fontos kérdések vetődnek fel, melyek az emberi létezés alapjait érintik, hiszen egyre kevésbé állja meg a helyét az az érv, hogy az MI-nek nincs célja intuíciója, kreativitása vagy nyelve.

A legfontosabb felmerülő kérdések a következők, melyekre megbízható és pontos választ kell adnunk a közeli jövőben:

- Hogyan alakíthatunk ki a mainál robusztusabb MI-rendszereket, amelyek engedelmeskednek az utasításainknak anélkül, hogy összeomlanának, a működésükben hiba lépne fel, és feltörhetetlenek legyenek?
- Hogyan oldhatjuk meg, hogy a fegyverek okosabbak legyenek, és kisebb legyen az esély, hogy ártatlan civilek életét oltás ki, mégse alakuljon ki fékezhetetlen fegyverkezési verseny a halálos automata fegyverek területén?
- Hogyan növelhetjük a jólétet az automatizáláson keresztül úgy, hogy nem fosztjuk meg az embereket a jövedelmüktől illetve életcéljuktól?



A legfontosabb kérdés a biztonság. A fejlődésünk során próbáltunk tanulni a hibáinkból. Az eddig feltalált technológiák előnyei jellemzően túlsúlyban voltak a hátrányaikhoz képest. Annak érdekében, hogy az ÁMI a jövőben kezelhető legyen az MI-biztonságtechnikai kutatások négy fontos kérésére kell fókuszálni az MI szoftverek szempontjából:

- **Verifikáció** – kiküszöbölni a „kék halált”. Megfelelően építettem meg a rendszert?
- **Validáció** – robotok tanuljanak meg óvatosnak lenni. A megfelelő rendszert építettem meg?
- **IT biztonság** – adatok és eszközök folyamatos teljes körű védelme, sértetlensége, folytonos rendelkezésre állása
- **Vezérlés** – gépek autonómiája, ellenőrizhetőség<sup>2</sup>

## 5. Társadalmi kihívások

A társadalom szempontjából is nagyon fontos kérdések merülnek fel az MI hatásaival kapcsolatban.

Hogyan hat majd az MI a munkánkra, a munkaerőpiacra, a jövedelmekre, milyen szakmát adjunk, a gyerekeink kezébe?

*Erik Brynjolfsson* „Digitális Athén” elmélete, optimista jövőt jósol az emberek és az MI kapcsolatnak. Eszerint az athéniak azért éltek jól, mert a munka nagy részét rabszolgák végezték, így több idejük maradt a művészetekre, sportra stb. Felmerül a kérdés, miért ne lehetnének a gépek a rabszolgáink, megteremtve a bőséget, a stresszmentességet és a fogyasztás korát. [1]

Az USA-ban a 70-es évekig ez működött, nőtt a jövedelem minden társadalmi szektorban igaz az eloszlás nem volt egyenletes. Ezután valami megváltozott. A gazdasági növekedés ugyan nem állt le, azonban a jövedelem stagnálni kezdett, a társadalom 90%-a részére és a többlet a társadalom 1%-át kitevő szupergazdag réteg kezébe került.

Az okokat a közgazdászok a világgpiaci trendekkel magyarázzák, de egyes kutatók szerint a fő ok a technológia fejlődése volt. Ez háromféleképpen járult hozzá az egyenlőtlenséghez.

- A régi szakmák lecserélődtek, a több készséget és szakmai tudást igénylő újakra, melyek a tanultabb rétegeknek kedveztek.
- Az automatizálás révén a bevételek többsége a gépek tulajdonosaihoz, vagyis a cégtulajdonosokhoz került. Jó példa erre a nagy detroiti hármás (Chrysler, Ford, GM) összbevétele 1990-ben ugyanakkora volt, mint a Szilícium-völgy nagy hármasáé (Google, Apple, Facebook). Utóbbiak kilencszer kevesebb embernek adnak munkát, de harmincszor érnek többet a tőzsdén. [4]
- A digitális gazdaság a sztárokat részesíti előnyben (egy jó alkalmazás mindent visz).

Egy biztosan kijelenthető, hogy a tanulás kifizetődik, és olyan szakmák lesznek a jobban fizetettek, melyek emberi interakciót, a társas intelligenciát, a kreativitást és a gyors problémamegoldó képességet igénylik. Emellett fontos tényező, a munkakörnyezet változatosságához, kiszámíthatatlanságához való alkalmazkodás képessége.

## 6. Jelen és jövő

A MI jelene is sok kérdést vet fel, a jövő pedig rajtunk múlik, hogy egy számunkra kedvezőbb vagy egy végzetes jövő felé megyünk. Jelenleg a legfontosabb feladat, a célok meghatározása és összehangolása. Nézzük meg ennek az elemeit:

2 forrás: AI Safety Research <https://futureoflife.org/landscape/>

- Az emberiségnek meg kell vitatnia és meg kell állapodnia arról, hogy az ÁMI milyen és kinek a céljait akarja elfogadni (pl. szabadság, hasznosság), és hogy hogyan kell alkalmazni ezeket a célokat a nem emberi jogokra (pl. az állatok estében).
- Az ÁMI és céljaink összehangolása érdekében három kihívással kell megküzdenünk.
  - Az ÁMI legyen tisztában az emberi célokkal, úgy hogy képes legyen megérteni, miért akarjuk őket, beleértve feltételezéseinket, rész céljainkat és prioritásainkat.
  - Az általános emberi célok elfogadása, például, hogy milyen értékrend szerint neveljük a gyerekeinket.
  - Annak biztosítása, hogy megtartják céljainkat abban az esetben is, ha tudásunkban, képességeinkben felülmúlnak minket.

A jövő tekintetében még több a bizonytalanság, lehet meg sem születik az Élet 3.0, lehet, hogy ez lesz az emberiség végzete. A lehetséges kimeneteleket a következő infografika foglalja össze.



2. ábra: Az Élet 3.0 lehetséges kimenetelei<sup>3</sup>

## 7. Összegzés

Az már biztos, hogy az MI képes intelligensebbé válni az embernél, de nem tudjuk megjósolni, hogyan fog viselkedni. Nem készültünk még fel olyan dolgokra, amelyek képesek akarva, vagy akaratlanul is kicselezni minket. A legjobb példa arra, amivel szembesülhetünk, az a

3 forrás: <https://readinggraphics.com/product/download-life-3-0-summary-in-pdf-audio-graphic/>  
Copyright © 2018 Skool of Happiness Pte Ltd.



saját evolúciónk lehet. Az emberek most nem azért irányítják a bolygót, mert mi vagyunk a legerősebbek, leggyorsabbak vagy a legnagyobbak, hanem azért, mert mi vagyunk a legokosabbak. Mi lesz, ha már nem leszünk a legokosabbak, tudjuk-e biztosítani, hogy továbbra is mi irányítsunk?

#### Irodalom

- [1] BRYNJOLFSSON E., McAfee A.: *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*, New York, Norton, 2014. ISBN13: 9780393350647
- [2] GOOD, I. J. (1965): „Speculations Concerning the First Ultrainelligent Machine”, in: *Advances in Computers*, vol 6, Franz L. Alt and Morris Rubinoff, eds, pp. 31-88, 1965, Academic Press. [https://doi.org/10.1016/S0065-2458\(08\)60418-0](https://doi.org/10.1016/S0065-2458(08)60418-0)
- [3] KURZWEIL R. (2013) : A szingularitás küszöbén, Amikor az emberiség meghaladja a biológiát, Ad Astra kiadó 2013. ISBN 978-615-5229-26-8
- [4] MANYIKA, J.: „Can technology and Productivity save the day?” c. ea. [https://futureoflife.org/data/PDF/james\\_manyika.pdf?x17135](https://futureoflife.org/data/PDF/james_manyika.pdf?x17135) letöltés ideje: 2022. március 21.
- [5] McLuhan, M. (2011): *The Gutenberg Galaxy*. Toronto: University of Toronto Press
- [6] MOORE G. E. (1965): Cramming more components onto integrated circuits In: *Electronics*, Volume 38, Number 8, April 19, 1965. <https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf> letöltve: 2022. 03 15.
- [7] TEGMARK, M. (2018): *Élet 3.0 Embernek lenni a mesterséges intelligencia korában* HVG-Könyvek, Budapest, 2018. ISBN 978-963-304-504-6

Modern robotikai technológiai ismeretek oktatása „Teljes spektrumú” oktatási módszerrel  
Application of “full-spectrum” teaching methods in the education of modern robotic  
technology

Hajdu Csaba  
Széchenyi István Egyetem Gépészmérnöki,  
Informatikai és Villamosmérnöki Kar Automatizálási Tanszék  
[hajdu.csaba@ga.sze.hu](mailto:hajdu.csaba@ga.sze.hu)

Szilágyi Zoltán  
Széchenyi István Egyetem Gépészmérnöki,  
Informatikai és Villamosmérnöki Kar Automatizálási Tanszék  
[szilagyi.zoltan@ga.sze.hu](mailto:szilagyi.zoltan@ga.sze.hu)

### Absztrakt

Napjainkban a feltörekvő technológiák és eszközök jelentős kihívást jelentenek az oktatási intézményeknek és oktatóknak egyaránt. Jelen cikk megkísérli a multidiszciplináris műszaki tudás átadását modern oktatási módszertan használatával egyváltozó tudásháttérű hallgatóságnak. A bemutatott módszertan alapképzésben indult bevezető robotikai kurzusban került felhasználásra. Jelen módszertan egyesíti a hagyományos pedagógiai módszereket a moderneekkel, mint a projekt-alapú feladatok, önértékelés és projektbemutatók, hasonlóan egy miniszimpóziumhoz. A hallgatók visszajelzése alapján az új módszertan ígéretes, az eredmények egy része igazolja előzetes elvárásainkat, ugyanakkor felfedi a javítási lehetőségeket más tárgyak esetében is.

**Kulcsszavak:** programozás oktatása, projektalapú oktatás, robotika oktatása, teljes-spektrumú oktatás

### Abstract

Nowadays, the emergence of new technologies and tools challenges educational institutions and educators alike with proper answers. This article proposes the use of modern educational methodology to solve the education of technology-oriented interdisciplinary knowledge for student groups of varied technical backgrounds. The presented methodology was used in an introductory robotics course for bachelor students. The current methodology combines the traditional methods with modern pedagogical approaches, such as project-oriented tasks, self-evaluation, and project presentation similarly to a mini-symposium session. Based on the feedback of the students, the new approach is promising, with some results confirming the initial expectations while also revealing new information on the enhancement to apply this methodology in further subjects.

**Keywords:** programming education, project education, student project, education of robotics, full spectrum education



## Bevezetés

A műszaki felsőoktatás a világon számos helyen, így hazánkban is jelentős mértékű kihívásokkal néz szembe. Egyik oldalról az emberiség technikai tudásának gyarapodása, a szakterületek ismeretanyagának szélesedése folyamatosan növeli és nehezíti a hallgatók számára elsajátítandó ismeretanyag mennyiségét. Másrészt viszont – és ez hazánkra kifejezetten érvényes – a hallgatók ismereti alapjainak és a képzéshez szükséges képességek hiánya tovább nehezíti az oktatásban részt vevők feladatát, az eredményes ismeret átadást, végsősorban elsajátítását. Ezen hatások szorításában a hallgatók könnyen válnak motiválatlanná, érdektelenné, és a diák-tanár partneri viszony könnyen változik diák-tanár ellentété. Tovább nehezíti a feladatot az a tényező is, hogy a képzésre szánható időtartam az ismeretanyag bővülésével együtt nem növekedett, hanem stagnált vagy inkább kis mértékben még csökkent is.

Jelen helyzet kikényszeríti az egyre nagyobb mértékű specializációt, mely komoly dilemma elé állítja a hallgatókat. A Széchenyi István Egyetem villamosmérnök képzésében résztvevő oktatóként magunk is szembesülünk a fent említett nehézségekkel. A technikai fejlődés előrehaladtával a villamosmérnöki ismeretek egyre nehezedtek az informatika és más magas szintű absztrakciós képességek elsajátításának igényével. A kihívásokra válaszul egy kísérletet próbáltunk tenni a feltárt ellentmondások feloldására és a nehézségek kezelésére.

## Választott tárgy

Kísérletünk tárgyául egy, a kihívások szempontjából minden feltételnek megfelelő kurzust választottunk. A villamosmérnök alapképzésen, az automatizálási szakirány hallgatóinak számára szabadon választott tárgyként került meghirdetésre az „**Autonóm és Intelligens Robotok**” című kurzus. Tekintve, hogy szakirányos és azon belül is szabadon választható tárgyról van szó, a tanszéki vezetéssel egyeztetve lehetőségünk volt összeállítani egy gyakorlati, projektalapú tanmenetet, melyet előzetesen a tárgyat felvenni kívánó hallgatókkal is megosztottunk. Továbbá kiscsoportos létszámmal (14 fővel) került a tárgy kiírásra, lehetővé téve a gyakorlati ismeret hatékonyabb átadását.

A tárgy témáját tekintve igen modern ismereteket tárgyal, ezért a hallgatók nagy számban érdeklődtek a téma iránt. Ugyanakkor a tényleges ismeretanyag elsajátításához csak keveseknek voltak meg a megfelelő előzetes ismeretei (pl. programozás, számítógépes-hálózatok) a tárgy anyagának elsajátításához.

A tárgy teljesítéséhez szükséges előismeretek:

- Objektum orientált programozás középszintű vagy annál magasabb ismerete
- Irányítástechnika, szabályzások átfogó ismerete

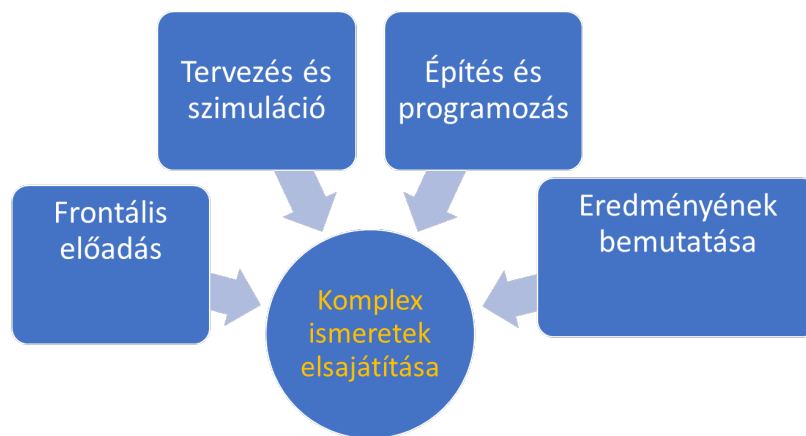
A tárgy tananyag tartalma, az elsajátítandó ismeretek:

- Robotirányítási rendszerek megismerése
- Python 3.10 programozási környezet használata
- Webots szimulációs szoftver használata
- Robot Operating System (ROS) 2 program könyvtár megismerése



## Módszertan

A választott oktatási módszerek vonatkozásában igyekeztünk elszakadni a hagyományos megoldásoktól. Bernát Péter és Zsakó László [1] által összegyűjtött módszertanokat elemezve jutottunk el a saját megoldásunkig. A kihívásokra és a hallgatói csoport összetételére figyelemmel, összeállítottunk egy teljes spektrumú oktatási csomagot. Ebben a diákok egyedi igényeire, valamint az ismeretelsajátítási szintekre figyelve, egymásra épülő szinteket alakítottunk ki. A valós eszközök oktatásban és a programozás oktatásban való használata megkönnyíti a készségek fejlődését, a megértés szintjének elmélyítését [2][3]. A tárgy során alkalmazott teljes oktatási módszertant foglalja össze az 1. ábra. A módszertant emellett még más elterjedt modern pedagógiai és előadói módszerek inspirálták egy informatikai kontextusban [4][5]. Megjegyzendő, hogy számos hazai felsőoktatási intézmény alkalmaz hasonló módszert az ismeretanyag átadására, különös tekintettel a gyakorlati ismeret megszilárdítására (pl. Budapesti Műszaki Egyetem témalaborjai), így ezen képzések tapasztalatai is inspirációt adtak.



1. ábra: Oktatási módszertan elemei

A **frontális előadás** képezte az ismeretátadás első szintjét. Az előadások alkalmával ismétlésre kerültek alapismeretek a robotika és a mechanika tárgyköréből. Ezt követően azon tananyagtartalmak következtek, melyek megalapozták az önálló projekt munkában végezhető feladat sikeres megvalósítását. Ide tartoznak a szoftverrendszerek és programnyelvek használatával kapcsolatos információk, valamint a témakör mélyebb szakmai ismeretei, rendszertervezési alapismeretek, irányítási algoritmusok és eszközök. A félév időtartamának 30%-át fordítottuk erre az oktatási tevékenységre.

Ezt követte az **önálló hallgatói projekt tervezése és szimulációja**, mely az első gyakorlati programozási feladatot jelentette. Ennek során a hallgatók csoportokba lettek sorolva és a csoportok önálló tervezési munka során kialakították a megvalósítandó mobil robot modellt, majd ennek szimulációs környezetben történő ellenőrzése és bemutatása következett. Itt kerültek első ízben ellenőrzésre a hallgatók szakmai elképzelései, megvalósíthatóság, tervezett viselkedés és egyéb más műszaki tartalmi szempontból. A hallgatók szembesülhettek elképzeléseik hiányosságaival, és a javítás során gyakorolhatták a módszeres gondolkodás és tervezés lépéseit. A szimulációkat Webots szimulációs szoftver [6] használatával valósították meg, melynek elsajátítása már a tananyagtartalom második elemét is magában foglalta, egy széles körben alkalmazott 3D szimulációs eszköz megismerését. Egy, a hallgatók által elkészített robotikai szimulációt mutat a 2. ábra.

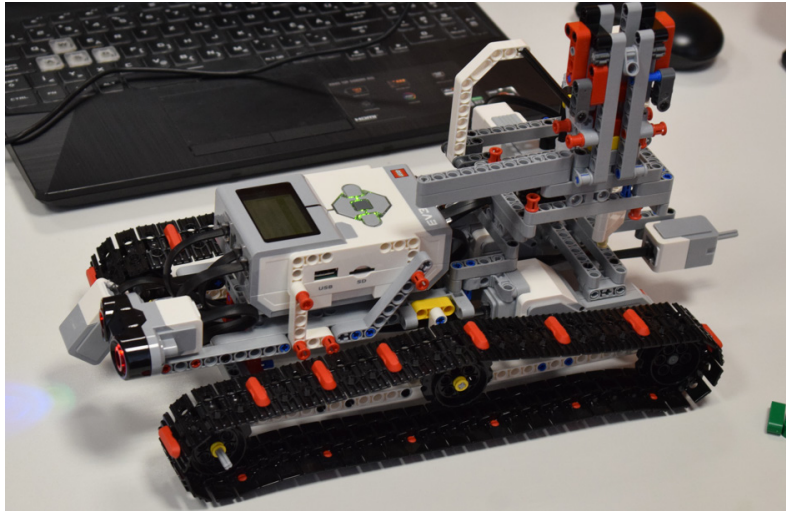


2. ábra: Hallgatók által elkészített robot Webots alapú szimulációja (vonalkövető robot)

A tervezés és szimulációs kontaktfoglalkozásokon túl, jelentős mértékű önálló elfoglaltságot, otthoni munkát is igényelt a hallgatók részéről. Ezt az évfolyam nagyon jól fogadta, számos esetben délutánonként kisebb hallgatói csoportok összegyűltek és közösen dolgozott a terveken, a szimuláción. Ez szabadabb gondolkodást, kötetlenebb feladatvégzést jelentet a hallgatók számára, erősítve a csapatszellemet, gyakoroltatva a közös munkavégzést szoros hierarchia kontrollja nélkül. A félév további 30%-át fordítottuk erre az oktatási elemre.

A harmadik szintet a **gyakorlati megvalósítás** jelentette, mely az ismeretek megerősítése és elmélyítése mellett az élményalapú oktatás eszközeként is jelen volt. A hallgatók a Lego Mindstorms EV3 építőkészletet használták az előzetesen megtervezett robot fizikai megvalósítására. Az építőkészlet elterjedt a gyakorlati oktatásban [7]. Az építéssel együtt a rendszer programozása, mint törzstéma itt került megvalósításra. Ez a módszertani elem egyben újabb visszacsatolást is jelentett a korábbi munkák minőségére és megfelelőségére vonatkozóan. Emellett az is fontos tanulságként szolgált a hallgatók számára, miszerint egy részletes szimuláció mellett is szükség van a valós robot megfelelő finomhangolására. Ezeket a tapasztalatokat tantermi körülmények között képtelenség átadni. A hallgatók Python 3 nyelven programozhatták a robotot, a robot vezérlőszámítógépére előzetesen konfigurált Linux operációs rendszeren<sup>1</sup>. Ezen felül használták a ROS2 keretrendszert [8], melynek elsajátítása szintén a tárgy központi eleme volt. A hallgatók által elkészített egyik robotot a 3. ábra mutatja.

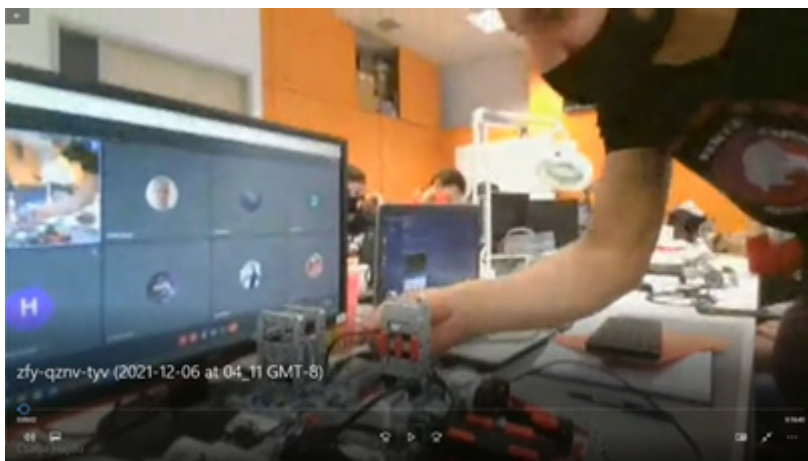
1 LEGO Mindstorms EV3 Linux: <https://www.ev3dev.org/>



3. ábra: Hallgatók által elkészített robot (tank robot)

Negyedik, egyben utolsó lépés volt az értékelés, mely jelen esetben nem a hallgatók kizárásával, hanem bevonásukkal történt meg. A bevonás nemcsak a további ismeret elsajátítást és gyakorlatszerzést célozta, hanem lehetőséget adott tágabb értelemben vett mérnöki és akár önismereti elemek beemelésére a kurzusba.

A hallgatóknak első lépésben be kellett mutatniuk a megvalósított projektjüket a társaiknak, és egyúttal értékelniük is kellett a feladat megoldását. A bemutatás egy online értekezlet és bemutató megvalósításával történt, melyen a hallgatók és oktató mellett más érdeklődők is részt vehettek. Az értékelés fő szempontja nem a tantárgyi követelményeknek való megfelelés volt, hanem a saját vállalásukhoz, tervükhöz képest elért eredmények, menet közbeni módosítások, kompromisszumok és kudarcok felmérése. Ezzel az önreflexiós lépéssel önmaguk számára foglalták össze a félév során elsajátított ismereteket és szembenéztek tudásuk gyarapodásával vagy épp annak bizonyos hiányosságaival. Az online bemutató egy pillanatképét mutatja a 4. ábra. Emellett természetesen tanári értékelést is kaptak a hallgatók.



4. ábra Hallgatói projekt prezentáció online felületen

## Eredmények értékelése – Hallgatói visszajelzés

A hallgatói vélemények megismerésének céljából kérdőíves felmérést végeztünk, ezt a kurzus vége után tettük közzé. Ügyeltünk arra, hogy ne legyen hosszú, vagy ne tartalmazzon ismétlődő kérdéseket. A kérdőívre visszakapott eredmények összefoglalását és a feltett kérdéseket az 1. táblázat mutatja.

1. táblázat: hallgatói kérdőív eredményének összegzése

Az oktatási módszerrel kapcsolatos visszajelzés									Átlag
Mennyire volt szokatlan a tárgy oktatási módszere?	3	2	3	1	2	2	4	4	2,625
Mennyire érezte jónak a tárgy oktatási módszerét?	5	3	5	5	5	4	5	5	4,625
A tárgy eredményessége a hallgató szempontjából									Átlag
Utólag, hogyan értékelné a tárgy eredményességét a saját maga számára?	5	3	5	5	5	4	5	5	4,625
Mennyire volt élményszerű az oktatás saját maga számára?	5	4	5	5	5	5	5	5	4,875
Mennyire felelt meg az előzetes elvárásainak a tárgy?	4	4	5	5	5	4	5	5	4,625
A tárgy tartalmi nehézségének megítélésének változása									Átlag
Előzetesen mennyire gondolta nehéznek, nehezen tanulhatónak a tárgy tartalmát?	3	3	3	1	3	3	4	4	3
Utólag mennyire tartja nehezen megtanulhatónak a tárgy tartalmát?	2	2	2	1	2	2	2	3	2

A 14 fős hallgatói csoport több mint fele, 8 fő adott választ a kérdéseinkre (57%), ami statisztikai szempontból nem mondható kifejezetten jónak. Általánosságban azonban jellemző, hogy a tárgyakhoz köthető kérdőíves felmérések kitöltésére nehéz rávenni a hallgatókat: az egyetemi rendszer automatikusan generál minden kurzus után kérdőívet, de ezek kitöltési hajlandósága és az értékelhető válaszok aránya csekély.

A válaszokból mégis sikerült néhány következtetést levonni, melyek között voltak meglepőek, és az elképzeléseinket visszaigazoló vélemények is. Első kérdéscsoportunk a módszertanra adott visszajelzést vizsgálta. Ebből számunkra kiderült, hogy kevésbé volt a hallgatók számára idegen a módszertan, annak ellenére, hogy más tárgyakban nem találkozhattak ezzel az oktatási módszerrel. Összességében kifejezetten jónak és kellemesnek ítélték meg az alkalmazott módszert. A kérdésekre adott 2,6-4,6 feletti átlagpár jól mutatja a „nem szokatlan” és „jó módszer” megítélést. Második kérdéscsoportunk a tárgy tartalmával és az elsajátított ismeretekkel volt kapcsolatos. Ebben a kérdéskörben minden válaszadó hallgató teljesen pozitívnak értékelték a tantárgyat (4,6 feletti átlagot számoltunk). Harmadik egyben utolsó kérdéspárunk, a tárgyat megelőző és a tárgy hallgatása utáni megítélésre, pontosabban a tárgy nehézségének megítélésére kereste a választ. Talán itt ért minket a legnagyobb meglepetés abban a tekintetben, hogy előzetes feltevésünk szerint a hallgatók

jóval nehezebbnek ítélik előzetesen a tárgyat. A válaszok alapján csak viszonylag kis mértékű változásként realizálódott, előzetesen nagyobb mértékű változást vártunk: átlagban 3-ra értékelték az előzetes nehézségét a tárgynak és 2-re ugyanezt utólagosan. Minimálisan, de jelzi azt a változást, amelyet kívántunk elérni, miszerint a nehezebb tárgyak elsajátítását is a hallgató pozitív élménnyel zárhatja és eredményes tanulási érzések maradjanak benne.

A félév végén a hallgatók az online bemutatók anyagából önszorgalomból egy video montázst is készítettek, melyet közzétettek a YouTube video-megosztó oldalon<sup>2</sup>.

## Összegzés

Jelen cikk bemutatott egy modern projektalapú megújítását egy gyakorlatorientált műszaki tárgy oktatásának. A tárgy során jelentős hangsúlyt kapott az élményszerű, kísérletező megközelítés a robotikához köthető tudás átadásában. Az oktatási technika kiemelte a robotika interdiszciplináris voltát és az egyes részterületek sajátosságait egy alapvetően projektalapú szemléletben. A szükséges elméleti alapok átadása hagyományos orális előadások formájában történt meg, míg a gyakorlati tudás megszerzése a hallgatók önálló munkáját igényelte. A tárgy módszertana bár a megszokottól eltérő volt, a hallgatók – visszajelzéseik alapján - nem érezték szokatlannak, alapvetően élményszerű volt a tudás megszerzése. Összességében elmondható, hogy a cikkben bemutatott megközelítés lehetővé tette a személyre szabott tanulást és megfelelő alapokat adott egy komplex, műszaki témában.

## Bibliográfia

- [1] Péter, Bernát, és Zsakó László. „Programozás tanítási módszerek- stratégia a kezdetekre”. Informatika szakmódszertani konferencia, 2019.
- [2] Nagy József. „XXI. század és nevelés”. Budapest: Osiris Kiadó, 2000.
- [3] Piaget Jean. „Az értelem pszichológiája”. Budapest: Gondolat, 1993.
- [4] János Kata. „Korszerű módszerek a szakképzésben”. Typotex Kft, 2007.
- [5] Molnár György. „A technológia és hálózatalapú alapú tanulási formák és attitűdök az információs társadalomban, különös tekintettel a felsőoktatás bázisára”, Információs Társadalom, 2012. <https://doi.org/10.22503/inftars.XII.2012.3.4>
- [6] Michel Olivier. „WebotsTM: Professional Mobile Robot Simulation”. International Journal of Advanced Robotic Systems. 1. <https://doi.org/10.5772/5618>.
- [7] Solymos Dóra. „LEGO robotok felhasználási lehetőségei az oktatásban”. Informatika szakmódszertani konferencia, 2019.
- [8] Yuya Maruyama, Shinpei Kato, és Takuya Azumi. 2016. Exploring the performance of ROS2. In Proceedings of the 13th International Conference on Embedded Software (EMSOFT ,16). Association for Computing Machinery, New York, NY, USA, Article 5, 1–10. <https://doi.org/10.1145/2968478.2968502>

---

2 <https://www.youtube.com/watch?v=40BzH4AH0Tw>

## E-tananyagfejlesztés virtuális 3D környezetben

T. Nagy László

*Debreceni Református Hittudományi Egyetem*[t.nagy.laszlo@drhe.hu](mailto:t.nagy.laszlo@drhe.hu)

Boda István Károly

*Debreceni Református Hittudományi Egyetem*[boda.istvan@drhe.hu](mailto:boda.istvan@drhe.hu)

Tóth Erzsébet

*Debreceni Egyetem*[toth.erzsebet@inf.unideb.hu](mailto:toth.erzsebet@inf.unideb.hu)**Absztrakt**

Az oktatás és a tanulás hatékony elektronikus tananyagai iránti igény az utóbbi időben szinte minden tudományterületen jelentősen megnőtt köszönhetően az elmúlt időszakban teret nyerő digitális oktatásnak.

Tanulmányunkban egy új és reményeink szerint ígéretes technológiát szeretnénk bemutatni az e-tananyagok fejlesztésére és prezentálására, amely hatékonyan alkalmazható különféle tanulási környezetekben. Ez a technológia a MaxWhere rendszer oktatási célú felhasználásán alapul. A MaxWhere egy néhány évvel ezelőtt kifejlesztett 3D-s prezentációs eszköz, amely látványos és könnyen használható három dimenziós virtuális környezetet kínál a megjeleníteni kívánt tartalmaknak. A rendszert kiválóan lehet használni az e-tananyagok fejlesztéséhez és bemutatásához mind a tanítás, mind a tanulás területén. Esetünkben a tananyag fejlesztésének és prezentálásának lehetőségeit egy kiválasztott informatikai téma kidolgozásával szeretnénk bemutatni és elemezni a MaxWhere 3D virtuális környezetében.

**Development of e-curricular material in a virtual 3D environment****Abstract**

The need for effective electronic curricular material in instruction and learning has in almost all fields significantly increased due to digital education that has recently expanded everywhere.

In this study, new and hopefully promising technology will be offered on how to develop and present e material that can be used effectively in different learning environments. This technology is based on the Maxwhere system applied for instructional use. The Maxwhere has been developed a few years ago as a 3D presentation tool that offers a spectacular and easily applicable 3D virtual environment for all contents to be presented. The system can be well used to develop and present e-materials in both teaching and learning. In our case, the possibilities of developing and presenting the curricular material will be shown with the help of a selected informatics topic which is elaborated and analyzed in a Maxwhere 3D virtual environment.

**Keywords:** multimedia, 3D space, virtual space, development of curricular material, Maxwhere



## Bevezetés

Az oktatás és a tanulás *hatékony elektronikus tananyagai* iránti igény az utóbbi időben szinte minden tudományterületen jelentősen megnőtt. Ez javarészt a világméretű COVID 19-járvány miatt bevezetett digitális oktatásnak köszönhető. Tanulmányunkban egy új és ígéretes technológiát szeretnénk bemutatni az e-tananyagok fejlesztésére és prezentálására, amely hatékonyan alkalmazható különféle oktatási vagy tanulási környezetekben.

A *MaxWhere Seminar System* egy néhány évvel ezelőtt Magyarországon kifejlesztett 3D-s prezentációs eszköz, amely látványos és könnyen használható háromdimenziós virtuális környezetet kínál keretként a megjeleníteni kívánt tartalmaknak. (MaxWhere 2021) A rendszert lehet használni e-tananyagok fejlesztéséhez, bemutatásához és böngészéséhez mind a tanítás, mind az önálló tanulás területén. A MaxWhere környezet igen sokoldalúan használható az oktatás keretein túl is, például marketing, vagy gazdasági területeken prezentációk, ismertetőik stb. bemutatására.

A MaxWhere környezet alapvető funkciói közül az alábbiakat szeretnénk röviden kiemelni:

- 3D tér: A környezetben kialakított térben lehetőség van 3 dimenzióban mozogni;
- Rugalmas irányítás lehetősége;
- A falakon lévő ún. *smartboard* táblákon lehet az információkat megjeleníteni, mintegy kiállítás-szerűen;
- Széleskörű kompatibilitás a különböző offline és online, statikus és dinamikus tartalmakkal;
- Prezentáció automatizálhatósága, több nézőpontból való megtekintés;
- A *smartboard* táblákon megjeleníthetők a szöveges tartalmak, videók, weboldalak, webalkalmazások, és különféle kollaborációs eszközök;
- Számos előre elkészített, különböző megjelenésű virtuális tér közül választhatunk.

## Tanulás és hálózatelmélet

Hálózatelméleti megközelítésben a tanulás során egyrészt új csomópontok adódnak hozzá a meglévő tudásrendszerhez, másrészt a csomópontok között új kapcsolatok (élek) kialakítása történik. Nyitott kérdés, hogy hálózatelméleti megközelítésben figyelembe kell-e venni a csomópontok közötti kötések erősségét vagy sem – ha az emberi agyra a konnekcionista („neurális háló”, „deep learning” stb.) metaforát alkalmazzuk, a válasz igen, ha viszont a számítógépes rendszer vagy (skálafüggetlen) hálózat (internet, web stb.) metaforát, akkor a válasz nem. A verbálisan szerveződő tudáshálóban a csomópontok a kisebb-nagyobb jelentéssel bíró egységek, „modulok” lehetnek (pl. adatok, fogalmak, kifejezések, mondatok, szövegek stb., ill. verbálisan értelmezett képek, hangok stb.). A csomópontok közötti élek ebben a megközelítésben a csomópontok között kialakított jelentésszervező (szemantikai, korreferenciális, intertextuális stb.) kapcsolatok, valamint a „szerteágazó” asszociatív kapcsolatok lehetnek. A tanulással folyamatosan építjük és/vagy karbantartjuk a tudáshálónkat. Ennek ellentétes folyamata a felejtés, amikor a már felépült rendszer kötéseinek száma csökken (vagy konnekcionista szemléletben „elhalványul”) vagy a kötések teljesen megszűnnek.

Ebben a megközelítésben könnyen belátható, hogy egy hálózat azon csomópontjai, amelyek kevés kapcsolattal rendelkeznek, könnyebben sérülhetnek („leválhatnak”), azaz a felejtés homályába veszhetnek. Ezzel szemben azon csomópontok, amelyek sok (és akár többszörös, redundáns) kapcsolattal rendelkeznek, sokkal ellenállóbbak a kapcsolatok sérülésével, azaz a felejtéssel szemben. Példaként említhető a „magolás” mint tanulásforma, ahol a tananyag egyes információelemeit úgy próbálja a tanuló megjegyezni, hogy nem, vagy csak nagyon hiányosan kapcsolja őket a meglévő ismereteihez. Mint tudjuk, ez a tudás hosszú távon



kevésbé maradandó; ezért mondhatjuk, hogy sérülékeny, hiszen a bemagolt információelemek között kevés kapcsolat van, vagy egyáltalán nincs.

Az „értelmes” tanulással ún. kognitív tartalékot képezünk agyunkban. SM betegeknél például a feltételezések szerint az iskolázottság bizonyos mértékben megvédhet a hanyatlástól (Esem. hu 2021/1/2). Egyes kutatások azt állapították meg, hogy a felejtés maga nem az információ (ill. adat) elvesztését jelenti, hanem az információ elérési képesség sérülését, vagy hiányát.

A közelmúltban végzett kísérletekben a Nobel-díjas biológus Tonegawa kutatócsoportjának sikerült elfeledett emlékeket újraéleszteni egerekben, amellyel azt bizonyítják, hogy nem az emlékképek sérülnek az agyban, hanem azok előhívási képessége (Ramirez et al. 2013). Ezzel összhangban áll az, hogy a vizsgálatok eredménye szerint az agysérülés, stressz, vagy az Alzheimer-kór következtében fellépő retrográd amnéziát az emlékek előhívási problémái okozzák (Ryan et al. 2015). Ezeknek az eredményeknek a birtokában feltételezhetjük, hogy az „elfelejtett” információt reprezentáló csomópontok nem törölődnek az agyból, pusztán azok az utak vesznek el, ill. sérülnek, amelyeken keresztül elérhetjük őket. (Ezt támasztja alá az is, hogy pl. retrográd amnézia esetén az időlegesen elfelejtett emlékek később visszatérhetnek.)

A virtuális 3D környezetben megvalósuló tanulásakor feltételezzük, hogy a tanulás az általunk használt virtuális környezetet valamilyen formában *leképezi* a tanuló mentális, kognitív rendszerébe. A tananyag tervezésekor és kialakításakor számunkra egy olyan hálózati modellre volt szükség, amely a tudásrendszer csomópontjainak és éleinek jelentést tulajdonít, és valamilyen formában jellemzi a kialakított hálózati architektúrát. Az alábbiakban körvonalazunk egy ilyen modellt.

Az internet mint spontán módon szerveződő információs hálózat fejlődése során folytonosan jelennek meg, ill. tűnnek el csomópontok („weblapok”) és/vagy a csomópontok között kialakított kapcsolatok (élek, „linkek”). Ebben a verbálisan és multimédiálisan szerveződő tudáshálóban a csomópontok jelentik a jelentéssel bíró egységeket (pl. numerikus vagy szöveges adatok, verbális és multimédiális szövegek, képek, videók stb.), a csomópontok közötti élek pedig szerteágazó (szemantikai, „asszociatív” stb.) kapcsolatokat valósítanak meg, amelyek lehetővé teszik a csomópontokban tárolt információ visszakeresését. Azonban a spontán szerveződés egyáltalán nem jelenti azt, hogy a világháló véletlenszerűen szerveződik. Az interneten léteznek „kis világok” (például a Wikipedia oldalai), olyan csomópontok, amelyek között nagy számú kölcsönös kapcsolatot találunk, és vannak „összekötők” (vagy középpontok, pl. az internetes keresők, lásd Google), amelyek rendkívül nagy számú kapcsolattal rendelkeznek, és – egyebek közt – összeköttetést biztosítanak a „kis világok” között. Általánosan fogalmazva a középpontok „a rendszer bármely két pontja között rövid utakat létesítenek” (Barabási 2013: 72).

Egy elektronikus tananyag kifejlesztésekor az egyes anyagrészeket csomópontoknak és a közöttük kialakított kapcsolatokat éleknek vagy linkeknek tekintve egy *tudáshálózatot* alakítunk ki. Ha az internetet a tanulás során kialakított tudáshálózat metaforájának tekintjük, akkor a *tanulást* úgy közelíthetjük meg, mint egy olyan mentális tevékenységet, amelynek során folyamatosan építjük és/vagy karbantartjuk a tudáshálózatot. Ebben a megközelítésben a tanulás az elektronikus tananyag *leképezését* valósítja meg a tanuló kognitív rendszerébe, vagyis a tananyag mentális képét hozza létre az emberi hosszú távú memóriában. Ez azonban nem egy az egyben történő megfeleltetést jelent, mivel a leképezett tudásanyag összekapcsolódva a memóriában tárolt további tudástartalmakkal (például a háttértudással) további csomópontokat kapcsol be a rendszerbe, új linkeket alakít ki (miközben a megszerzett tudás „elmélyül”), és az is lehetséges, hogy a tanuló egyes tudáselemeket átértelmez, vagy éppen elutasít. Mindezt visszacsatolva a tananyagfejlesztésbe az elektronikus tananyag tartalma és hatékonysága folyamatosan fejleszthető.

## Tanulás 3D virtuális környezetben

A következőkben a virtuális 3D környezet néhány pozitívumát szeretnénk röviden bemutatni a tananyag megjelenítése, elrendezése, és használata szempontjából.

Általánosságban elmondható, hogy a 3D környezetnek számos – jelentős – előnye lehet az etananyagok szempontjából, amelyek közül emeljük ki a következőket:

- A tananyagok szegmentálása, elrendezése és tartalmi szervezése szempontjából a virtuális 3D tér kibővített lehetőségekkel rendelkezik a szokásos 2D környezetekhez.
- A tananyag virtuális 3D környezetben történő bemutatása vélhetően izgalmasabb és ezért könnyebben emészthető az átlag tanuló számára, mint a megszokott, és sok esetben „egysíkú” tartalom, amelyet a hagyományos médiák kínálnak.
- A tananyag megfelelő elrendezése és megjelenítése a 3D térben nagymértékben segítheti a tájékozódást és a tanulás kognitív folyamatát, pl.
  - az információk észlelése, és ezáltal a tananyag értelmezése könnyebb lehet;
  - a 3D térben *egyidejűleg* megjelenített tartalmi egységek közötti kapcsolatok nagymértékben segíthetnek az értelmezésben;
  - a virtuális térben bemutatott tananyag vizuális rögzítése **automatikus folyamat** lehet.
- A 3D környezet által nyújtott fokozott vizuális élmény különösen motiválhatja az „új” fiatal generációk tagjait (pl. „net”-generáció, Z generáció, kognitív entitások generációja<sup>1</sup>).
- Elméletileg a tananyag felépítése a releváns kognitív egységek hálózataként vizsgálható (például fejezetek, bekezdések, hivatkozások, diagramok stb.). Ebből a szempontból a tananyag egyes információegységei minél több kapcsolatban vannak egymással explicit vagy implicit módon, a befogadók annál több kapcsolatot hozhatnak létre saját kognitív rendszerükben a befogadni kívánt tananyaggal kapcsolatban.
- Következésképpen a memorizált információk szintetizálása, megőrzése és visszahívása sokkal hatékonyabb lehet, ha az egyes elemek vagy részek közötti kapcsolatok (tudáskapcsok, értelmi/értelmezési kötések) száma magasabb, és a memorizálandó egységek közötti (bejárási)utak rövidebbek.

## Tananyagfejlesztés 3D környezetben

Cikkünkben a tananyag fejlesztésének és prezentálásának lehetőségeit egy kiválasztott informatikai téma kidolgozásával szeretném bemutatni a MaxWhere Seminar System 3D virtuális környezetében (1. ábra).

---

1 generation of cognitive entities (vö. Baranyi–Csapó–Sallai 2015: vii et passim)

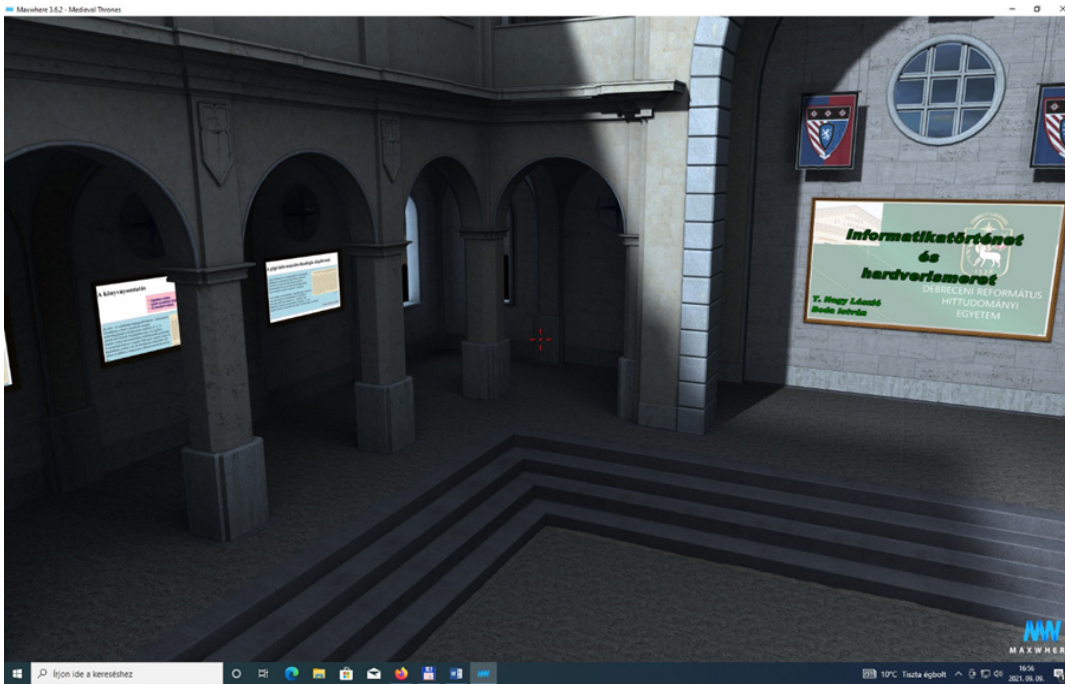


1. ábra. A MaxWhere virtuális 3D környezetében fejlesztett informatikai tananyag kezdőképernyője

A kiválasztott témánk egy, a Debreceni Református Hittudományi Egyetem Kölcsey Ferenc Tanítóképzési Intézetének Matematika és Informatika Tanszéke által gondozott tantárgy tananyaga, mely az informatikai műveltségterület képzési hálójában szerepel. A tantárgy címe *Informatikatörténet és hardverismeret*. A tárgy tananyaga az előzőekben már létezett elektronikus formában MS PPT prezentációként. Ahhoz hogy a MaxWhere 3D rendszerébe betölthető legyen, előbb elkészítettük az eredeti tananyag hipertextes (HTML/CSS) változatát, amivel máris átkerült az offline térből az online világba.

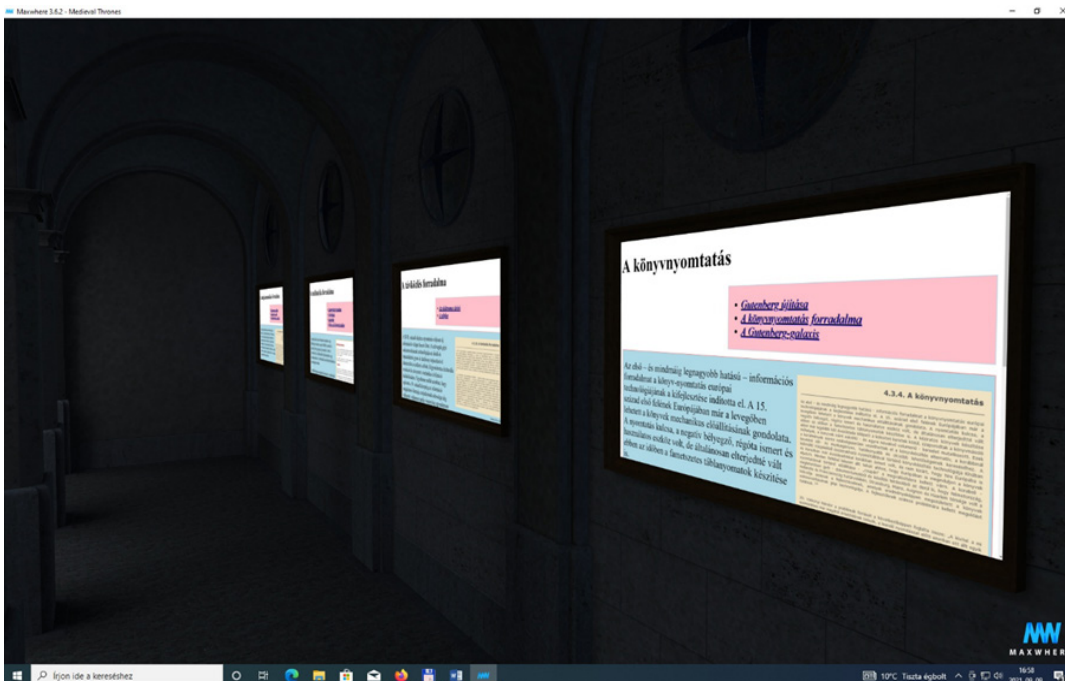
A fejlesztésben a tantárgy tananyagának 26 fejezete készült el weboldalként, HTML formátumban, amelyeket külön böngészőablakokban jelenítettük meg (2. és 3. ábra). A MaxWhere-ben való megjelenítés nem feltétlenül igényli a háttérben a HTML formátumot, a keretrendszer ugyanis lehetővé teszi számos, különböző kiterjesztésű állomány megjelenítését a virtuális 3D térben elhelyezett böngészőablakokban (ún. *smartboardok*ban), így akár közvetlenül egy PPT prezentációt is megjeleníthetünk. A weblap fejlesztésére fordított plusz munkát két okból éreztük indokoltnak. Az egyik ok, hogy a HTML oldalak könnyen, bármikor elérhetőek az érdeklődők (például a hallgatók) számára, és az ilyen formában elkészített online tartalom bármikor, a 3D környezettől függetlenül is fejleszhető, változtatható. A másik ok az, hogy a MaxWhere virtuális környezet igazi – vagy mondhatjuk teljes – *interaktivitását* a weblapok interaktív lehetőségei révén tudjuk igazán kihasználni, és érzékelni.





2. ábra. A MaxWhere virtuális 3D környezetében fejlesztett informatikai tananyag két fejezete a 3D tér földszinti folyosóján a kezdőképernyőtől balra

A hipertextualizálás már önmagában is egy új dimenziót adott az eredeti nyersanyagnak, hiszen a „passzív”, főleg lapozható diák helyett egyebek közt interaktív tartalomjegyzéket, hivatkozásokkal és képekkel ellátott hiperszöveget alkottunk. A linkeken keresztül a tananyag tartalmi egységeit összekapcsoltuk, és emellett számos kiegészítő információ és metaadat is elérhetővé vált.



3. ábra. A MaxWhere virtuális 3D környezetében fejlesztett informatikai tananyag négy fejezete a 3D tér földszinti folyosóján egy másik perspektívából



A tananyag „természetes”, a tartalomjegyzék sorrendjében történő szekvenciális bejárása az általunk javasolt tanulási sorrendet követi, de maga az anyag elvileg tetszőleges sorrendben bejárható a felhasználók számára. Ez a tanulás révén – az előzőekben már részletesebben taglalt - megszerzett tudás hatékonyságát jelentősen növeli, mert a kialakított tudásháló annál erősebb, „robusztusabb”, minél nagyobb redundanciát tartalmaz.

### Összegzés, következtetések

A tananyagfejlesztés során a *tudáshálózat metaforából* indultunk ki, amely az emberi (szubjektív) tudás szerveződését az internethez hasonló komplex, skálafüggetlen hálózatok szerveződéséhez hasonlítja. A tudáshálózat metafora alapján a tanulást úgy képzelhetjük el, hogy az agy komplex tudáshálózatában folyamatosan új tudáselemek, csomópontok adódnak hozzá a meglévő rendszerhez, amelyek a már létező elemekhez kapcsolódnak, továbbá a már meglévő csomópontok közt is keletkezhetnek új kapcsolatok. *A tananyagfejlesztés során egy olyan rendszert próbálunk kialakítani, amely követi a fenti mintát.* A skálafüggetlen módon szerveződő komplex hálózatok viszonylag ellenállók azzal szemben, ha véletlenszerűen veszünk ki egyes elemeket (pl. csomópontokat, éleket; vö. Barabási 2013: 217), ezért ha a tanulás során egy hasonló struktúrát képezünk le az emberi agyba, feltehetjük, hogy az elsajátított tudás viszonylag stabilan „ellenáll” majd a felejtésnek. Ebből a szempontból különösen fontos, hogy a tananyagban szereplő egyes tudáselemek közt minél több (értelmes, releváns) kapcsolatot hozzunk létre. (Juhász et. al. 2016)

*A virtuális valóság a Dale féle tapasztalati (tudás- vagy tanulási) piramisban (Cone of Experience) különösen értékes tapasztalatszerzési formaként szerepel, amely jellegéből fakadóan összekapcsolja a („virtuálisan”) közvetlen tapasztalatszerzést az absztrakt tartalmak verbális és/vagy multimediális megjelenítésével. (Dale 1969) Ezáltal a virtuális valóságra a modell alapján mind a tudásátadás, mind és információ-megőrzés szempontjából rendkívül hatékony rendszerként tekinthetünk.*

### Irodalomjegyzék

Barabási, Albert-László 2013. *Behálózva. A hálózatok új tudománya.* [Linked. The New Science of Networks.] Budapest: Helikon K.

Baranyi Péter – Csapó Ádám – Sallai Gyula 2015. *Cognitive Infocommunications (CogInfoCom).* Springer. <https://doi.org/10.1007/978-3-319-19608-4>

Bates, A.W. (Tony) 2019. *Teaching in a digital age. Guidelines for designing teaching and learning.* Vancouver BC: Tony Bates Associates.  
<https://pressbooks.bccampus.ca/teachinginadigitalagev2/> (2022-02-04)

2.6. Connectivism.

<https://opentextbc.ca/teachinginadigitalage/chapter/3-6-connectivism/> (2022-02-04)

Boda István, T. Nagy László (2021) Informatikatörténet és hardverismeret, DRHE 2021  
<http://bodaistvan.hu/infalap/infotort/infotort.html> (2022 05 30)

Dale, Edgar 1969. *Audiovisual Methods in Teaching.* New York: Dryden Press.

Esem.hu 2018. 04. 17. Az iskolázottság megvédhet a kognitív hanyatlástól.

<https://esem.hu/az-iskolazottsag-megvedhet-a-kognitiv-hanyatlastol/> (2022-02-04)



- Esem.hu 2017. 07. 06. Használd az agyadat, hogy késleltesd a progressziót!  
<https://esem.hu/hasznald-az-agyadat-hogy-kesleltesd-a-progressziot/> (2022-02-04)
- Juhász Sándor – Elekes Zoltán – Gyurkovics János 2017. A tudáshálózatok időbeli változásának vizsgálati lehetőségei. In: Lengyel I. (szerk.) 2017. *Két évtizedes a regionális tudományi műhely Szegeden: 1997–2017*. Szeged: JATEPress, 161–175.
- Maxwhere (2021) MaxWhere - Virtual spaces with the benefits of reality.  
<https://www.maxwhere.com/> (2022-02-04)
- Ramirez, S. – Liu, X. – Lin, Pei-A. – Suh, J. – Pignatelli, M. – Redondo, R. L. – Ryan, T. J. – Tonegawa, S. 2013. Creating a False Memory in the Hippocampus. *Science* Vol 341. No. 6144. 387–391.  
<https://tonegawalab.mit.edu/research/publications/> (2022-02-04)
- Ryan, T. J. – Roy, D. S. – Pignatelli, M. – Arons, A. – Tonegawa, S. 2015. Engram cells retain memory under retrograde amnesia. *Science* Vol 348. No. 6238. 1007–1013.  
<https://tonegawalab.mit.edu/research/publications/> (2022-02-04)

Digitális átállás – Minőség – lehetőségek az EQAVET terén  
Digital Transition – Quality – Opportunities in EQAVET

Palencsárné Kasza Marianna

*Digitális Jólét Nonprofit Kft., Digitális Szakképzési és Felnőttképzési Módszertani Központ  
(Budapest)*

[kasza.marianna@djnkft.hu](mailto:kasza.marianna@djnkft.hu)

**Absztrakt:** A 21. század új kihívása, hogy az ipar 4.0 követelményeihez igazított szakképzési rendszert kell kialakítani. A magyar kormány által elfogadott Szakképzés 4.0 stratégia a gazdaság változó kihívásaira reagál, pontos képet ad a helyzetről és felvázolja a tervezett beavatkozásokat. A szakképzés minőségének javítása érdekében valamennyi szakképző intézményben minőségirányítási rendszert vezetnek be, amelynek elemei kompatibilisek lesznek az EQAVET rendszerrel, és azonos indikatív jellemzőket használnak, így az intézmények tevékenysége és eredményei európai szinten összehasonlíthatóvá válnak. Az EQAVET magyarországi bevezetése során a digitalizációs követelmények minden önértékelési ciklusba bekerültek. Az intézmények és a tanárok értékelését és fejlődését támogatja a Digitális Névjegyrendszer (DNR), az innovatív oktatási technológiák használatának elősegítésével megvalósuló hatékony tanulás önértékelési eszköze (SELFIE) és a DigCompEdu önértékelési rendszer.

**Abstract:** A vocational training system adapted to the requirements of Industry 4.0 needs to be developed. The Vocational Training 4.0 Strategy adopted by the Hungarian Government responds to the changing challenges of the economy, providing a precise picture of the situation and outlining the planned interventions. To improve the quality of vocational education and training, a quality management system will be introduced in all vocational education and training institutions, the elements of which will be compatible with the EQAVET system and will use the same indicative characteristics, thus making the activities and results of the institutions comparable at European level. During the implementation of EQAVET in Hungary, digitalisation requirements have been included in every cycle. The Digital Badge System (DBS), the Self-reflection on Effective Learning by Fostering the Use of Innovative Educational Technologies (SELFIE) and the DigCompEdu self-assessment system support the assessment and development of institutions and teachers.

**Keywords:** Vocational Education and Training 4.0; intervention, EQAVET, SELFIE, DigCompEdu, digitalisation, quality management

## 1. Bevezetés

A cikk célja, hogy a jelenleg meglévő jogszabályok, digitális fejlesztési célok, eszközök és a minőségi kritériumok lehetséges összehangolását bemutassa. A 21. század fontos kihívása a szakképzés területén is a digitalizáció és a szakképzés egyre magasabb szinten való megvalósítása. Ehhez kíván áttekintést adni az olvasó részére, hogy a jelenleg is hatályos rendelkezések hogyan támogatják az oktatás digitalizációt és a minőségi oktatás

lehetőségeit. A cikk bemutatja, hogy a Magyarországon elfogadott Szakképzés 4.0 stratégia tartalmi jellemzői az Európai Unióban kiadott szakmai ajánlások és a 2022. szeptember 1-től bevezetett EQAVET elveken alapuló minőségirányítási rendszer együttesen hogyan tudja összehangolni a szakképzés digitalizációját. Bemutat néhány eszközt, ami alkalmas a szakképző intézmények digitalizációjának szintjének mérésére. A mérési eredmények beépítésével pedig a minőségfejlesztésre.

A szakképzés alapvető feladata, hogy a képzett szakemberek a munkaerőpiacon megállják a helyüket, azaz a vállalkozások működésében hasznosítható tudással rendelkező szakembereket bocsásson ki a munkaerőpiacra.

Az Európai Parlament 2016-ban megfogalmazott állásfoglalása szerint: „Az ipar 4.0 a termelési folyamatok olyan szervezését írja le, melynek keretében az eszközök önállóan kommunikálnak egymással az értéklánc mentén: a jövő egy olyan „okos” gyárat hozva létre ezzel, amelyben a számítógép-vezérelt rendszerek nyomon követik a fizikai folyamatokat, létrehozzák a fizikai valóság virtuális mását és decentralizált döntéseket hoznak önszervező mechanizmusok alapján.”[6]

A szakképzésnek alkalmazkodnia kell az új kihívásokhoz, vagyis fel kell készülnie arra, hogy a szakmával és szakképesítésekkel kibocsátott szakemberek rendelkezzenek az ipar 4.0 elvárásait teljesíteni tudó 21. századi kompetenciákkal.

A „Szakképzés 4.0 - a szakképzés és felnőttképzés megújításának középtávú szakmapolitikai stratégiája, a szakképzési rendszer válasza a negyedik ipari forradalom kihívásaira” stratégiát 2019-ben fogadta el a Kormány. [11] A stratégia tartalmaz egy helyzetértékelést és oktatási helyzetképet a szakképzésre és a felnőttképzésre vonatkozóan, melyben a szakképzési és felnőttképzési rendszer bemutatása, és a rendszer legnagyobb kihívásai és kulcsproblémái kerültek leírásra. Bemutatja a 2020. szeptember 1-től életbe lépő új szakképzési és felnőttképzési rendszert és a fejlesztési célkitűzéseket, valamint konkrétan meghatározott 40 db beavatkozást tartalmaz, mely mindegyikéhez eredmény-indikátor is meghatározásra került.

A Szakképzés 4.0 stratégiában a következő beavatkozások vonatkoznak a digitalizáció támogatására:

- 3.: Minden képzésbe be kell épülnie az Ipar 4.0 követelményeinek és a szakmában elvárt digitális tartalmaknak
- 14.: Élményalapú pálya-tanácsadási rendszer működtetése a Digitális Közösségi Alkotóterek lehetőségeit is bevonva.
- 17.: Digitális eszközök és korszerű oktatási technológiák fejlesztése
- 18.: Minden ágazatban digitális tananyagok beszerzése és korszerű, aktuális állapotban tartása
- 34.: A pedagógus-továbbképzések megújítása a szakmai képzésben: vállalati helyszíni szakmai képzések, a digitális ismeretek és a módszertani kultúrájának fejlesztő továbbképzések erősítése
- 40.: A szakmai képzettséget szerettek nyomon követése a Digitális Munkaerő-piaci Program segítségével

A fenti 6 beavatkozás jól szemlélteti, hogy a szakképzésre és a felnőttképzésre vonatkozóan elkezdődött a digitális módszerek, eszközök és technológiák beépítése, ami a munkaerőpiaci elvárások miatt is fontos kritérium és elvárás, hiszen a jövő szakemberei így tudják majd saját szakmai tudásukat felhasználni az innováció jegyében. A fenti célokat támasztja alá a 2019-ben elfogadott, a szakképzésről szóló 2019. évi LXXX. törvény preambuluma is, melyben szintén szerepel a digitális tudás fejlesztése.



„Az Országgyűlés ...

- a tudásalapú gyakorlati tanulásra, a digitális tudás fejlesztésére, a kreatív és széles látókörű gondolkodásra, továbbá a változásokhoz való rugalmas alkalmazkodás képességére oktató, a szakképzés és a gazdasági szféra közötti együttműködés erősítésére alapozó, valamint valós piaci esélyt és biztos megélhetést biztosító szakképzettség, illetve szakképesítés megszerzését elősegítő szakképzési rendszer kialakítása céljából” [2]

## 2. Minőségirányítás a szakképzésben – EQAVET (szakképzés Európai Minőségbiztosítási Referencia Keretrendszere – European Quality Assurance Reference Framework)

A Szakképzési törvény előírja, hogy a szakképző intézmény és a duális képzőhely a tevékenységét minőségirányítási rendszer alapján végezze, melynek kötelező bevezetését 2022. augusztus 31-éig írja elő. [2]

A minőségirányítási rendszer jellemzői az EQAVET alapján kerültek kialakításra. Az Európai Bizottság javaslatára az Európai Parlament és a Tanács 2009. június 18-án egy új európai jogszabályt – ajánlást – fogadott el a szakképzés Európai Minőségbiztosítási Referencia Keretrendszerének létrehozásáról, melyben meghatározták az EQAVET alapelveket. [5]

A következő lépés a Bruges-i Közlemény: „A szakképzés kiválóságának, minőségének és relevanciájának előmozdítása érdekében fejleszteni kell a nemzeti minőségbiztosítási rendszereket. A szakképző intézményeknek visszajelzést kell kapniuk arról, hogy a szakképzés keretében oklevelet szerzett diákok milyen mértékben kerülnek alkalmazásra.” [7]

A Rigai Szakképzési Nyilatkozat [4]: az oktatás és képzés terén folytatott európai együttműködés kiemelt területei között szerepel a szakképzésre és a felnőttkori tanulásra vonatkozó 2020-ig megvalósítandó prioritások egyikeként az oktatás és képzés minőségének és hatékonyságának javítása.

Az EQAVET mint kidolgozott keretrendszer az Európai Unió Tanácsának ajánlása a fenntartható versenyképességet, a társadalmi méltányosságot és a rezilienciát célzó szakképzésről dokumentumban jelenik meg. [3] Az ajánlás melléklete az EQAVET indikatív (elvárások) jellemzőit és a javasolt mutatókat tartalmazza, mely alapján a magyarországi minőségirányítási rendszer kidolgozására is sor került.

Az EQAVET a minőségbiztosítási, minőségirányítási rendszerekre jellemző PDCA<sup>1</sup> ciklus szerint 4 szakaszra bontja a szakképző intézmények tevékenységét. A fogalmakat az „Önértékelési kézikönyv szakképző intézmények számára” következőképpen határozza meg. [8]

- **Tervezés (P):** A tervezés tükrözi az érdekelt felek közös stratégiai jövőképét, és egyértelműen meghatározott célokat/célkitűzéseket, intézkedéseket és mutatókat tartalmaz.
- **Megvalósítás (D):** A megvalósítási terveket az érdekelt felekkel konzultálva állítják össze, és egyértelműen meghatározott elveket tartalmaznak.
- **Értékelés (C):** Az eredményeket és a folyamatokat rendszeresen ellenőrzik, értékelik, és ezt mérésekkel támasztják alá.
- **Felülvizsgálat (A):** Az intézmény hasznosítja az értékelés eredményeit, megvalósítja a szükséges visszacsatolásokat és fejlesztéseket.

1 PDCA: P-plan (tervezés), D-do (megvalósítás), C- chek (ellenőrzés, értékelés), A-act (felülvizsgálat, beavatkozás)

## 2. Az EQAVET és a digitalizáció kapcsolata

Az „Önértékelési kézikönyv szakképző intézmények számára” módszertani útmutató indikatív jellemzőket határoz meg a 4 részfolyamathoz kapcsolódóan, melyben a digitalizációval kapcsolatos elvárások is megjelennek.

A tervezés szakaszában az egyik indikatív jellemző: a szakképző intézmények által kitűzött helyi célokban tükröződnek az európai, országos és regionális szakképzési szakpolitikai célok. Az országos stratégiát a SZAKKÉPZÉS 4.0 -át alapul véve a fentebb felsorolt 6 db beavatkozás, mint a digitalizációt képviselő szempontot is figyelembe kell vennie az intézményeknek.

A megvalósítás szakaszában az egyik indikatív jellemző: a szakképző intézmények a digitális technológiák és az online tanulási eszközök használatának segítségével előmozdítják az innovációt a tanítási és tanulási módszerek terén, az iskolában és a munkahelyen egyaránt. A megvalósítás során az oktatás folyamatába kell beépíteni a digitális technológiákat és az online tanulási eszközöket, mely a tanításra és tanulásra egyaránt hatással lesz.

Az értékelés szakaszának egyik indikatív jellemzője: a nemzeti és regionális szabályozás/ keretrendszer szerint vagy a szakképző intézmény kezdeményezésére időszakonként önértékelésre kerül sor, amely kiterjed a szakképző intézmények digitális felkészültségének és környezeti fenntarthatóságának az értékelésére is. Az intézményeknek mérniük, értékelniük kell a digitális felkészültségüket, ezekre a későbbiekben konkrét módszertani eszközök kerülnek bemutatásra.

A felülvizsgálat szakaszának egyik indikatív jellemzője: széles körben és nyilvánosan hozzáférhető a felülvizsgálat eredményeiről szóló információk (digitálisan/online csatornákon). Az értékelés eredményét digitalizált formában és online elérhetőséggel meg kell osztani a partnerek számára (tanulók, szülők, duális képzők, fenntartó), hogy pontos képet kapjanak az intézmény működéséről.

### 2.1. Eszközök az intézményi digitalizáció méréséhez

A szakképző intézmények rendelkezésére állnak különféle digitalizációs szintet mérő önértékelő eszközök, a Digitális Névjegrendszer (DNR) és a Self-reflection on Effective Learning by Fostering the Use of Innovative Educational Technologies (SELFIE).

#### 2.1.1. DNR

A DNR olyan komplex intézményi visszajelző- és fejlesztőeszköz, mely alkalmas az iskolák digitális fejlettségi szintjének meghatározására. A rendszer célja sokrétű, hiszen a visszajelzést és fejlesztési javaslatokat nyújtó funkciók révén strukturált formában nyújt információkat a köznevelési intézmény digitális érettségéről és a fejlődés lehetséges lépéseiről. Az EFOP-3.2.15-VEKOP-17-2017-00001 „A köznevelés keretrendszeréhez kapcsolódó mérési-értékelési és digitális fejlesztések, innovatív oktatásszervezési eljárások kialakítása, megújítása” projekt keretében a Digitális Jólét Nonprofit Kft. Digitális Pedagógiai Módszertani Központja alakította ki DNR-t. [1]

A DNR-ben kialakításra került egy Önértékelő modul, melynek célja az intézmény digitális érettségére vonatkozó pillanatkép elkészítése.



Az önértékelő eszköz öt tématerülete:

- Vezetés és menedzsment
- Digitális pedagógiai kultúra
- Szakmai fejlődés
- Iskolai digitális kultúra
- Infrastruktúra

E tématerületeken keresztül vizsgálja meg az eszköz, hogy hol tart a digitális átmenet. A kérdések megválaszolását a kitöltés során fogalommagyarázatok, illetve a Digitális Iskola Kézikönyv vonatkozó részei segítik. [1]

A kérdéssor kitöltése és az intézmény megfelelő tényszerű adatainak (Monitoring modul) feltöltése után az intézmény kap egy visszajelzést, hogy az egyes területeken milyen fejlettségi szintet azonosított be. A meghatározott szintek:

1. Belépő
2. Útkereső
3. Haladó
4. Szakértő
5. Mester

A Hogyan tovább? modul segítségével pedig a saját fejlettségi szintjéhez kaphat az intézmény fejlesztési lehetőségekhez támogatást.

### 2.1.2. SELFIE

A SELFIE online önértékelő eszköz elkészítésének tervét az Európai Bizottság 2018 januárjában terjesztette elő, a Digitális Oktatási Cselekvési Terv (Digital Education Action Plan – DEAP) egyik kezdeményezéseként. [10]

Modulok:

- Iskolai önértékelés
- Oktatói/pedagógus önértékelés
- Duális képzőhelyek (WBL) önértékelése

A SELFIE önértékelő eszköz által felmért területek:

- Vezetőség
- Együttműködés és kapcsolati hálózatok
- Infrastruktúra és eszközök
- Szakmai továbbképzés
- Pedagógia: Tantermi oktatás
- Értékelési gyakorlatok
- A diákok digitális kompetenciája

A felmérés teljes körű, hiszen az oktatók, vezetők, tanulók és duális partnerek is kitöltik. Az információk alapján az eszköz létrehoz egy jelentést, vagyis egy „pillanatfelvételt” („szelfit”), amely megmutatja az iskola erősségeit és gyengeségeit a digitális technológiák tanítása és tanuláshoz való használata területén. Minél több válaszadó (tanulók, oktatók, szülők, duális partnerek) vesz részt a kérdőív kitöltésében, annál pontosabb képet kap az intézmény a SELFIE segítségével. [10]



### 2.1.3. DigCompEdu

Pedagógusok Digitális Kompetenciájának Európai Keretrendszere (DigCompEdu) egy tudományosan megalapozott keretrendszer, mely pontosan meghatározza, mit értünk a pedagógusok digitális kompetenciáján és általános referenciakeretet nyújt a pedagógusok digitáliskompetencia-fejlesztésének támogatásához. A DigCompEdu 6 kompetenciaterületet (1.ábra), benne 22 kompetenciaelemet határoz meg. A keretrendszer arra összpontosít, hogy miként használható fel a digitális technológia a tanítás-tanulási folyamat hatékonyságának növelésében és megújításában. Az önértékelő eszköz lehetőséget nyújt az oktatóknak a digitális technológia tanítási célú használatához kapcsolódó erősségei és gyengeségei azonosítására. Az egyes kompetenciaterületekhez 6 különböző jártassági szint kapcsolódik:

- A1-Belépő,
- A2-Felfedező,
- B1-Beépítő,
- B2-Gyakorlott,
- C1-Irányító,
- C2-Újító.

Az önértékelő eszköz adaptációja az EFOP-3.2.15-VEKOP-17-2017-00001 projekt keretében valósult meg a Digitális Pedagógiai Módszertani Központ vezetésével. [9]



1. ábra DigCompEdu 6 kompetenciaterülete, forrás: <https://digcompedu.dpmk.hu/start.php>

### Összegzés

Összefoglalásul megállapítható tehát, hogy a szakképzés területén minden készen áll a minőségi fejlődésre. A minőségirányítási rendszer bevezetésével a szakképző intézmények még jobban tudják szabályozni folyamataikat és az eredményeik folyamatos mérésével, önértékeléssel a hibáikat javítani. A digitalizáció szintjét mérni képes mérőeszközök is rendelkezésre állnak, melynek használatáról szabadon dönthetnek. A feltételek tehát adottak egy intézményi szintű digitalizációs fejlesztésre.



## Szakirodalom

- [1] A Digitális Iskola Kézikönyve, Digitális Névjegy Rendszer, az EFOP-3.2.15-VEKOP-17-2017-00001 „A köznevelés keretrendszeréhez kapcsolódó mérési-értékelési és digitális fejlesztések, innovatív oktatásszervezési eljárások kialakítása, megújítása” című kiemelt projekt keretében jött létre. [dnr.dpmk.hu](http://dnr.dpmk.hu) <https://www.digitalisnevjegyrendszer.hu/page.php?pid=77#mi> utolsó hozzáférés: 2022.05.27.
- [2] A szakképzésről szóló 2019. évi LXXX. törvény.
- [3] A TANÁCS AJÁNLÁSA (2020. november 24.) a fenntartható versenyképességet, a társadalmi méltányosságot és a rezilienciát célzó szakképzésről (2020/C 417/01) ([https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32020H1202\(01\)&from=HU](https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32020H1202(01)&from=HU)) utolsó megtekintés: 2022.06.30.
- [4] A Tanács és a Bizottság 2015. évi közös jelentése az oktatás és a képzés terén folytatott európai együttműködés stratégiai keretrendszerének végrehajtásáról (Oktatás és képzés 2020), Az oktatás és a képzés terén folytatott európai együttműködés új prioritásai (2015/C 417/04) [https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:52015XG1215\(02\)&from=EN](https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:52015XG1215(02)&from=EN) utolsó megtekintés 2022.06.30.
- [5] AZ EURÓPAI PARLAMENT ÉS A TANÁCS AJÁNLÁSA (2009. június 18.) a szakoktatás és szakképzés európai minőségbiztosítási referenciakeretének létrehozásáról (EGT-vonatkozású szöveg) (2009/C 155/01) ([https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32009H0708\(01\)&from=EN](https://eur-lex.europa.eu/legal-content/HU/TXT/PDF/?uri=CELEX:32009H0708(01)&from=EN)) utolsó megtekintés: 2022. 06.30.
- [6] Industry 4.0 Policy Department Economic and Scientific Policy, Study for the ITRE Committee, 2016.
- [7] Közlemény a szakképzésért felelős európai miniszterek, az európai szociális partnerek és az Európai Bizottság 2010. december 7-én Bruges-ben a koppenhágai folyamat 2011–2020-as időszakra vonatkozó stratégiai megközelítésének és prioritásainak felülvizsgálatára tartott találkozójáról. [file:///C:/Users/kasza.marianna/Downloads/Bruges\\_Communique\\_HU-1.pdf](file:///C:/Users/kasza.marianna/Downloads/Bruges_Communique_HU-1.pdf) utolsó megtekintés:2022.06.30.
- [8] Önértékelési kézikönyv a szakképző intézmények számára, A szakképzésért felelős miniszter által 2022.02.08-án jóváhagyott szakmai dokumentum/módszertani útmutató. Hatályos 2022.02.09-től. [https://api.ikk.hu/storage/uploads/files/onertekelesi\\_kezikonyv\\_2022\\_02\\_09pdf-1645096289357.pdf](https://api.ikk.hu/storage/uploads/files/onertekelesi_kezikonyv_2022_02_09pdf-1645096289357.pdf) utolsó megtekintés: 2022.06.30.
- [9] Pedagógusok Digitális Kompetenciájának Európai Keretrendszere (DigCompEdu)
- [10] SELFIE (Self-reflection on Effective Learning by Fostering the use of Innovative Educational Technologies) European Education Area, Quality education and training for all
- [11] SZAKKÉPZÉS 4.0 – A SZAKKÉPZÉS ÉS FELNŐTTKÉPZÉS MEGÚJÍTÁSÁNAK KÖZÉPTÁVÚ SZAKMAPOLITIKAI STRATÉGIÁJA, A SZAKKÉPZÉSI RENDSZER VÁLASZTÁSÁNAK EGYEDIK IPARI FORRADALOM KIHÍVÁSAIRA, Innovációs és Technológiai Minisztérium, 2019.

## Nemzetközi kitekintés a felsőoktatási könyvtárak világára: a EUGLOH könyvtári workshopja

Nagy Gyula  
SZTE Klebelsberg Könyvtár  
[gyula.nagy@ek.szte.hu](mailto:gyula.nagy@ek.szte.hu)  
ORCID: [0000-0002-8391-2851](https://orcid.org/0000-0002-8391-2851)

The EUGLOH Library Workshop attempted to provide a comprehensive overview of the new tasks of university libraries. Many new challenges and issues have been taken on in recent years: some of these come from the field of Open Science or analogue-to-digital transformation, while others are related to the general research support (e.g. advice on publications) or more specific issues (e.g. research data management). Experience has shown that in a progressive university environment, new demands and expectations are constantly emerging for libraries: university rankings, scientometrics, innovative user training and many more.

**Keywords:** academic library, EUGLOH, cooperation, workshop, staff mobility

### A EUGLOH program

A EUGLOH (European University Alliance for Global Health)<sup>1</sup> program alapvető tematikáját a globális egészség tágran értelmezett témaköre jelenti. A EUGLOH egy innovatív megközelítésű európai felsőoktatási szövetség, melynek elsődleges célja a globális egészséggel kapcsolatos oktatásban és képzésben való részvétel különböző interdiszciplináris tevékenységek által.

A legfontosabb célkitűzések között szerepel az európai felsőoktatás, kutatás és innováció vonzerejének és versenyképességének növelése. Egy közös európai egészségügyi térség alapjainak megteremtése, illetve egy európai kampusz felépítése magas szintű integrációval, amelyet közös eljárások és struktúrák tartanak fenn. Az olyan európai értékek előmozdítása, mint a szolidaritás, az esélyegyenlőség, a befogadás, az emberi jogok tiszteletben tartása és a jóléthez való teljes hozzáférés, továbbá egy olyan világszínvonalú felsőoktatási szövetséggé való átalakulás, amely releváns válaszokat adhat a globális egészségügyi és jóléti kihívásokra való összpontosításon keresztül. A megfogalmazott célok elérésének egyik kiemelt eszköze az egyetemi hallgatók és dolgozók mobilitásának növelése, ezáltal az európai felsőoktatás minőségének javítása.



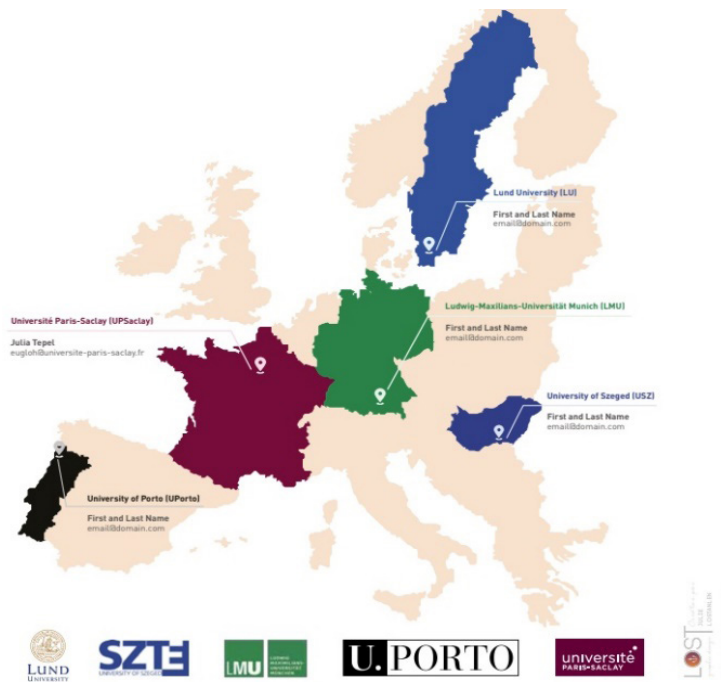
1. ábra A EUGLOH szövetség logója

---

1 European University Alliance for Global Health – <https://www.eugloh.eu>

A Szegedi Tudományegyetem 2019 nyarán csatlakozott a kezdeményezéshez, a 2019. június 28-i keltezésű sajtóközlemény<sup>2</sup> szerint: „A Szegedi Tudományegyetemet is beválasztotta az Európai Bizottság azon felsőoktatási intézmények közé, amelyek az első „európai egyetemi” szövetségek részét képezik. Ezen intézmények révén az európai felsőoktatás vonzóbbá válik, fokozódik az intézmények, a diákok és a dolgozók közötti együttműködés. Az SZTE az európai egyetemek élvonalába tartozó négy partnerintézményével a globális egészség kihívásaira keresi a választ.”

A EUGLOH szövetség tagjait vezető európai egyetemek alkotják: Université Paris-Saclay, Lund University, Ludwig-Maximilians-Universität München, University of Porto, University of Szeged. Az összefogás nemrégiben újabb tagokkal bővült: University of Alcalá, University of Hamburg, University of Novi Sad, Tromsø University.



2. ábra A EUGLOH szövetség tagjai

A szövetség mindennapi életének konkrét megvalósulása főként különböző szervezett rendezvényeken, látogatásokon és eseményeken keresztül történik. Ezek között vannak, amelyek az online térben zajlanak, illetve vannak személyes jelenlétet igénylők is, a pandémia idején előbbiek különösen hangsúlyos szerepet töltek be, ugyanakkor a résztvevő egyetemek közötti aktív kapcsolattartás és a földrajzi távolság is indokoltá teszi a virtuális rendezvényeket.

2 Az SZTE is részt vesz az európai felsőoktatás jövőjének alakításában - <https://u-szeged.hu/sztehitek/2019-junius/szte-is-reszt-vesz>

Show events Upcoming

Sort by Start date
Order Ascending

---

Filters:

Duration
Location
Type of event

Target group
Host
Subject

ECTS
Registration

3. ábra A EUGLOH események szűrését lehetővé tévő keresőprogram a szövetség hivatalos honlapján

A deklarált céloknak megfelelően, különösen fontos szerep jut a különböző mobilitási programoknak, melyeket a résztvevő intézmények individuálisan, vagy éppen egymással kooperálva szerveznek időről időre. A mobilitási programok között szintén előfordulnak online és személyes jelenléttel megvalósult alkalmak, sőt ezek kombinációja is gyakori.

< > today March 2022 month list year

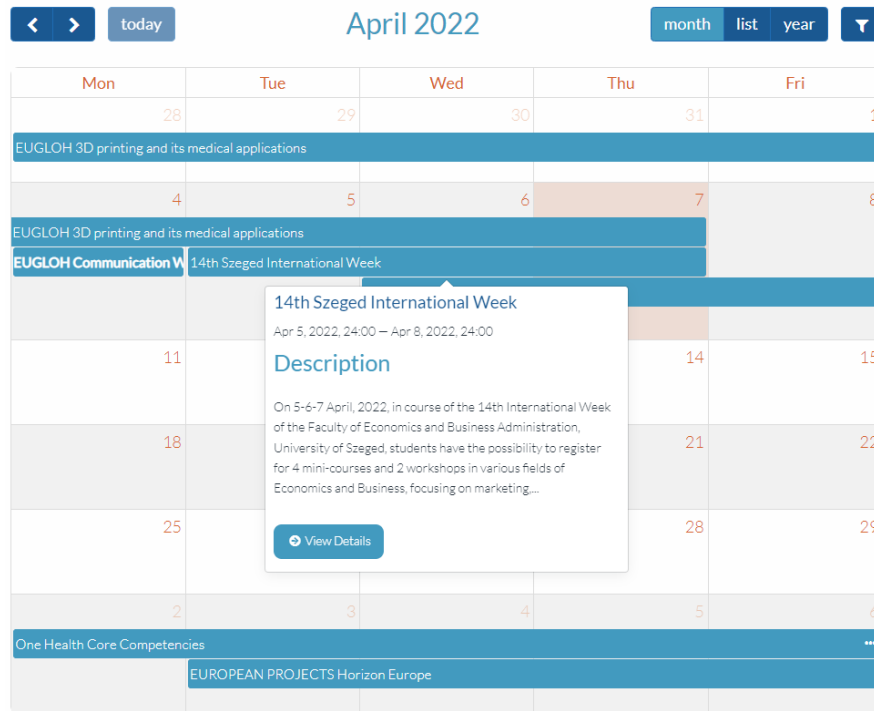
March 7, 2022		
all-day		Innovation Eco-systems and Entrepreneurship
5:00pm - 6:00pm		EUGLOH Business Talk with Infor
March 8, 2022		
all-day		Innovation Eco-systems and Entrepreneurship
1:00pm - 2:30pm		EUGLOH E-Conference series: Women
March 9, 2022		
all-day		Innovation Eco-systems and Entrepreneurship
March 10, 2022 <span>Thursday</span>		
all-day		Innovation Eco-systems and Entrepreneurship

Categories

- Block course
- Conference
- Discussion
- EUGLOHRIA
- Lecture
- Lecture Series
- MOOC
- Meeting
- Networking event
- Outreach event
- Semester course
- Seminar
- Summer/winter school
- Training
- Webinar
- Workshop
- Other

4. ábra EUGLOH programnaptár 2022 március elejéről

Fontos kiemelni, hogy a mobilitások során az együttműködés felelősei igyekeznek bevonni az egyetemi polgárok összes szintjét, hogy minél szélesebb bázisú munkatársi közösség tudjon bekapcsolódni a program sikeres megvalósításába. Így a hallgatók, oktatók, kutatók és az egyes funkcionális egységek munkatársai egyaránt a célcsoportba tartoznak, a potenciális résztvevők körét gyarapítva.



5. ábra Események 2022 áprilisában

Egy ilyenfajta mobilitás, az online és személyes jelenlétet vegyesen megvalósító esemény zajlott február közepén is, melynek könyvtárunk kollégái is részesei, sőt szervezői lehettek. Az ilyen módon „Blended Staff Mobility Opportunity for EUGLOH Partners” névre hallgató esemény online lába 2022. február 15-16. között került megrendezésre a Zoom platformon keresztül.

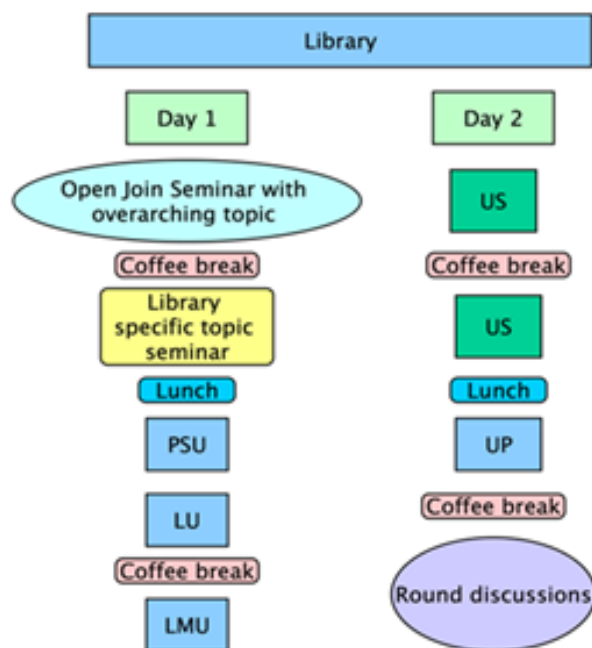
Egy közös plenáris előadás (The Future of Higher Education in Europe) után a konferencia résztvevői öt párhuzamos szekcióban hallgathattak előadásokat és vehettek részt különféle műhelybeszélgetésekben. Az esemény öt párhuzamos tématerülete a következő volt: karrier szolgáltatások (career services), kommunikáció és nyilvános szerepvállalás (communication & public engagement), oktatásfejlesztés (educational development), nemzetközi kapcsolatok és hallgatói mobilitás (international relations & student mobility) és végül saját területünk, a könyvtár (library).

### EUGLOH Library Workshop

A könyvtári terület kétnapos programjának a Szegedi Tudományegyetem volt a szervezője, a könyvtár munkatársai határozták meg a tematikát és a program szerkezetét is. A EUGLOH tagegyetemek 5-5 fővel képviseltették magukat a könyvtáros szekcióban, így az online konferencia összesen 25 fő részvételével zajlott.

A könyvtári workshop előkészítése 2021. október-novemberben indult, a szervezőmunka főként online meetingek segítségével, illetve e-mailen keresztül zajlott.





6. ábra A 2022. február 15-16-i EUGLOH Library Workshop felépítése

Célkitűzése szerint a EUGLOH Könyvtári Workshop megkísérelt átfogó áttekintést nyújtani az egyetemi könyvtárak új feladatairól, hiszen az elmúlt években folyamatosan számos új kihívás és kérdés merült fel Európa-szerte az akadémiai szféra könyvtári területein. Ezek egy része a nyílt tudomány vagy az analóg-digitális átalakulás területéről származik, míg mások az általános kutatástámogatáshoz (pl. publikációs tanácsadás) vagy speciálisabb kérdésekhez (pl. kutatási adatok kezelése) kapcsolódnak.

A tapasztalatok azt mutatják, hogy a progresszív egyetemi környezetben folyamatosan új igények és elvárások jelennek meg a könyvtárakkal szemben, például releváns válaszok adásának képessége az egyetemi rangsorok, a tudománymetria, az innovatív felhasználói képzések és még sok egyéb terület vonatkozásában. Nem véletlen tehát, hogy az egyes előadások is ezen témacsoportok köré szerveződtek.

Maga a kétnapos esemény egy közös webináriummal indult a Lund University jóvoltából, ahol az európai felsőoktatás jövője volt a téma. Itt három résztvevő beszélgetett egyfajta kerekasztal formátumban egy moderátor közreműködésével. Ez a plenáris webinárium teljesen nyitott volt minden EUGLOH tag számára, körülbelül 120 résztvevő hallgatta meg a beszélgetést. Ezt követték az öt tématerület önálló workshopjai, így a könyvtári workshop is.

Ennek első tematikus szekciója az egyetemi rangsorok és a felsőoktatási könyvtárak kapcsolatával foglalkozott (Linking points between university rankings and libraries). Az eseményt egyfajta eszmecsereként szántuk, ahol minden résztvevő egyetem 15 percet kapott, hogy megossza tapasztalatait és ismertesse saját intézménye álláspontját az egyetemi rangsorokkal kapcsolatban, majd ezután nyílt vita következett az egyes beszámolókból elhangzott pontokkal kapcsolatban.



A beszélgetés az alábbi irányított kérdések mentén szerveződött: Lényegesek-e az egyetemi rangsorok egy európai egyetem életében? Melyek a legtekintélyesebb egyetemi ranglisták, és mely rangsorokat részesítsük előnyben? Van-e könyvtáraknak feladata az egyetemi rangsorokkal kapcsolatban? Hogyan támogathatja egy egyetemi könyvtár az egyetemi rangsorokkal kapcsolatos projekteket?

Az interaktív műhelybeszélgetést előadások sora követte: minden résztvevő számára nagyjából egy háromnegyed órás időtartam állt rendelkezésére, amelynek során prezentációkon keresztül mutatták be saját könyvtáruk kezdeményezéseit, elképzeléseit az előzetesen egyeztetett témák mentén, ezután pedig a résztvevők rendre kérdéseket tehettek fel az elhangzott előadásokkal kapcsolatban.

Az első tematikus előadást az Université Paris-Saclay tartotta, Felhasználáson alapuló új szolgáltatások (New services based on uses) címmel. Az előadók a könyvtárak olvasói igényeknek megfelelő infrastruktúra-fejlesztéséről számoltak be. A párizsi egyetemnek több mint harminc könyvtára van, ezek között olyan is van, amelyik kifejezetten a látássérült olvasók igényeit szolgálja ki.

Az épületek felújítását megelőzően felméréseket végeztek az egyetemi polgárok körében, hogy pontosan sikerüljön felderíteni, hogy használóik mire tartanak igényt. Ehhez a virtuális valóság technológia által biztosított lehetőségeket is felhasználták, amelynek segítségével az olvasók bejárhatták a tervezett épületeket és elmondhatták, hogy mit változtatnának meg a terek elrendezésében. Ez különösen hasznosnak bizonyult a mozgáskorlátozott olvasók esetében, mert ezzel a módszerrel meg tudtak győződni róla, hogy a kerekesszékekben ülők számára is megfelelő magasságban helyezkednek el a jelzetek és minden szolgáltatás elérhető. Emellett kamerák segítségével figyelték meg az olvasók mozgásmintázatát a belső terekben, továbbá az asztalokba, valamint a székekbe épített szenzorokkal állapították meg, hogy melyek azok a helyek, ahová szívesen ülnek le, majd megvizsgálták, hogy mi az oka, hogy ezeket a pontokat előnyben részesítik a hallgatók. A projekt keretében különböző típusú bútorokat próbálhattak ki az olvasók, hogy kiválaszthassák a legmegfelelőbbeket a könyvtár számára.

A látássérültek számára fenntartott könyvtár kialakításánál a fizioterápia szakos hallgatókkal közreműködve tervezték meg a teljes épületet. Itt kisebb tereket alakítottak ki, hogy csak kevesebb ember tartózkodjon egy helyen, mert ebben az esetben az olvasók különösen érzékenyek a zajokra, míg a számítógépeken felolvasó program érhető el fejhallgatóval, speciális billentyűzet, nagyító monitor, szkener állnak rendelkezésre, illetve az asztalokat is kifejezetten az ablakok mellé igyekeztek pozícionálni. Az előadók beszámoltak arról, hogy jelenleg is éppen egy új oktatási központot építenek, melyet ősszel terveznek megnyitni.

A következő előadást a Lund Egyetem tartotta, ők a könyvtári hálózatban megtalálható intézmények közötti kollaborációs munkáról (Working in parallel: collaborative work across institutions), valamint egy nagyszabású digitalizálási projektről (Digitisation of all Swedish Print) számoltak be. A svéd egyetem könyvtári hálózata decentralizált modellben működik, mintegy 22 könyvtár tartozik a lund rendszerbe, ezek elszórva találhatóak a városban, illetve a Malmöben és a Helsingborgban lévő kampuszokon.

2021-ben egy kutatástámogatást fejlesztő projektbe kezdett hét kiemelt könyvtár, amelynek célja egy új modell kifejlesztése és tesztelése volt az egyes tudományágak, illetve az azokat reprezentáló intézmények közötti együttműködésben. A projektet a Svéd Felsőoktatási Intézmények Szövetségének kutatókönyvtári szakértői hálózata kezdeményezte. Az előadás második részében a központi könyvtár igazgatója arról számolt be, hogy 2020 januárjában Svédország öt legnagyobb egyetemi könyvtára a Svéd Nemzeti Könyvtárral közös projektjében a 15. század óta megjelent összes svéd nyomtatvány digitalizálását tűzte ki célul, természetesen a nagyközönség számára is elérhető formában.

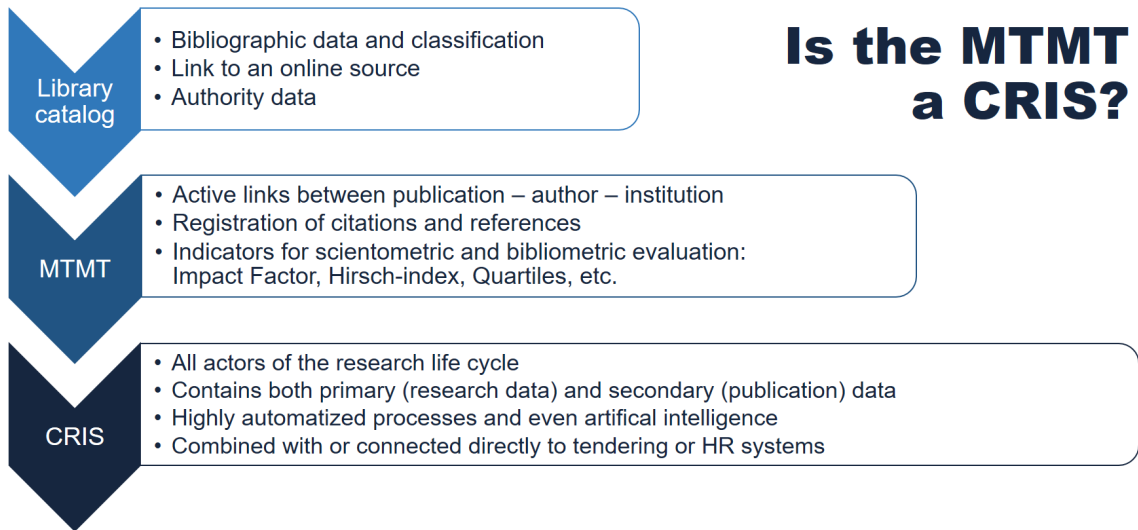
A harmadik előadást a Ludwig-Maximilians Egyetem munkatársai tartották Münchenből, ők a nyílt tudománnyal kapcsolatos szolgáltatásaikról (Openness in Academic Libraries – Examples from the University Library LMU) beszéltek. A német könyvtárosok komoly elköteleződést mutatva támogatják a nyílt tudományt, mottójuk, hogy „az információ követeli a szabadságot”. Hat pont köré csoportosítják az Open Science vonatkozású szolgáltatásaikat: Open Access, Open Data, Open Educational Resources, Open Innovation, Open Peer Review, Open Source.

Az egyetemi könyvtár több EPrints alapú repositóriumot is üzemeltet, illetve 2018-ban egy Open Science Center elnevezésű szervezeti egységet is alapítottak a terület történéseinek hatékonyabb összefogása érdekében. Mintegy 13 Open Access folyóiratuk van, emellett megtudhattuk azt is, hogy az egyetemen nincs központi publikációs finanszírozó szerv, az egyetemi karoknak saját maguknak kell gondoskodniuk a publikációs díjak költségeinek fedezéséről. Az LMU könyvtára a FOLIO elnevezésű, nyílt felhőalapú könyvtári integrált rendszer fejlesztésében is részt vesz, ezért ezzel kapcsolatban is hallhattunk egy rövid beszámolót.

A második napon házigazda-szervezőként a Szegedi Tudományegyetem két előadást tarthatott: első előadásunk a Magyar Tudományos Művek Tárát mutatta be (MTMT: the Hungarian CRIS), majd pedig különböző szintű felhasználóképzési módszereinkről (User training opportunities at all levels: from Bachelor to researcher) számoltunk be. Az előadás segítségével egy általános áttekintést szerettünk volna nyújtani könyvtárunk felhasználóképzési portfóliójáról. Különösen a Kölcsönözz ki egy könyvtárost!<sup>3</sup> szolgáltatás keltette fel a kollégák érdeklődését, ezzel kapcsolatban számos kérdés merült fel. Ahogyan a kutatás- és publikációnyilvántartással kapcsolatos előadás során is több észrevétel hangzott el.

---

3 Kiss Márta, Szűcs Judit. Kölcsönözz ki egy könyvtárost! Új szolgáltatás szakdolgozóknak a Klebelsberg Könyvtárban. In: Tick, József; Kokas, Károly; Holl, András (szerk.) *NETWORKSHOP 2018 konferenciakötet* Budapest, Magyarország : Hungarnet (2018) p. 64-70. <https://doi.org/10.31915%2FNWS.2018.9>



## Is the MTMT a CRIS?

7. ábra Előadás az MTMT szerepéről a magyar kutatási ökoszisztémában

Not only books, but „**Borrow a librarian**” as well - [szakdolgozat.ek.szte.hu](http://szakdolgozat.ek.szte.hu)

**Theses topics** we have been approached with

- Rice flour pasta production technology
- Kindergarten motivational techniques in the education and upbringing of the digital generation
- Piercing for adolescents
- Galvanic wastewater treatment
- From waste to artwork
- Development of hypotonic drinks
- Plato's influence on Tolkien



8. ábra Előadás a felhasználóképzés különböző szintjeiről az SZTE Klebelsberg Könyvtárban: a Kölcsonözz ki egy könyvtárost szolgáltatás bemutatása

Végül az utolsó önálló előadásban a Portói Egyetem kollégái a digitális könyvtárakban rejlő lehetőségekről (Digital libraries: much more than an interplay with technology) beszéltek. Az egyik előadó már negyven éve a könyvtár dolgozója, így előadásában visszatekintett a könyvtár történetére és a lezajlott számos változásra.

Az egyetemek saját tapasztalatait bemutató előadások után egy újabb másfél órás témamegbeszélés következett. Ez a záró panel egy moderált, ugyanakkor nyílt kerekasztal-beszélgetés volt. Ennek során olyan kérdések kerültek tárgyalásra, mint például az egyes intézmények strukturális felépítéséből adódó különbségek. A magyar példán kívül ugyanis az összes többi résztvevő könyvtár decentralizált modellben működik, ezért az ezzel kapcsolatos nehézségekről és előnyökről mindenki kifejtette álláspontját. Emellett szó esett a hagyományos analóg gyűjtemények szerepéről, illetve jövőjéről napjaink digitális világában, az egyetemi könyvtárak felhasználóképzéseinek kitüntetett fontosságáról, illetve az egyes kutatástámogató szolgáltatások bevezetése során tapasztalható tanulságokról is.

Összességében elmondható a kétnapos, szegedi szervezésű nemzetközi partnerek részvételével zajló EUGLOH könyvtári workshopról, hogy egy kifejezetten kellemes hangulatú, inspiráló eseményt sikerült megvalósítani, amely remek perspektívát és gondolatébresztőt jelentett arra vonatkozóan, hogy milyen fejlődési irányok látszanak kibontakozni más nagy európai felsőoktatási intézmények könyvtáraiban.

## Az európai természettudományi gyűjtemények digitális integrációja: határ a csillagos ég

Babocsay Gergely  
Magyar Természettudományi Múzeum Mátra Múzeuma  
[babocsay.gergely@nhmus.hu](mailto:babocsay.gergely@nhmus.hu)

### Abstract

Natural history collections harbour billions of biological and geological specimens worldwide. The collections of CETAF (Consortium of European Taxonomic Facilities) have amassed ca. 1.5 billion specimens, representing about 80% of the known bio- and geodiversity of our planet. This is a unique material that can help to solve endless scientific problems related to the environment and biodiversity. Accessing every bit these collections physically has become almost impossible and costly. Collections, however, assemble into digitally accessible research infrastructures (RIs) worldwide making digitised specimens available online. One of these is the CETAF initiated Distributed System of Scientific Collections (DiSSCo), an RI under the ESFRI framework. DiSSCo is in a preparatory phase, but will be fully operational in 2026. The Hungarian Natural History Museum is part of the process of this realisation, but it can become a full member of DiSSCo only if it is included in the national RI roadmap.

### Bevezetés

A természettudományi múzeumok speciális környezettudományi kutatási infrastruktúrák, amellett, hogy hagyományosan az élővilággal és a földtörténettel kapcsolatos közművelésben úttörő és a mai napig élenjáró intézményekről van szó. A világ természettudományi múzeumai 2,5-3 milliárd biológiai és kőzettani példányt őriznek<sup>1,2</sup>, csak az Európai Természettudományi Gyűjtemények Konzorciumába (CETAF, Consortium of European Taxonomic Facilities) tartozó 22 ország több mint 70 intézménye 1-1,5 milliárdot<sup>3</sup>. Ez utóbbiak az ismert fajok 80%-át őrzik a gyűjteményeikben. A hazai gyűjtemények közül a Magyar Természettudományi Múzeum és tagintézményei a legnagyobbak, az általuk őrzött példányok száma meghaladja a 10 milliót. A természettudományi gyűjtemények minden egyes példánya információt hordoz önmagáról, fajáról, arról az időpillanatról és arról a környezetről, amelyből és ahonnan begyűjtötték őket. Általuk nem csak az élővilág mai állapotáról és környezetéről kaphatunk információkat, hanem a múlt evolúciós és geológiai eseményei és környezeti változásai is nyomon követhetők. Ebben az értelemben tehát unikális és megismételhetetlen tudományos anyagról van szó, amely más formában nem áll rendelkezésre az emberiség és a tudomány számára. A gyűjtemények mérete azonban olyan hatalmasra duzzadt, hogy hozzáférhetőségük biztosítása gyakorlatilag lehetetlenné vált (csak a Smithsonian rovargyűjteménye 1976 és '86 között éves átlagban 100 ezer példányt kölcsönzött ki<sup>4</sup>, és az általuk hordozott információk kiaknázása, és így általában a társadalmi hasznosulásuk messze elmarad a bennük rejlő potenciáltól. Éppen ezért a természettudományi gyűjtemények anyagainak digitalizálása és a digitális hozzáférés biztosítása mára elkerülhetetlenné vált, különösen annak fényében, hogy a biológiai sokféleség hanyatlása soha nem látott ütemben zajlik, miközben az emberiség egyre nagyobb, az élővilág felől érkező kihívásokkal (özönfajok, mezőgazdasági, erdészeti kártevők, újonnan megjelenő fertőző betegségek stb.) kénytelen szembe nézni. A digitalizáció révén a gyűjtemények tudományos célú hozzáférhetősége csillagászati



léptékben fog növekedni, és ezzel egy időben az oktatás számára is eddig el nem érhető utak nyílnak meg<sup>5</sup>.

## Túl a 21. századon; digitalizált természettudományi gyűjtemények

A közgyűjtemények anyagainak digitalizációjával kapcsolatos igények több irányból is érkezőek. Különösen az európai szinten, társadalmi hasznosulásuk igényét a fenntartói szándékok fejezik ki legmarkánsabban, amelyek előtérbe helyezik a kultúrához való hozzáférés minél szélesebb körű megvalósulását, szem előtt tartva az esélyegyenlőséget, a közösségépítést, a részvételiséget, a magyar kulturális identitás erősítésének vagy a tudásalapú társadalom szempontjait. A hazai közgyűjtemények digitalizációjának szándéka megjelent a Digitális Nemzet Fejlesztési Programban (1486/2015. (VII. 21.) kormányhatározat), részben mint uniós cél, részben mint össztársadalmi igény, annak érdekében, hogy a gyűjteményekben őrzött tárgyak minél szélesebb körű hozzáférése biztosított legyen.

A természettudományi gyűjtemények már több mint 150 éve a biológiai sokféleség kutatásának legfontosabb infrastruktúrái, anyagaik elsősorban az evolúcióbiológia, a biogeográfia, az ökológia, a taxonómia vagy a földtan területén szolgáltattak pótolhatatlan, mással nem kiváltható kutatási anyagot<sup>6,7</sup>. A gyűjteményi anyag azonban csak korlátozottan hozzáférhető a hatalmas példányszám, a korlátozott fizikai tér, a véges számú személyzet, a logisztikai nehézségek, az állományvédelmi szempontok, vagy pénzügyi korlátok miatt. A digitalizálás révén a gyűjteményi adatok hasznosulásának növekedése a társadalom minden szegmensében gyorsulni fog, ahogyan azt pl. a Finn Biodiverzitás-információs Eszköz (FinBIF) életbe lépése is demonstrálta, amelyről 175.000 új felhasználója 100\*10<sup>6</sup> mennyiségű adatpontot töltött le 2017 és 2019 között. A letöltők e-mail címéből kiderült, hogy többségük civil felhasználó volt<sup>8</sup>. A Londoni Természettudományi Múzeum Adatportáljáról 2015-ös élesítése óta a felhasználók 22 milliárd (!) adatot töltöttek le<sup>9</sup>. Ezek a számok nyilvánvalóan jelzik, hogy a biológiai és geológiai sokféleséggel kapcsolatos adatok nagy érdeklődésre tartanak számot, és hogy gyűjtemények adataihoz biztosított széleskörű digitális hozzáférés lehetősége eddig elképzelhetetlen távlatokat nyit meg azok hasznosulása előtt, akár a tudományos kutatás, akár a részvételi tudomány és az oktatás területén.

A mit, miért, kinek és hogyan digitalizáljuk kérdés ma még nem minden aspektusában válaszolható meg könnyen, de a természettudományi gyűjtemények esetén konszenzus övezi azt az álláspontot, hogy a teljes állomány valamilyen szintű digitális leképezésére van szükség. A tudományos célú digitalizálás mélységei és rétegei függenek a gyűjtemény jellegétől (1a,b ábra), a rendelkezésre álló technikáktól és technológiáktól, valamint a tudományos ismeretek és az elemzési módszerek fejlődésétől. Az intakt, alkoholos gerinces példányok mikro-CT-s digitális leképezése<sup>10</sup> által a példányok fizikai roncsolása nélkül juthatunk értékes csonttani bélyegekhez<sup>11</sup>. Speciális igények közé tartozhat a lepkeszárnyak nanostruktúrájának leképezése és spektrális elemzése<sup>12</sup>. Az ezekből származó digitális objektumok felhasználása a tudományos elemzésen túl például a biodizájn számára lehetnek kreatív bemenetek<sup>13</sup>. A nagy kihívást jelentő rovardobozok tartalmának digitalizálására is egyre kifinomultabb fototechnikák és rendszerek állnak rendelkezésre<sup>14</sup>. A különféle természettudományi gyűjteménytípusok digitalizálásának aspektusait, kihívásait, lehetséges irányait és a jó gyakorlatokat az Európai Unió által finanszírozott ICEDIG elnevezésű projekt eredményei foglalják össze<sup>15</sup>. Rendkívül fontos szempont, hogy a digitális adatok FAIR-ek, azaz megtalálhatók, elérhetők, együttműködők és újrahasználhatók (Findable, Accessible, Interoperable and Reusable) legyenek<sup>16</sup>. Ez utóbbi teremti meg a gyűjteményekben rejlő információk maximális társadalmi hasznosulását.



## A digitalizált európai természettudományi gyűjtemények, mint kutatási infrastruktúra

A világon több kezdeményezés is indult a természettudományi gyűjtemények anyagainak digitális mobilizálása érdekében. Az Egyesült Államokban az Integrated Digitized Biocollections (iDigBio), Ausztráliában a CSIRO National Research Collections Australia (NRCA) tekinthetők előrehaladott projektnek ezen a területen. A CETAF kezdeményezésére az európai természettudományi gyűjtemények is összefogtak, hogy egy nagy digitálisan elérhető kutatási infrastruktúrát (KI) alkossanak. A cél eléréshez az Európai Kutatási Infrastruktúrák Stratégiai Fóruma (European Strategy Forum on Research Infrastructures, ESFRI) keretein belül fogalmazták meg a céljaikat, és Distributed System of Scientific Collections (DiSSCo) néven megkezdték a leendő KI felépítését, amely 2018-ban felkerült az európai KI-útitervebe. A DiSSCo jelenleg, mint ESFRI-projekt, előkészítési fázisában van, és a tervek szerint 2026-ban válik teljes mértékben működő KI-vá (DiSSCo.eu). A DiSSCo-ba eddig 21 országból 130-nál is több intézmény (köztük a Magyar Természettudományi Múzeum) jelentkezett be. A fentebb említett ICEDIG projekt célja volt, hogy az előkészítő fázisban felvázolja azokat a fő kereteket, amely mentén a DiSSCo digitális architektúrája és működési mechanizmusai felépíthetők és kialakíthatók. Bár a projekt célja kutatási célú infrastruktúra hadrendbe állítása, a DiSSCo alapvetően a fizikai gyűjteményi állomány digitális hozzáférhetőségét fogja megteremteni, amely hatalmas potenciált jelent a kutatás, a közművelődés és az oktatás számára. Egyben nagyon jelentős állományvédelmi szempontok is megvalósulnak<sup>17</sup> általa, hogy a fizikai hozzáférés igényei csökkennek. A digitalizált gyűjtemények alapját a digitális vagy kiterjesztett példány (digital/extended specimen) fogja képezni (2. ábra), amely jóval több, mint maga a fizikai példány, amely továbbra is a gyűjtemények része marad, és soha nem kerül megsemmisítésre. A kiterjesztett példány egységet képez, amely a fizikai példány digitális leképezéséből (legyen az fénykép, röntgen- vagy ultrahangos felvétel stb.), a hozzá tartozó gyűjteményi, gyűjtési/megfigyelési adatokból valamint az egyéb hozzá kapcsolható adatokból (DNS-szekvencia, izotópadatok, irodalmi hivatkozások fellelhetősége stb.) áll. A kiterjesztett példány információtartalma annak felhasználásával egyre nő, mivel újabb és újabb kontextusokban rakódnak rá/extrahálódnak belőle további adatok. A kiterjesztett példányt stabil, állandó egyedi azonosító<sup>18</sup> azonosítja, amely egyetlen online portálon keresztül utat nyit a hozzá tartozó összes fellelhető információhoz. Ezzel párhuzamosan törekvések indultak a biológiai sokféleséget leíró taxonómiai publikációk gépi olvasást lehetővé tevő, szemantikus annotációt tartalmazó szerkesztésére<sup>19</sup> és a már korábban megjelent publikációkból történő adatbányászatra is, ami a publikált adatokat automatikusan lesz képes összekötni az érintett példányokkal az adataggregátorok számára.

### Összefoglalás

A DiSSCo hadrendbe állásával az európai természettudományi gyűjtemények „one stop shop” rendszerben egyetlen platformon keresztül lesznek elérhetők a legkülönbözőbb felhasználók számára. Általa felgyorsulhat a biológiai sokféleség feltárása, az új fajok leírásának üteme, miközben a fizikai hozzáférés igénye csökken, ami jelentős utazási illetve kölcsönzési (posta-) költségmegtakarítást jelent. Kiszélesedik a használók köre, különösen a közösségi tudomány művelői és az oktatás számára jelent majd könnyebbséget és több lehetőséget. Megnyílik a terep más tudományterületek, az ipari fejlesztés, az építészet vagy a művészetek számára is. A bioinspiráció lehetősége egy kattintással elérhető lesz mindazok számára, akik inspirálódni szeretnének a gyűjteményekben felhalmozott élőlények vagy kőzetek által. A lehetőségek tárháza tehát kimeríthetetlen, de a gyűjtemény-digitalizáció a hazai természettudományi gyűjteményekben egyelőre nehézkesen indul, bár történnek előrelépések. A folyamat felgyorsításához hatalmas lehetőségnek tekinthető a Magyar Természettudományi Múzeum és tagintézményeinek DiSSCo-tagsága, de a teljes jogú tagság eléréséhez az MTM-nek fel

kellene kerülni a hazai ESFRI Kutatási Infrastruktúra útitervbe. Amennyiben ez megtörténik, a gyűjteményeinek hasznosulása előtt végtelen lehetőségek tárulnak fel.

## Bibliográfia

- 1 Duckworth, W.D., Hugh H., Genoway, H.H. & Rose, C.L. 1993. Preserving natural science collections: chronicle of our environmental heritage. National Institute for Conservation of Cultural Property; Washington DC, USA.
- 2 OECD 1999. Megascience Forum Working Group on biological informatics, Final Report. Hozzáférés: 2022. 06. 14. <https://www.oecd.org/science/inno/2105199.pdf>
- 3 CETAF. Hozzáférés: 2022. 06. 14. <https://cetaf.org/>
- 4 Suarez A.V. & Tsutsui N. D. 2004. The Value of Museum Collections for Research and Society. *BioScience*, 54: 66-74. [https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
- 5 Bakker, H., Willemsse, L., van Egmond, E., Casino, A., Gödderz, K., Vermeersch, X. 2018. Inventory of current criteria for prioritization of digitization. ICEDIG Deliverable D2.1. <https://doi.org/10.5281/zenodo.2579156>
- 6 Babocsay G. 2015. A kutató fogott egy gyönyörű madarat, azután megölte... MTM Blog: Mire jók a gyűjtemények? Hozzáférés: 2022. 06. 14. [https://mttmuzeum.blog.hu/2015/10/30/a\\_kutato\\_fogott\\_egy\\_gyonyoru\\_madarat\\_azutan\\_megolte](https://mttmuzeum.blog.hu/2015/10/30/a_kutato_fogott_egy_gyonyoru_madarat_azutan_megolte)
- 7 Miller, S.E., Barrow, L.N., Ehlman, S.M., Goodheart, J.A., Greiman, S. E., Lutz, H.L., Misiewicz, T.M., Smith, S.M., Tan, M., Thawley, C.J., Cook, J.A. & Light, J.E. 2020. Building natural history collections for the twenty-first century and beyond. *BioScience*, 70: 674–687. <https://doi.org/10.1093/biosci/biaa069>
- 8 Schulman, L.; Lahti, K.; Piirainen, E.; Heikkinen, M.; Raitio, O.; Juslén A. 2021. The Finnish Biodiversity Information Facility as a best-practice model for biodiversity data infrastructures. *Scientific Data*, 8: 137.; <https://doi.org/10.1038/s41597-021-00919-6>.
- 9 Natural History Museum, London 2020. Hozzáférés: 2022. 06. 14. <https://www.nhm.ac.uk/press-office/press-releases/scientists-rack-up-record-downloads-from-the-nhm-data-portal-dur.html>
- 10 Weinell, J.L., Paluh, D.J., Siler, C.D. & Brown, R.M. 2020. A new, miniaturized genus and species of snake (Cyclocoridae) from the Philippines. *Copeia*, 108: 907–923. <https://doi.org/10.1643/CH2020110>
- 11 Keklikoglou, K., Faulwetter, S., Chatzinikolaou, E., Wils, P., Brecko, J., Kvaček, J., Metscher, B. & Arvanitidis, C. 2019. Micro-computed tomography for natural history specimens: a handbook of best practice protocols. *European Journal of Taxonomy* 522: 1-55. <https://doi.org/10.5852/ejt.2019.522>
- 12 Kertész K., Bajja Zs. Deák A., Piszter G., Rázga Zs., Bálint Zs. & Biró L. P. 2021. Additive and subtractive modification of butterfly wing structural colors. *Colloid and Interface Science Communications*, 40: 100346. <https://doi.org/10.1016/j.colcom.2020.100346>
- 13 Aish, A. & Sun, J-S., 2020. Bioinspire-Museum: Scoping Paper, Muséum national d'Histoire naturelle, Paris, 24 pp.



- 14 Short, A.E.Z., Dikow, T. & Moreau, C.S. 2018. Entomological collections in the age of big data. *Annual Review of Entomology*, 63: 513–530.  
<https://doi.org/10.1146/annurev-ento-031616-035536>
- 15 ICEDIG: <https://icedig.eu/content/deliverables> Hozzáférés: 2022. 06. 14.
- 16 Wilkinson M., Dumontier, D. & Mons, B. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3: 160018.  
<https://doi.org/10.1038/sdata.2016.18>
- 17 Bánki Zs. & Ruttkay Zs. 2020. Digitalizált gyűjtemények oktatási hasznosítása. In: Pacsika M. (szerk.) *Digitális múzeumi tartalmak a köznevelés szolgálatában*. Múzeumi iránytű, 24: 75-96.
- 18 CETAF Information and Science Technology Commission. Hozzáférés: 2022. 06. 14.  
<https://cetafidentifiers.biowikifarm.net/wiki/>
- 19 Chester, C., Agosti, D., Sautter, G., Catapano, T., Martens, K., Gérard, I., & Bénichou, L. 2019. EJT editorial standard for the semantic enhancement of specimen data in taxonomy literature. *European Journal of Taxonomy*, (586).  
<https://doi.org/10.5852/ejt.2019.586>

## Ábrák



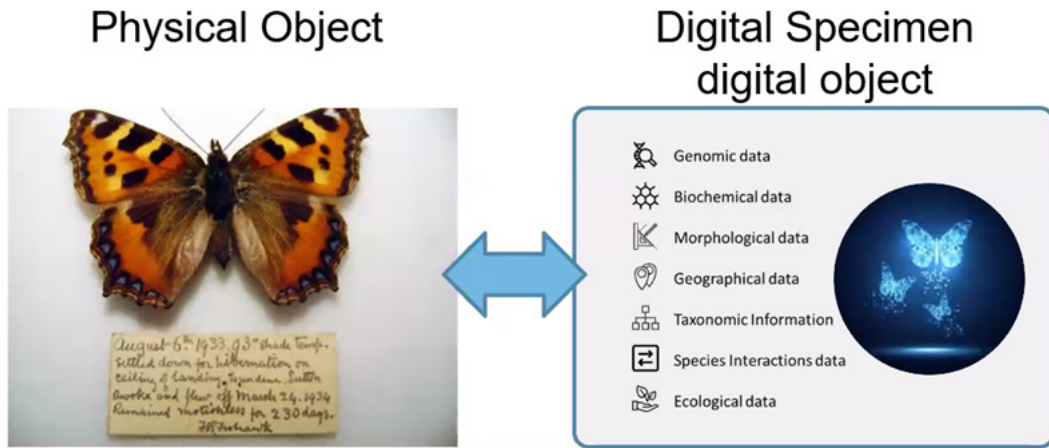
1 ábra. a.) A Németh Márton féle tojásgyűjtemény a Mátra Múzeumban.

Digitálizálása esetén a több, mint 32 ezer darabot számláló gyűjtemény kiváló anyag lenne az evolúcióbiológia oktatására vagy akár kisebb részvételi tudományos kutatási projektek számára.

Fotó: Babocsay Gergely.

b.) Szalamandrapéldányok a Mátra Múzeum Gerincesgyűjteményében. Az alkoholos gyűjtemények digitalizálása például speciális nehézségeket támaszt, mivel több példány található egy üvegben.

Fotó: Magyar Balázs.



2. ábra. A kiterjesztett példány, amely a fizikai példány digitális leképezése, és a hozzá tartozó gyűjteményi, gyűjtési/ megfigyelési, és az egyéb hozzá kapcsolható adatok (Forrás: [dissco.eu](http://dissco.eu)).



Egyenlőtlenségek a tudományos kutatás területén.  
Az amatőr kutatók szerepe

Somorjai Noémi

[somnoemi@gmail.com](mailto:somnoemi@gmail.com)

ORCID: [0000-0002-9379-2339](https://orcid.org/0000-0002-9379-2339)

Discrepancies in scientific research. The role of civil scientists.

Due to the rise of medical research, recent decades have shown an explosion of medical and health-related information. In addition to salient research topics, there are areas in information provision that are markedly lacking. Is it possible to observe these processes ethically, as an outsider, from the vantage point of the information science and librarian profession? There is a wealth of internationally available knowledge, which we can process scientifically, yet if we let it perish in the shadow of prosperous medical research, in domains will we most likely suffer losses.

**Keywords:** scientific research, discrepancy, citizen science, disparities, challenges, research ethics

## 1. Az információrobbanás mértéke

Az orvostudományi kutatások területén világszerte hatalmas információ-robbanás figyelhető meg. Egy 2011. évi tanulmány<sup>1</sup> szerint az orvosi és egészségtudományi információ duplázódási ideje 1950-ben még 50 év volt, 1980-ban már csak 7 év, 2010-ben 3,5 év, 2020-ra pedig 73 napot prognosztizáltak. A hazai orvosi bibliográfiai adatokat összevetve a PubMed orvosi adatbázis adatmennyiségével egyes területeken kiugró aránytalanságokat tapasztalhatunk. Eleddig nem volt az informatikai és könyvtári szakma mindennapi gyakorlatának része, hogy monitorozza és összevesse a hozzáférhető szakirodalmi források szegmenseit, az elemszámának eltéréseit. Mondhatnánk, nem a mi dolgunk, hogy például az irodalomtudomány mely forrásokhoz nyúl és melyeket hagy figyelmen kívül. Az orvos- és egészségtudományi adatbázisokat vizsgálva azonban nem jelenthetjük ki, hogy ez mellékes szempont volna. Mit tehetünk tehát mi, informatikai szakemberek, hogy felhívjuk az kiugró egyenlőtlenségekre a szakma figyelmét?

### 1.1 Horror vacui

A filozófiában és a fizikában egyaránt ismert tény, hogy a természet irtózik az ürességtől. Tökéletes vákuum földi viszonyok között nem létezik, mert a környezetben lévő anyagi részecskék a szívóhatás miatt akadálytalanul beáramlanak az üres térbe. Mi a helyzet az információs vákuum keletkezésével, fennmaradásával és megszűnésével? Vajon az informatikában másképp viselkedik a vákuum jelensége, tartósan fennmaradhat-e informatikailag „üres” állapot? A kérdés megválaszolásához hazai forrásunk a Magyar Orvosi Bibliográfia rendszerváltás óta fejlesztett nyílt hozzáférésű, elektronikus változata, nemzetközi forrásunk a PubMed, a National Library of Medicine, (USA) szabad elérésű orvosi és élettudományi adatbázisa. Vizsgáljuk a jelentősebb hazai könyvtárak attitűdjét egy ajándékküldemény kapcsán a kiadvány állományba vétele, OPAC-katalógizálása,

1 Peter DENSEN, „Challenges and opportunities facing medical education”, *Transactions of the American Clinical and Climatological Association* 122 (2011): 48–58.



hozzáférhetővé tétele, vagy annak elutasítása mentén. Szót ejtünk a nemzetközi könyvtári, könyvtárosi attitűdökről egy IFLA-felmérés előzetes eredményei kapcsán. Betekintünk a hazai tudományos kutatási nyilvántartás jelenleg hozzáférhető adataiba az NKFIH, a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal támogatott projektjeit véve alapul. Röviden vizsgálódunk a hazai egyetemek és kutatóhelyek kutatási preferenciáit tekintve a doktori témakiírások és kutatások területén. Végül a felmerülő kérdésekre keressük a választ.

## 2. A Magyar Orvosi Bibliográfia<sup>2</sup> adatainak vizsgálata 2017 és 2022 között

A MOB adatbázis online verziója 1993-tól tartalmazza a feldolgozott cikkek bibliográfiai tételeit, jelenleg 128 678 tételt. Próbaképpen 2017 áprilisában és 5 évvel később, 2022 áprilisában kulcsszavas keresést végeztem a magatartástudományhoz köthető legfontosabb kutatási területekről. A tágan értelmezhető találati halmazt (cím, tárgyszó, absztrakt) ez a módszer eredményezi. Az „étkezés” és „evés” kulcsszavas keresés találati halmazát összevontam, a duplikátumokat szűrtem. A találati halmazokat az 1. táblázatban vettem össze. A találatok alapján három csoportba sorolhatók a kutatási területek. Az agykutatás, valamint a gyógyszerkutatás kiugróan kutatott terület a táblázat szerint, majd ezt követik az átlagos szinten fejlődő tudományterületek, mint az alváskutatás, bioetika, evészavarok, kommunikáció és pszichoterápia. A legkevésbé vizsgált kutatási területek az antropológia, autizmus, homoszexualitás, hospice, szexualitás és szociológia.

	2017	2022	gyarapodás 2017–2022 (cikk)	gyarapodás 2017–2022 %-ban	átlagos gyarapodás/ év (cikk)
agy	1611	1695	84	5,2	16,8
alvás	536	770	234	43	46,8
antropológia	110	118	8	7,2	1,6
autizmus	43	81	38	88,3	7,6
bioetika	931	1009	78	8,4	15,6
étkezés, evés	676	933	257	37,9	51,4
gyógyszer	1969	2000*	n.a.	n.a.	n.a.
homoszexualitás	27	36	9	33,3	1,8
hospice	90	113	23	25,5	4,6
kommunikáció	573	813	240	41,9	48
pszichoterápia	653	824	171	26,2	34,2
szexualitás	181	211	30	16,6	6
szociológia	252	265	13	5,2	2,6

1.táblázat

Az adatbázisban a keresési felső limit 2000 tétel.

2 „Magyar Orvosi Bibliográfia, Bibliographia Medica Hungarica,” GYEMSZI, [adatbázis], hozzáférés: 2022. június 13., <https://mob.aeek.hu/simplesearch.jsp>



A MOB cikkadatbázis, nem terjed ki a szakterületen megjelent szakkönyvekre, e-bookokra, elektronikus megjelenésű publikációkra. Számos területen magasabb már az online kiadott folyóiratok és könyvek száma, mint a nyomtatottaké. Az agy kutatás hazai eredményeit főképp nemzetközi szaklapokban publikálják. A kapott eredmények tehát nem feltétlenül tükrözik a tudományterületek teljes publikációs spektrumát. A homoszexualitás és transzneműség témakörében 2022-ig 36 tételt találunk a MOB adatbázisban. 35 év alatt nagyjából évente 1 cikk íródott e szexuális tudományterületen, de mindössze ezek fele tartalmaz a gyógyításban használható információkat. A homoszexualitással foglalkozó cikkek aránya az online Magyar Orvosi Bibliográfia adatbázisban 0,0279 százalék. A kutatási területek publikációs számát tekintve a növekedési ütem erősödése ellenére is a legalacsonyabb kutatási hajlandóság a homoszexualitás területén mutatkozik hazánkban. A tanulmány további részében a vizsgálat fókuszát erre a negligált témakörre, a LGBTI/homoszexualitás kutatási területére szűkíttem.

### 3. PubMed adatbázis<sup>3</sup>

2022 tavaszán keresést végeztem a PubMed adatbázisban a homoszexualitás kutatásában használatos tudományos tárgyszavakat használva. Először kulcsszóként vizsgáltam a fogalmak előfordulását, majd a Medical Subject Heading teaurusz segítségével tárgyszavas kereséseket végeztem. Az 2. táblázat a PubMed adatbázisban 2022 áprilisában végzett keresés eredményeit mutatja. Az adatbázis közli a tárgyszavakhoz tartozó legkorábbi kiadású indexelt publikáció megjelenésének időpontját is. A kapott találati halmazok exponenciálisan növekvő görbét mutatnak a PubMed találati lista mellett szokásosan közölt grafikonon az idő függvényében.

	Kulcsszavas keresés	Tárgyszavas keresés (MeSH)
* Lesbian (1946–)	* 26.809	* ( 7.850)
* Gay (1853–)	* 52.860	* (21.327)
* Bisexuality (1920–)	* 21.412	* ( 4.980)
* Transgender (1992–)	* 10.614	* (10.158)
* Intersexual (1920–)	* 49.968	* ( 1.067)
* Homosexuality (1853–)	* 46.527	* (34.186)

2.táblázat

A PubMed-ben kapott találati halmaz aránya a teljes adatbázis tételszámához képest 0,3276%, szemben a MOB-ban eredményül kapott 0,0279%-os aránnyal. A két adatot összevetve több mint tízszeres eltérést találunk. A feldolgozás üteme és a hazai tudományos információáramlás alig tart lépést a nemzetközi szakirodalmi adatok integrálásával a homoszexualitás területén, mely az OECD legújabb adatai szerint<sup>4</sup> a lakosság 2,7 százalékát érinti. A rejtőzködés és a téma tabusítása miatt azonban az adat többszörösével érdemes számolniuk a döntéshozóknak.

3 „National Library of Medicine,” National Institute of Health, [adatbázis], hozzáférés: 2022.06.13., <https://pubmed.ncbi.nlm.nih.gov/>

4 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, *Society at a glance 2019: OECD social indicators*, (Paris: OECD Publishing, 2019), 9–10.

#### 4. Ajándékküldemény a jelentősebb hazai könyvtáraknak

2014 novemberében a Háttér Társaság ajándékként egy kiadványt<sup>5</sup> juttatott el a jelentősebb 54 hazai könyvtár számára a szexuális irányultság és nemi identitást érintő diszkrimináció csökkentését illetően. A kiadvány postázása után 3 illetve 10 hónappal, majd kettő és négy évvel később ismételt kereséseket végeztem a megajándékozott könyvtárak nyilvános katalógusában, monitorozva a kiadványok feldolgozottságát, szakozását, visszakereshetőségét, elhelyezését, és többes példány beszerzését.

A kiküldés után három hónappal a kiadványt feldolgozta és a nyilvános katalógusban kereshetővé tette 16 könyvtár, a könyvtárak 29 százaléka. Tíz hónappal utána feldolgozásra került és a nyilvános katalógusban kereshetővé vált további 13 könyvtárban, ez a könyvtárak további 24 százaléka. Az expediálás után két évvel végzett kereséskor nem találtam változást a korábbi állapothoz képest, végül 4 évvel később további négy könyvtár tette kereshetővé a nyilvános katalógusában a kiadványt. Ez a szám a könyvtárak további 7 százalékát jelentette. Így mindösszesen a könyvtárak 60 százaléka vette állományba és tette visszakereshetővé az ajándékküldeményt, 37%-uk pedig nem tette láthatóvá az olvasók számára a dokumentumot. A maradék 3% nem rendelkezett nyilvános katalógussal, illetve nem volt használható az OPAC-ja. Többes példányokat két könyvtár szerzett be, de mindössze egy könyvtár alkalmazta a tartalmi feltáráshoz a kísérőlevélben ajánlott 34 tárgyszót. Tíz könyvtárból érkezett köszönőlevél, viszont közülük mindössze hét vette leltárba az ajándékot, és három könyvtár megköszönte, de nem tette az olvasók számára elérhetővé a kiadványt.<sup>6</sup>

#### 5. Az IFLA LMBT+ Olvasók Szakmai csoport felmérése<sup>7</sup>

Az IFLA LMBT+ olvasók információs ellátását célzó, speciális érdeklődésű szakmai csoport az eddigi jó gyakorlatok összegyűjtése és a minimum szolgáltatások körét felölelő irányelv (guideline) kidolgozása céljából online kérdőívet<sup>8</sup> állított össze és tett közzé az IFLA által használt világnyelveken és további fordításokban, így angolul, arabul, franciául, németül, kínaiul, olaszul, oroszul, spanyolul és törökül. A felmérés azt vizsgálta, mennyiben ismerik fel és mennyiben tudják kielégíteni a könyvtárosok lesbikus, meleg, biszexuális és transznemű olvasók speciális, az általánostól merőben eltérő információigényét, és mennyiben képesek tájékoztatni az olvasóközönséget az LMBT emberek életét érintő hétköznapi kérdésekről. Az anonimizált kérdőív többek között a gyűjteményfejlesztésben az LMBT+ szak- és szépirodalmi források elérhetőségére, a célcsoport számára szervezett programok és szolgáltatások alkalmazására, a védett terek meglétére kérdezett rá. Érdeklődtek az LMBT könyvtári munkavállalók támogatásáról, érzékenyítő tréningek és továbbképzések elérhetőségéről is.

---

5 HÁTTÉR TÁRSASÁG, Összefoglaló jelentés az Európa Tanács Miniszteri Bizottsága szexuális irányultság, illetve nemi identitás alapján történő diszkrimináció leküzdését célzó intézkedésekről szóló CM/Rec(2010)5. számú ajánlásának magyarországi végrehajtásáról, hozzáférés: 2022.06.13., <https://hatter.hu/kiadvanyaink/osszefoglalo-jelentes-europa-tanacs>

6 HÁTTÉR TÁRSASÁG, DOMBOS Tamás és SOMORJAI Noémi, „Némileg tartunk az LMBT tartalomtól”, *Könyvtár Információ Társadalom* 18, 30. sz. (2019): s.d. hozzáférés: 2022.08. 26., [http://kithirlevel.hu/index.php?kh=nemileg\\_tartunk\\_az\\_lmibt-tartalomtol](http://kithirlevel.hu/index.php?kh=nemileg_tartunk_az_lmibt-tartalomtol)

7 „Survey of LGBTIQ+ Library Resources and Services,” International Federation of Library Associations and Institutions, hozzáférés: 2022.06.13., <https://www.ifla.org/news/survey-of-lgbtq-1-library-resources-and-services/>

8 „IFLA Survey of LGBTIQ+ Library Resources and Services,” IFLA LGBTQ+ Users Special Interest Group, hozzáférés: 2022.06.13., [https://stcloudstate.co1.qualtrics.com/jfe/form/SV\\_6WjZRGtQZ5bd4s5](https://stcloudstate.co1.qualtrics.com/jfe/form/SV_6WjZRGtQZ5bd4s5)



A kérdőív kitöltésére 2020 márciusától októberig volt lehetőség. A kérdéscsoportok az alábbiak voltak:

- \* Tájékoztató szolgálatok: szaktájékoztató könyvtárosok LGBT kérdésekről, LGBT gyűjtemény, külső információforrás/kapcsolat, amennyiben a könyvtár nem rendelkezik megfelelő LGBT-gyűjteménnyel
- \* Gyűjtemények: van-e LGBT szak- és szépirodalom, ha nincs, miért és hogyan pótolja a könyvtár, mik az akadályok
- \* Programok: célcsoportok, ki finanszírozza, a könyvtár és/vagy az LGBT közösség
- \* Könyvtári helyek: biztonságos tér, elvárt viselkedés, internetes szűrőprogram, mosdóhasználat, nem-semleges mosdó (gender-neutral bathroom)
- \* Vezetés és munkavállalók: Egyenlő tájékoztatás elve, készült-e ez irányú deklaráció, LGBT dolgozók, LGBT képzések
- \* Demográfiai információk: kitöltő beosztása, ország és könyvtártípus: nyilvános, egyetemi, iskolai, szak, kormányzati, levéltár, egyéb

Az adatfelvétel időszakában 236 fő töltötte ki a kérdőívet 8 nyelven. Az adatok feldolgozása és az irányelvek kidolgozása jelenleg folyamatban van.

## 6. NKFIH-EPR Kutatás-nyilvántartási adatbázis

A Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal Elektronikus Pályázati Rendszerben<sup>9</sup> visszakereshető kutatási pályázatok adatai szerint 2000-től 9626 kutatási projekt került be a rendszerbe. Ebből 61 foglalkozik szexualitással, ami magában foglalja a növény- és állatvilág, illetve az ember szexuális életével kapcsolatos összes projektet. A 61 témából mindössze 2 foglalkozik humán szexualitással, és abból az egyik fókuszál a homoszexualitásra, történeti nézőpontból, vagyis a teljes adatbázis közelítőleg tízezred része, 0,01%-a.

Más szempont szerint vizsgálva a 9626 kutatási kérdést 103 projektben említik a 'gender' fogalmat, százalékban kifejezve 1,07%, ezek közül pedig 15 foglalkozik a társadalmi nem kérdésével, pontosabban érinti a kérdést, százalékosan 0,15%.

## 7. Doktori adatbázis

A Doktori adatbázis<sup>10</sup> 2022 tavaszán készült lekérdezéskor 173.101 témakiírást tartalmazott, ezek között szerepelnek ismétlődő témakiírások is. (Az adatbázis nem nyújt lehetőséget konkrétan lekérdezésre.) Az összes témakiírásból 140 (ismétlődő) témacím foglalkozik szexualitással, ebből orvosi, pszichológiai vagy egészségügyi szempontból 37, a szexualitás/homoszexualitás témára szűkítve 5 témakiírás. Az orvosi témakörökben meghirdetett összes kutatási téma 0,021%-a foglalkozik szexualitással és homoszexualitással. Kiugróan alacsonynak tekinthető ez az arány.

9 „Elektronikus Pályázati Rendszer,” Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal, [Adatbázis], hozzáférés: 2022.06.13., <https://nkfi.gov.hu/palyazoknak/palyazatkezo-ertekelo/elektronikus-rendszer>

10 „Országos Doktori Tanács,” Országos Doktori Tanács, [Adatbázis], hozzáférés: 2022.06.13., <https://doktori.hu/>

## 8. Mit tehetnek az amatőr kutatók az elhanyagolt kutatási területek felzárkóztatásáért

Orvosi könyvtárosként feladatunk a részrehajlástól mentes, egyenlőségen alapuló tájékoztatás. A könyvtáros etikai kódex kimondja: „A használóknak joguk van ... a lehető legjobb könyvtári ellátáshoz és az egyenlő elbánáshoz, a könyvtáros ... törekszik az esélyek kiegyenlítésére. A könyvtáros ... mindent megtesz azért, hogy a használó szabadon és korlátozás nélkül hozzáférhessen az információkhoz.”<sup>11</sup>

Mit tehetünk, ha az információ rendelkezésünkre áll, csak nem anyanyelvünkön, hanem egy világnyelven: angolul? Mit tegyünk akkor, ha az információhiány súlyosan érinti a vizsgált olvasói csoport esélyeit és életkilátásait? Walker kutatásai szerint<sup>12</sup> első helyen az élet fenntartásához, az egészségmegőrzéshez szükséges információt keresik az LMBT csoportok tagjai, fiataljai. Éppen őket hogyan hagyhatnánk ki az információáramlásból? Bármi legyen is a LMBT csoporthoz tartozók társadalmi megítélése, a tájékoztatási 'FLOW'-ból való kirekesztődésük nem csak az érintetteknek, de az egész társadalomnak GDP-ben kifejezhető veszteséget okoz. Egyre több tanulmány foglalkozik a kérdéssel makrogazdasági szinten és kimutatják, hogy a teljes nemzeti jövedelemben mekkora hiányt eredményez munkaidő-veszteségben, termelékenység kiesésben, az emberi erőforrások alacsony fokú kihasználásában, ha nem ismerik eljogilag az LMBT+ embereket. Lee Badgett és munkatársai<sup>13</sup> kimutatták, nagyjából 2000 USD növekedést hoz az éves GDP-ben fejenként, ha integrálják a társadalomba az LMBT személyeket.

### 8.1. Saját kutatások

A szakirodalmi feltárást amatőr kutatóként kezdtem 2007-ben. Először a magyar nyelvű, majd az angol nyelvű könyvtári és orvostudományi ismereteket gyűjtöttem. A hazai szakirodalom nem hozott eredményt, könyvtári forrásokban LMBT olvasók ellátásáról nem volt találat, a MOB adatbázis 15 évvel ezelőtt mindössze 20 cikket adott, aminek elenyésző töredéke volt releváns egészségmegőrzési szempontból. Az angol nyelvű könyvtári források két monográfiát emeltek a fókuszba, az egyik<sup>14</sup> a felnőtt korosztály, a másik<sup>15</sup> a serdülőök könyvtári és információs ellátását helyezte szélesebb látószögbe. A feltárás nyomán 2012-ben posztert<sup>16</sup> nyújtottam be a Könyvtáros Világkongresszusra az LMBT olvasók könyvtári ellátásáról. A poszter jelentős érdeklődést keltett a résztvevők körében. A következő évben az IFLA Acquisition and Collection Development Section, az IFLA Szerzeményezési és Gyűjteményfejlesztési Szekció alá betagozódva megalakult az LGBTQ+ Users Special Interest Group, az LMBTQ+ (leszbikus, meleg, biszexuális, transznemű) Olvasók Speciális Szakmai Érdeklődésű Csoport.

---

11 „A magyar könyvtárosság etikai kódexe,” Magyar Könyvtárosok Egyesülete, hozzáférés: 2022.06.13., <https://mke.info.hu/wp-content/uploads/2010/08/alairtkodex.pdf>

12 Janine WALKER and Jo BATES, „Developments in LGBTQ provision in secondary school library services since the abolition of Section 28”, *Journal of Librarianship and Information Science* 48, 3. sz. (2016): 269–283. DOI: [10.1177/0961000614566340](https://doi.org/10.1177/0961000614566340)

13 Mary Virginia LEE BADGETT, Kees WAALDIJK, and Yana van der MEULEN RODGERS, „The relationship between LGBT inclusion and economic development: Macro-level evidence,” *World Development* 120, (2019): 1–14. DOI: [10.1016/j.worlddev.2019.03.011](https://doi.org/10.1016/j.worlddev.2019.03.011)

14 Ellen GREENBLATT, ed., *Serving LGBTIQ Library and Archives Users* (Jefferson: McFarland, 2010)

15 Hillias J. MARTIN and James R. MURDOCK, *Serving Lesbian, Gay, Bisexual, Transgender, and Questioning Teens*, (New York: Neal-Schuman Publishers, 2007)

16 SOMORJAI Noémi, „Libraries for Lesbian, Gay, Bisexual and Transgender Library Users”, Poster. World Library and Information Congress: 78th IFLA General Conference and Assembly, hozzáférés: 2022.06.14., [https://www.researchgate.net/publication/237216771\\_IFLA\\_poster\\_Helsinki\\_2012\\_somorjai\\_03](https://www.researchgate.net/publication/237216771_IFLA_poster_Helsinki_2012_somorjai_03)



A magyar orvosi irodalomban tapasztalható információhiány az angol nyelvű adatbázisok és más források feldolgozására inspirált. Ennek eredményeiből számos publikáció született. A hazai tudományos kutatás és információfeldolgozás késedelmét ellensúlyozza, hogy 2018-tól néhány orvosegyetemi tankönyvbe is integrálódik a LGBT témakör, valamint az Orvosi Hetilapban megjelent egy összefoglaló cikk<sup>17</sup>, ami több karon is ajánlott irodalomként szerepel. A használati statisztikát tekintve a megtekintések és letöltések száma alátámasztja a cikk megjelentetésének időszerűségét. Az első doktori disszertáció védeése a témából 2022-re várható.

## 9. Felmerülő kérdések

Laikusok és IT szakemberek számára is látható, hogy léteznek gyengén kutatott tudományos szakterületek. Az információhiány társadalmi és gazdasági szintéren egyaránt problémát okoz. Az elvégzett anyaggyűjtés, feldolgozás és közzététel folytán több kérdés is megfogalmazható az információtudományi szakma számára.

- \* Mit tegyünk az **exponenciálisan növekvő** kutatási adatokkal, melyek a nyilvánosan elérhető nemzetközi adatbázisokban **feldolgozatlanul** szunnyadnak, a hazai kutatóhelyek és érintettek látókörét elkerülve?
- \* Van-e felelősségük a könyvtárosoknak a nemzetközi szinten hozzáférhető adatok hazai kutatási körképbe integrálását tekintve?
- \* Mit tehetünk az egészségügyi információzárlat feloldása és az információáramlás felzárkóztatása érdekében?

Az egészségtudományokban az egészségmegőrzéshez köthető információ kellő időben történő integrálása az ellátók minőségi munkáját és az ellátottak minőségi ellátáshoz köthető jogát is érinti. A bioetika és a könyvtárosi etika szakmai szempontjai felerősítik egymást. Amíg a megfelelő orvosi ellátáshoz szükséges információ beépül a tantervekbe és a mindennapi életbe, a könyvtárosoknak ajánlott figyelemmel kísérni a szokásostól eltérő információigényű, információs korlátok közé szorított olvasókat. Az információ gyógyít, hiánya viszont csökkenti a családi kapcsolatok továbbélését, az egészséges életévek számát, a munkaképességet, mindemellett növeli a stresszhez köthető betegségek kialakulását, az önkárosító magatartás veszélyét és az öngyilkossági gondolatok előfordulását. Nem mindegy, meddig késlekedünk. Megfontolandó lenne a Magyar Könyvtárosok Egyesületében az LGBT+ Olvasók Könyvtáros Szakmai Csoport létrehozása az információs vákuum megszüntetésére és az LGBT+ olvasók könyvtári és információ ellátási színvonalának konvergálása céljából.

## Bibliográfia

DENSEN, Peter. „Challenges and opportunities facing medical education” *Transactions of the American Clinical and Climatological Association* 122, (2011): 48–58.

Elektronikus Pályázati Rendszer. Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal.  
<https://nkfih.gov.hu/palyazoknak/palyazatkezo-ertekelo/elektronikus-rendszer>.

GREENBLATT, Ellen, ed. *Serving LGBTIQ Library and Archives Users*. Jefferson: McFarland, 2010.

<sup>17</sup> SOMORJAI Noémi, „A nemi kisebbségekhez tartozó személyek orvosi ellátásának aktuális kérdései”, *Orvosi Hetilap* 162, 32. sz. (2021): 1267-74. DOI: [10.1556/650.2021.32119](https://doi.org/10.1556/650.2021.32119)



- „Összefoglaló jelentés az Európa Tanács Miniszteri Bizottsága szexuális irányultság, illetve nemi identitás alapján történő diszkrimináció leküzdését célzó intézkedésekről szóló CM/Rec(2010)5. számú ajánlásának magyarországi végrehajtásáról,” Háttér Társaság, 2013. <https://hatter.hu/kiadvanyaink/osszefoglalo-jelentes-europa-tanacs>.
- Háttér Társaság, Dombos Tamás és Somorjai Noémi. „Némileg tartunk az LGBT tartalomtól,” Könyvtár Információ Társadalom 18, no. 29 (2019) [http://kithirlevel.hu/index.php?kh=nemileg\\_tartunk\\_az\\_lmbt-tartalomtol](http://kithirlevel.hu/index.php?kh=nemileg_tartunk_az_lmbt-tartalomtol).
- MARTIN , Hillias J. and James R. MURDOCK. *Serving Lesbian, Gay, Bisexual, Transgender, and Questioning Teens*. New York: Neal-Schuman Publishers, 2007.
- „IFLA Survey of LGBTIQ+ Library Resources and Services,” IFLA LGBTIQ+ Users Special Interest Group, 2019, [https://stcloudstate.co1.qualtrics.com/jfe/form/SV\\_6WjZRGQTQZ5bd4s5](https://stcloudstate.co1.qualtrics.com/jfe/form/SV_6WjZRGQTQZ5bd4s5).
- WALKER, Janine and Jo BATES. „Developments in LGBTQ provision in secondary school library services since the abolition of Section 28.” *Journal of Librarianship and Information Science* 48, no. 3 (2016): 269–283. <https://doi.org/10.1177/096100061456663>.
- „A magyar könyvtárosság etikai kódexe,” Magyar Könyvtárosok Egyesülete, 2006, <https://mke.info.hu/wp-content/uploads/2010/08/alairtkodex.pdf>.
- Magyar Orvosi Bibliográfia, Bibliographia Medica Hungarica. GYEMSZI. <https://mob.aEEK.hu/simplesearch.jsp>.
- LEE BADGETT, Mary Virginia, Kees WAALDIJK, and Yana van der MEULEN RODGERS. „The relationship between LGBT inclusion and economic development: Macro-level evidence.” *World Development* 120, (2019): 1–14. <https://doi.org/10.1016/j.worlddev.2019.03.011>.
- National Library of Medicine. National Institute of Health. <https://pubmed.ncbi.nlm.nih.gov/>.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *Society at a glance 2019*. Paris: OECD, 2019.
- Országos Doktori Tanács. <https://doktori.hu/>.
- International Federation of Library Associations and Institutions. „Survey of LGBTIQ+[1] Library Resources and Services.” <https://www.ifla.org/news/survey-of-lgbtiq-1-library-resources-and-services/>.
- SOMORJAI, Noémi. „Libraries for Lesbian, Gay, Bisexual and Transgender Library Users.” Poster presented at the *World Library and Information Congress: 78th IFLA General Conference and Assembly*. Helsinki, August 2012. Also available at [https://www.researchgate.net/publication/237216771\\_IFLA\\_poster\\_Helsinki\\_2012\\_somorjai\\_03](https://www.researchgate.net/publication/237216771_IFLA_poster_Helsinki_2012_somorjai_03).
- SOMORJAI, Noémi. „A nemi kisebbségekhez tartozó személyek orvosi ellátásának aktuális kérdései.” *Orvosi Hetilap*, 162, no. 32 (2021): 1267–1274. <https://doi.org/10.1556/650.2021.32119>.

## Robotok a könyvtárban: Hogyan válhat a robotika a könyvtári mindennapok részévé?

Molnár Dániel

*Bács-Kiskun Megyei Katona József Könyvtár, informatikus-könyvtáros*  
[molnar.daniel@kjm.k.hu](mailto:molnar.daniel@kjm.k.hu)

Dani Erzsébet

*Debreceni Egyetem BTK, Könyvtár- és információtudományi Tanszék, tanszékvezető*  
*egyetemi docens*  
[dani.erzsebet@arts.unideb.hu](mailto:dani.erzsebet@arts.unideb.hu)

### Robots in the library: how can robotics become part of everyday library life?

As a library constantly looking for new opportunities, we have launched our robotics services with the mission of knowledge transfer. Our main aim is to introduce the basics of robotics: how it has become part of our everyday life and the opportunities it offers, which anyone can take part in, in order to prepare for the automation challenges of today and the near future.

We also run a weekly club session called Library Robotics Hour, mainly for school children. Here, participants can learn about Lego robots and the basics of robot programming, develop their analytical thinking and problem-solving skills in a playful way, and learn about the use and possibilities of smart devices in a controlled environment. The more experienced participants can deepen their knowledge with creative tasks requiring independent ideas. As we have two types of robots that can be programmed through different applications, they can learn how similar results can be achieved through different approaches.

Our shorter, one-session sessions for pre-school children or even for older age groups aim to raise interest in the history and potential of robotics, mainly through examples from previous studies or everyday life.

The exhibited robots can also be tried out, so that the theoretical knowledge can be quickly implemented into practice.

Our hands-on and experience-like solutions have brought robotics closer to many people in our library, and in smaller libraries and schools country-wide.

**Keywords:** library, robotics, Lego, presentation

## Bevezetés

Az egyre nagyobb teret nyerő automatizálás, a mesterséges intelligencia, és a hétköznapi életben is egyre több mindenre használatos okoseszközök világában minduntalan újabb és újabb ismeretek, kompetenciák elsajátítására van szükség. A gyerekeket olyan munkalehetőségek kívánalmaira kell felkészíteni, amik ma még akár nem is léteznek, de az idősebb nemzedékeknek is lehetőséget kell adni rá, hogy az újabb technikai lehetőségekkel, szemléletmódokkal találkozzanak, megismerjék őket, és ne idegenkedjenek tőlük. A Bács-Kiskun Megyei Katona József Könyvtárban, tudásátadó szerepének megfelelően erre a kihívásra adott egyik válaszuk a robotikai alapismeretek bemutatásának beemelése a könyvtári szolgáltatások közé.

## Építsünk szolgáltatást!

Eleinte nyolc, Lego Mindstorms EV3, majd miután az azokkal való munka alapján úgy láttuk, hosszabb távon is szeretnénk a robotikára alapozni, három Lego Mindstorms Inventor típusú robot került beszerzésre. Az, hogy Lego robotokkal dolgozunk, meghatározta, milyen irányban indulhattunk el, milyen jellegben valósíthattuk meg elképzeléseinket. A Lego robotok többféle motorja, szenzorai kifejezetten sok gyakorlati példa megvalósítására alkalmasak, programozásuk elsajátítása bár egyszerű, nagy mélységekkel rendelkezik.

Kérdés volt, milyen mód jelenjenek meg a robotok a közönség számára. Felmerült, hogy adott irodalmi művekre épülő, azok robotokkal megvalósuló feldolgozásával próbálkozzunk, amely során egyszerre lehet mélyíteni a szövegismeret, és megtanulni a programozás alapjait, de robotjaink fantasztikus filmekbe illő, karakteres látványviláguk miatt nehezen építhetők tetszőleges karakternek, így végül egy technikaibb megközelítésű, nem más művekre épülő irányba indultunk.

A már beszerzett robotok önálló megismerése mellett az interneten fellelhető gyakorlati példákat is áttekintettük, egyrészt inspirálódva belőlük, másrészt hátha rávilágítanak olyan lehetőségekre, amik az egyéni felkészülésből kimaradtak. A Youtube-on több hasznos videó található arról, hogyan integrálható sikeresen a robotika a könyvtári szolgáltatások körébe, ahogy a különféle írásos beszámolókból, részletes szakköri dokumentációkból tanulva is gondolkodhattunk azon, milyen szolgáltatási forma is lenne legközelebb lehetőségeinkhez.

A Lego nyújtotta szabadsággal élve robotjaink több összeépítési módját is kipróbáltuk, melyik formájuk lesz a legalkalmasabb arra, hogy a legegyszerűbben, a lehető legtöbb mindent át tudjunk adni az érdeklődőnek úgy, hogy mindenki egyforma módon fejlődhessen. Az EV3 robotoknál végül három, eltérő tulajdonságokkal rendelkező típus maradt meg:

- A robotok alapszintű kipróbálására, népszerűsítésére egy EV3RSTORM és egy R3PTAR nevű, látványos, játékos robot került összeépítésre. Ezekhez egy programozásra nem, csak távirányításra szolgáló alkalmazás is elérhető, ami a velük találkozó csoportoknál hamar elérte, hogy több résztvevő is többet szeretett volna tudni, így szinte láthatatlan módon váltak orientálhatóvá a mélyebb ismeretek megismerése felé.



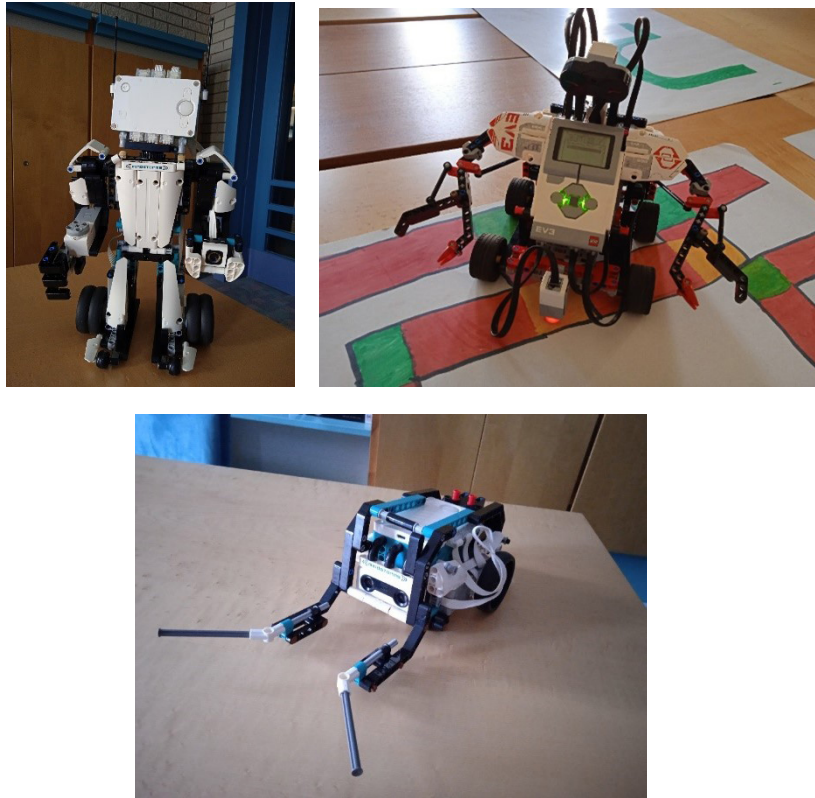
1. ábra: A látványosabb robotok

- A tanulásra, a programozás elsajátítására az EV3MEG nevű robotból készült el hat darab, a megadott építési útmutatóhoz képest egy apró saját hozzáadással, hogy minden szenzora kihasználható legyen. Kiscsoportos foglalkozások esetében mindenki ezt a robotot kapja meg, így egyforma mód fejlődhetnek, és nem lesz nehézség abból, hogy különböző összeépítésű robotok esetleg eltérő módon reagálnak.



2. ábra: A tanulásra alkalmas robot

- A későbbi időpontban beszerzett három Inventor típusú robot minden példánya más-más módon lett összeépítve, így összesen hat fajta robot kipróbálására van lehetőség. Az Inventor robotok, így, hogy egy-egy darab áll rendelkezésre belőlük, általában egyéni munkához kerülnek felhasználásra, saját, kreatív ötletek kidolgozásához.



3. ábra: Az Inventor robotok

A robotok építésével együtt a programozásukat is meg kellett tanulni. A programozófelület egy, kifejezetten a robotokhoz készült, ingyenes alkalmazáson keresztül érhető el, mind asztali számítógépes, mind táblagépes változatban. A választás a táblagépes változatra esett, egyrészt az eszközök mobilitása miatt, másrészt a később megvalósult foglalkozások élményszerűbbé tétele miatt.

Miután sikerült a robotok programozását a későbbi gyakorlati foglalkozások anyagát megelőlegező feladatokon keresztül megfelelő szinten elsajátítani, a következő feladat a megszerzett tudás továbbadásának módjának kitalálása lett. A következő kérdéseket kellett végiggondolni:

- Kik legyenek a célcsoport?
- Legyen-e a foglalkozás része a robotok építése is?
- Milyen helyszínt válasszunk?
- Hány alkalom szükséges ahhoz, hogy mindent, amit meg szeretnénk tanítani, átadhassunk?
- Akik már előzetes ismeretekkel rendelkezve jönnek hozzánk, hogyan tehetjük számukra is vonzóvá a látogatást?
- Megyei hatókörű könyvtárként, több, a városban található fiókkönyvtárral is, hogyan lehetne szolgáltatásunkat mobilizálni?
- Hogyan érhetjük el a célközönséget?

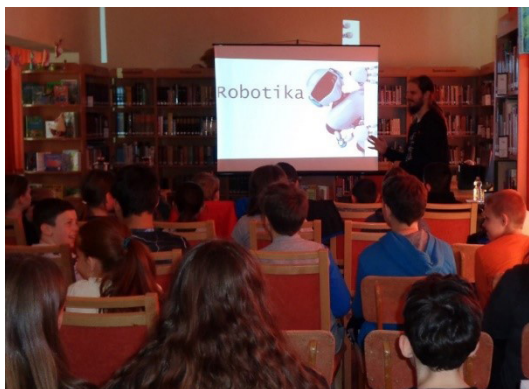


A kérdésekre a következő válaszokat adtuk:

- Elsődleges célcsoportként a fiatalabbakat, főképp a továbbtanulás és pályaeorientáció előtt álló általános iskolásokat neveztük meg, akiknek a Lego robotok világa tökéletes belépő lehet a robotika, a programozás, az automatizáció világába. Így egy első lépcsőfokként adhatunk nekik olyan tudást, élményeket, amikre későbbi tanulmányaikban, szabad idejük önálló eltöltésében építhetnek. Ezen kívül a többi korosztályt sem szerettük volna kizárni, így olyan foglalkozásformákat igyekeztünk megalkotni, ami akár a még fiatalabbak a nagyobb óvodások, vagy akár a felnőttek, nyugdíjasok igényeire is adaptálhatók.
- Bár a robotok építése mindenki számára izgalmas, és szórakoztató cselekvés, és működésükről is sokat lehet megtanulni megalkotásuk közben, végül úgy döntöttünk, a résztvevők előre megépített robotokkal dolgozhatnak. Mivel sok alkatrészből álló, sokszor nem a legegyszerűbben építhető készletekről van szó, amiket, hogy jó élmény legyen, mindenkinek a saját ütemében kéne építenie, ez viszont könnyen vezethet oda, hogy nagy egyéni különbségek adódhatnak a résztvevők között.
- Több, lehetséges helyszín végig gondolása után könyvtárunk Digitális Labor elnevezésű, zárt, vetítési lehetőséggel ellátott, több asztallal rendelkező helyisége lett. Itt kényelmesen lehet együtt, külön kis csoportokban, vagy egyénileg is zavartalanul dolgozni. Nagyobb létszámú csoportok esetén, például iskolai osztályoknál a Tinivilág nevű tágas, szintén vetítővel ellátott egységünk a helyszín. A könyvtári környezetben zajló élménytanulás erősítheti az érzést, hogy a könyvtár egy „jó hely”, és az így szerzett pozitív tapasztalatok a további szolgáltatásaink megismerésére sarkallhatnak.
- A megtartandó foglalkozási lehetőségek esetében két irányba indultunk el. Kifejlesztésre került egy egyalkalmas, vetítéssel kísért, játékos változat. Ez esetben a vetítés egy kultúr- és technikatörténeti keretbe helyezi a robotika témáját, az ókori elképzelésektől kezdve a fantasztikus művek világán át a kortárs, akár hétköznapi alkalmazási lehetőségekig. Ezt a kiállított robotok játékos, a távirányító alkalmazáson keresztüli kipróbálása követi, egyéni igény szerint akár egyszerűbb programozási feladatokkal bővítve. Ez a program a könyvtárban többször meg is lett hirdetve, és bárki eljöhett rá, illetve iskolás és táborozó csoportok számára is meg lett tartva. A másik irány egy klub jellegű sorozat lett, Könyvtári Robotóra címmel, ami heti rendszerességgel, könyvtári tagság mellett látogatható. Itt igény szerint lehet irányítottan, segítséggel tanulni, de akár önállóan is lehet felfedezni a robotok lehetőségeit, így azoknak, akiknek vannak előzetes ismeretei, de nem rendelkeznek saját robottal, is vonzó lehet a látogatás. Azoknak, akiknek segítségre van szükségük az ismeretek elsajátításához, készült egy öt alkalomra szabott foglalkozássorozat, és egy hozzá tartozó részletes segédlet, ami az alapoktól kezdve, egyre nehezebb, összetettebb feladatokat tartalmaz, kihasználva az EV3MEG robot minden motorját és érzékelőjét. Ezeket természetesen nem egyedül, hanem a foglalkozást felügyelővel együtt oldják meg, és a végére képesek lesznek az önálló feladatmegoldásra, és akár saját ötletek megvalósítására is a későbbi látogatások alkalmával.
- Mivel a robotok viszonylag kis méretűek, és a velük való munka is hordozható eszközön zajlik, így könnyen megoldható, hogy alkalomszerűen a könyvtár épületén kívül is lehessen dolgozni velük. Ilyenkor az egyalkalmas, előadással és kipróbálási lehetőséggel egybekötött változatot tudjuk felajánlani. Kistelepülési könyvtárakba már több alkalommal eljutottunk ilyen módon, így azokhoz is közelebb tudjuk hozni a robotika élményét, akik lakóhelyük lehetőségei miatt nehezebben ismerkedhetnek meg vele. Ezen kívül az egyik helyi általános iskolában is több alkalommal tartottunk bemutatókat.
- Mivel egy olyan, újszerű szolgáltatást próbáltunk bevezetni, ami potenciálisan olyan



érdeklődőknek is vonzó lehet, akik nem, vagy csak ritkán látogatják a könyvtárat, így a hagyományos marketinglehetőségeken (plakát, városi programajánlóban megjelenés) kívül egy néhány perces, saját ötlet alapján elkészült kisfilm forgattunk könyvtárunk Kreatív Stúdiójával, több lelkes munkatársunk bevonásával. A film Könyvtári robotexpedíció címmel megtekinthető a Youtube-on is<sup>1</sup>. A nem hagyományos szolgáltatásbemutatóként megvalósított, hanem rövid történettel bíró videó beváltotta a hozzá fűzött reményeket, felkeltette az érdeklődést a Robotóra iránt.



4. ábra: Foglalkozások kistéleplési könyvtárban, és az egyik kecskeméti általános iskolában

## Rakjuk össze, amit tudunk!

Azon kívül, hogy kellemes időtöltés és játékos jelleggel bír, a Lego robotokkal való foglalkozás több, a komolyabb robotoknál, gépeknél, programozási feladatoknál felmerülő fogalmat, problémát, az algoritmikus gondolkodásmódot fejlesztő megoldást bemutat. Néhány példa ezekre:

- A robotok több, különböző módon működő motorral mozognak, és ezek is többféleképpen állíthatók be: más lesz az eredmény, ha fordulat, szög, vagy másodperc alapján adjuk meg nekik az értékeket, ahogy a motorok egyszerre többféleképpen történő programozása is sok kimenetelre adnak lehetőséget.
- Többféle - infravörös, színérzékelő és nyomógombos - szenzor is rendelkezésre áll. Ezekkel megtanulhatják, hogyan lehet a különféle külső ingerekkel feladatokra bírni a robotokat. A szenzorok is többféleképpen programozhatók, és néha egy apró változtatás is teljesen más eredményt hoz a program lefutásában, mint előzetesen várható. Így nagy tere van a kísérletezésnek, a különböző lehetőségek akár újra és újra kipróbálásának, és az így tanultak a továbbiakban beépítésének.
- A robot folyamatos visszajelzést küld a programozóalkalmazásnak, – a töltöttségi állapotát, a motorok és a szenzorok aktuális állását, miket érzékelnek, az egyes komponensek aktuálisan a robot központi egységének éppen melyik bemenetébe vannak csatlakoztatva – és ezeknek a visszajelzéseknek a folyamatos figyelemmel kísérése sok, enélkül elsőre érthetetlennek tűnő problémát kiküszöböl. Így megtanulható, hogy a programozás megtörténte és a célhardver megléte önmagában még nem mindig hoz kielégítő eredményt, az elsőre teljesen automatikusnak tűnő feladatok esetén is szükség van kontrollra.
- Lehetőség van a robotok párhuzamos programozására – azaz egy programban több, egyszerre zajló, egymástól független utasítássorozat fut. Ezzel gyakorolható az összetettebb, több szálon való gondolkodást fejlesztő feladatok elvégzése.

1 [https://www.youtube.com/watch?v=Aeu-ZPO\\_xD0&t=56s](https://www.youtube.com/watch?v=Aeu-ZPO_xD0&t=56s)

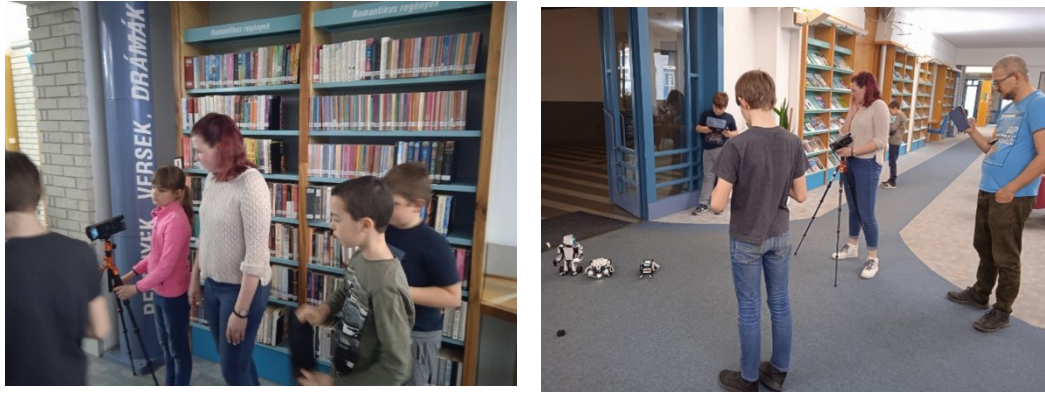
- A két robottípus, az EV3 és az Inventor bár első ránézésre nagyon hasonlítanak, és hasonló elemekből épülnek fel, másfajta központi egységgel bírnak, és nem ugyanazzal az alkalmazással programozhatók. Az EV3-k programozására használt alkalmazás – amelynek általunk használt változata sajnos 2021 nyaratól nem támogatott, a Google Play áruházból már nem letölthető, de továbbra is szabadon használható – grafikus blokkjai után sokaknak jelent nagyváltást az Inventor alkalmazásának használata, hiszen az eltérő felület mellett valamennyire eltérő logika alapján nyújtják a programozási lehetőségeket, viszont egymás után használva a két robot-és alkalmazástípust hamar megoldható lesz hasonló eredmények elérése különböző felületeken keresztül.

Ahogy az előbbieken során is megemlítsük, nagy hangsúlyt fektetünk arra, hogy mindenki az egyéni ötleteit megvalósíthassa, fejlesztve a problémamegoldó készséget, a kreativitást, bátorítva a résztvevők közötti együttműködést is az egyes felmerülő akadályok megoldására. Jó példa erre, amikor a résztvevők saját feladatokat találnak ki, akadálypályákat, a robotok által követhető útvonalakat alkotva, amelyeket utána mind maguk, mind társaik megpróbálnak oldani. Rendszeresen szokták az elkészült, saját megoldásokat tartalmazó programok alapján a robotok tudását összemérni: kisebb, sokszor spontán robotversenyek is kialakulhatnak: kinek a robotja erősebb, gyorsabb, kinek a programozási megoldása a jobb. Az így létrejött csoportkohézió fejleszti az együtt tanulást, a segítőkészséget, a feladatmegosztást, a csoportban való munkára való készséget.



5. ábra: A Robotóra pillanatai

Az egyéni kezdeményezéseknek, és a megszerzett tudás gyakorlatban bemutatásának legfrissebb példája a klub keretén belül a robotokkal való filmkészítés. Ebben segítségünkre – ahogy a népszerűsítő kisfilmnél is történt – könyvtárunk Kreatív Stúdiójának munkatársa van. A Kreatív Stúdió rendszeres lehetőséget biztosít a filmkészítés, a forgatókönyvírás, a kamera használatának és a vágás folyamatának megismerésére. A projektben részt vevők maguk találták ki a film történetét, hogy melyik robotnak milyen szerepet szánnak, és ezt hogyan kívánják megvalósítani. Eldöntendő kérdés, hogy a felvételek alatt a robotok előre programozott utasításokat hajtsanak végre, vagy emberi beavatkozással, távirányítással működtessék őket, vagy akár egyidejűleg, az adott jelenettől függően mindkét megoldással éljenek. A korábbi feladatokhoz képest kihívás, hogy az eszközök nem csak elszigetelten, vagy bár együtt, de csak egy tulajdonságot összevetve dolgoznak, hanem a narratív szempontokat is figyelembe kell venniük.



6. ábra: A filmforgatás pillanatai

### Miután minden elem összeállt

Az első robotikai foglalkozásokra könyvárunkban 2020-ban került sor, eleinte egyalkalmas, bemutató jellegű eseményekként. Ezt egy meghívásos alapon összeállt pilot csoporttal való, kísérleti jellegű sorozat követte, akik az addig elkészült, a programozás gyakorlását segítő feladatokat oldották meg több alkalmon keresztül. Miután a pilot csoporttal történő munkát sikeresnek ítéltük, és úgy véltük, robotikai szolgáltatásunk indulásra kész, a pandémiás helyzet keresztülhúzta terveinket. Így végül csak 2021. őszén – a nyári szünetben és az iskolakezdés első heteiben nem tartottuk szerencsésnek a kezdést – sikerült elindítani a Könyvtári Robotórát. Azóta rendszeres visszajárók és egyszeri érdeklődők is egyaránt folyamatosan látogatják az eseményt. A rendszeresen járóknál a fejlődés látványos, a robotok programozását könnyedén el tudták sajátítani, problémamegoldó és együttműködő képességeik is fejlődtek. Többen közülük kifejtették, hogy hosszabb távon is szívesen foglalkoznának robotikával, és célzottan keresnek az ilyen irányú továbbtanulási, vagy további iskolán kívüli képzési lehetőségeket. Így elmondhatjuk, hogy célunk, hogy a robotizáció megismerésének első lépcsőfoka legyünk, és azokhoz is közelebb hozzassuk ezt a tudományágat, akiknek amúgy nem feltétlen lenne lehetősége első kézből találkozni vele, sikeresnek tekinthető.

Digitalizáció a szakképzésben  
A Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek  
felülvizsgálata

Digitalisation in vocational education and training  
Review of digital competence proficiency levels for professions in the Register of  
Vocational Occupations

Horváthné Felföldi Helga  
*Digitális Jólét Nonprofit Kft. (Budapest)*  
[felfoldi.helga@djnkft.hu](mailto:felfoldi.helga@djnkft.hu)

### Absztrakt

A Szakképzés 4.0 Stratégia alapján a digitális készségek elsajátítása alapkövetelmény a szakképzésben is, ennek támogatása a Kormány kiemelt célja. A Digitális Jólét Nonprofit Kft. (a továbbiakban: DJ NKft.) keretében működő Digitális Szakképzési és Felnőttképzési Módszertani Központban (a továbbiakban: DSZFMK) a Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek vizsgálata a Magyar Kereskedelmi és Iparkamara és a Nemzeti Agrárgazdasági Kamara megbízásából előzetes kísérleti fejlesztésként valósult meg. A projekt célja a 12/2020. (II. 7.) Korm. rendelet 1. sz. melléklete szerinti 2021. 03. 01-jén hatályos Szakmajegyzékben szereplő szakmák képzési és kimeneti követelményeinek elemzésével a szakmák digitáliskompetencia vizsgálata a DSZFMK által meghatározott módszertan alapján. A módszertan alapját képezte a GINOP-6.1.2-15-2015-00001 „Digitális szakadék csökkentése” című kiemelt projekt keretében a DJ NKft. által kidolgozott DigKomp Állampolgári Digitáliskompetencia-keret. A kísérleti fejlesztés eredményei segíthetik a hatékony oktatást, értékelést, a minőségi tananyagok, taneszközök elkészítését a Szakképzés 4.0 Stratégia céljainak megfelelően.

**Kulcsszavak:** digitális kompetencia, digitális kompetencia keretrendszer, DigKomp rendszer, digitáliskompetencia-profil, szakmai digitális tevékenység, digitáliskompetencia-készlet

### Abstract

According to the Vocational Education and Training 4.0 Strategy, the acquisition of digital skills is a basic requirement in vocational education and training, and the Government's priority is to support this. At the Digital Welfare Nonprofit Ltd. (hereinafter: DJ NKft.) operating within the Digital Vocational Training and Adult Education Methodology Centre (hereinafter: DSZFMK), the digital competence proficiency levels of the professions listed in the Register of Occupations were tested as a preliminary pilot development on behalf of the Hungarian Chamber of Commerce and Industry and the National Chamber of Agriculture. The aim of the project was to analyse the training and output requirements of the professions included in the Register of Occupations in force on 01.03.2021, according to Annex 1 of Government Decree 12/2020 (II. 7.), and to examine the digital competence of the professions on the basis



of the methodology defined by the DSZFMK. The methodology is based on the Hungarian Digital Competence Framework for Citizens developed by DJ NKft. within the framework of the GINOP-6.1.2-15-2015-00001 priority project „Reducing the Digital Divide”. The results of the pilot development can help effective teaching, assessment, preparation of quality teaching materials and tools in line with the objectives of the VET 4.0 Strategy.

**Keywords:** digital competence, digital competence framework, DigKomp System, digital competence profile, digital professional activity, digital competence kit

## 1. Bevezető

A Szakmajegyzékben szereplő szakmák digitáliskompetencia jártassági szintjeinek felülvizsgálata a Digitális Jólét Nonprofit Kft. keretében működő Digitális Szakképzési és Felnőttképzési Módszertani Központban a Magyar Kereskedelmi és Iparkamara és a Nemzeti Agrárgazdasági Kamara megbízásából előzetes kísérleti fejlesztésként valósult meg.

A fejlesztés 2021 májusában kezdődött és 2021. november 30-ával zárult. A fejlesztés célja a 2021. március 1-jétől hatályos Szakmajegyzékben szereplő 175 szakma képzési és kimeneti követelményeinek (a továbbiakban: KKK) vizsgálatával a szakmák szakmai digitáliskompetencia-készleteinek elkészítése, és a digitáliskompetencia jártassági szintjeik beazonosítása volt.

## 2. Előzmények

A fejlesztés távlati stratégiai célja a Kormány által elfogadott Szakképzés 4.0 – a szakképzés és felnőttképzés megújításának középtávú szakmapolitikai stratégiája, a szakképzési rendszer válasza a negyedik ipari forradalom kihívásaira című stratégia [1] (a továbbiakban: Szakképzés 4.0) céljainak teljesítése, az egyre inkább digitalizálódó munkaerőpiaci igényhez történő igazodás. A Szakképzés 4.0 3. beavatkozása: Minden képzésbe be kell épülnie az Ipar 4.0 követelményeinek és a szakmában elvárt digitális tartalmaknak, mely alatt előírja, hogy „Ágazatonként szükséges definiálni, hogy a digitalizáció milyen tudást, képességeket és készségeket igényel. Minden szakmai képzésbe be kell épülnie a digitális tartalmaknak, a kimeneti követelményekben meg kell jelennie a gyakorlatban alkalmazható digitális tudásnak.” [2]

2019 novemberében megjelent a szakképzésről szóló 2019. évi LXXX. törvény [3], mely rendelkezik arról, hogy a szakmákhoz KKK-at kell előírni és kötelezően alkalmazni. A KKK-k a szakmai oktatásban az ellenőrzési, a mérési és az értékelési rendszer kialakítását és működését biztosítják.

2020 februárjában megjelent a szakképzésről szóló törvény végrehajtásáról szóló 12/2020 (II.18) Korm. rendelet [4], melynek 1. számú melléklete tartalmazza a Szakmajegyzéket. A Szakmajegyzékbe új elemként került be minden egyes szakmának a Digitális Kompetencia Keretrendszer szintet jelölő számértéke. 2019-ben, amikor a szakmák Digitális Kompetencia Keretrendszer szintjei meghatározásra kerültek, a 2017-ben közzétett European Digital Competence Framework for Citizens, vagy röviden DigComp 2.1 Állampolgári Digitáliskompetencia-keret [5] leírása állt rendelkezésre.

A Digitális Kompetencia Keretrendszer fejlesztéséről és bevezetésének lépéseiről szóló 1341/2019. (VI. 11.) Korm. határozat [6] értelmében a GINOP-6.1.2-15-2015-00001 „Digitális szakadék csökkentése” c. kiemelt projektben a Digitális Jólét Nonprofit Kft.



által elkészült a Digitális Kompetencia Keretrendszer, azaz a DigKomp Rendszer szakmai koncepciója. A koncepciónak kiemelt eleme az a hazai állampolgári digitáliskompetencia-keret, mely kompatibilis az európai uniós DigComp 2.1 ajánlással. A DigKomp Állampolgári Digitáliskompetencia-keret kidolgozásakor az adaptálás során a hazai környezetben releváns digitális készségkészletet, illetve ezekre vonatkozó példatár készült az 5 kompetenciaterület, 21 kompetenciaeleme, 8 jártassági szintje mentén, amely alapul szolgált a fejlesztésben a KKK-k digitáliskompetencia vizsgálatához.

### 3. DigKomp Állampolgári Digitáliskompetencia-keret<sup>1</sup>

A DigKomp Állampolgári Digitáliskompetencia-keret 5 területe és a hozzájuk tartozó 21 eleme:

#### 1. Információk és adatok kezelése, használata

- 1.1. Adatok, információk és digitális tartalmak böngészése, keresése és szűrése
- 1.2. Adatok, információk és digitális tartalmak kiértékelése
- 1.3. Adatok, információk és digitális tartalmak kezelése

#### 2. Kommunikáció és együttműködés

- 2.1. Digitális technológiával támogatott interakció
- 2.2. Digitális technológiával támogatott megosztás
- 2.3. Digitális eszközökkel támogatott társadalmi jelenlét és szerepvállalás
- 2.4. Digitális eszközökkel támogatott együttműködés
- 2.5. Netikett
- 2.6. A digitális személyazonosság kezelése

#### 3. Digitális tartalmak létrehozása

- 3.1. Tartalom fejlesztése
- 3.2. Digitális tartalmak integrálása és átalakítása
- 3.3. Szerzői jogok és licencek
- 3.4. Programozás /algoritmikus gondolkodás/

#### 4. Biztonság

- 4.1. Az eszközök védelme
- 4.2. A személyes adatok és a magánszféra védelme
- 4.3. Az egészség és a jóllét védelme
- 4.4. A környezet védelme

#### 5. Különbéféle problémák kezelése

- 5.1. Technikai problémák megoldása
- 5.2. Igények és technológiai válaszok megfogalmazása
- 5.3. A digitális technológia kreatív alkalmazása
- 5.4. A digitáliskompetencia-hiány azonosítása

1 GINOP-6.1.2-15-2015-00001 „Digitális szakadék csökkentése” c. kiemelt projekt eredménytermék



## A DigKomp Állampolgári Digitáliskompetencia-keret (DigKomp ÁDKK) 8 jártassági szintje:

	Alapszint		Középszint		Haladó szint		Mesterszint	
Szintek a DigKomp ÁDKK-ban	1	2	3	4	5	6	7	8
A feladatok összetettsége	Egyszerű feladatok	Egyszerű feladatok	Jól meghatározott, rutinszerű feladatok, egyértelmű problémák	Feladatok, jól meghatározott, nem rutinszerű problémák	Változatos feladatok és problémák	A tevékenységek többségének optimális megválasztása	Komplex problémákra adható megoldások testreszabása	Rendkívül összetett, sok tényező által befolyásolt problémák önálló megoldása
Önállóság	Segítséggel	Önállóan, szükség esetén segítséggel	Önállóan	Önállóan, saját igények szerint	Mások segítése	Másokhoz való alkalmazkodás komplex helyzetben	Integrálás a szakmai gyakorlat fejlesztése és mások segítése érdekében	Új ötletek és folyamatok ajánlása az adott területen
Kognitív terület	Emlékezés	Emlékezés	Megértés	Megértés	Alkalmazás	Értékelés	Létrehozás	Létrehozás

1. ábra: A jártassági szintekhez kapcsolódó fontosabb kulcsszavak<sup>2</sup>

Például a 3-as középszintű jártassági szinttel rendelkező személy önállóan old meg jól meghatározott, rutinszerű (egyszerű) feladatokat, problémákat, képes követni a lépéseket, érti a megoldásokat.

### 4. Fejlesztés folyamata

A fejlesztés két szakaszban valósult meg. Az első szakaszban pilotként a megbízást adó két kamara által kiválasztott 8 szakma vizsgálata történt meg, majd ezt követte a további 167 szakma vizsgálata.

A fejlesztéshez módszertani útmutató és sablon került kidolgozásra az új szakképzési struktúrának, kimeneti szabályozásnak megfelelően.

A fejlesztés megvalósításához többségében olyan szakértők kerültek felkérésre, akik részt vettek a KKK-k kidolgozásában, ezzel biztosítva a KKK-k és a tanulási eredményalapú megközelítés alapos ismeretét. A fejlesztést a szakértők felkészítése előzte meg, melynek keretében a DigKomp rendszer, az elemzések folyamata, módszertana és a fejlesztés ütemezése került bemutatásra.

A pilotot követően workshopot szerveztünk a fejlesztés tapasztalatairól, mely alapján véglegesítésre került a módszertani útmutató és sablon. Jelentős mértékben nem történt változás, csupán az elméletben kidolgozott módszertani eljárás finomhangolására volt szükség.

<sup>2</sup> GINOP6.1.2-15-2015-00001 „Digitális szakadék csökkentése” c. kiemelt projekt eredménytermék



A fejlesztés időpontjában hatályos Szakmajegyzék 24 ágazatban 175 szakmát tartalmazott, ebből 132 szakma esetében nem volt szakmairány, 43 szakmához azonban 143 szakmairány tartozott. A szakmák és a különböző szakmairányok magas száma indokoltá tette, hogy közel 200 szakértő vegyen részt a fejlesztésben.

Először a 23 ágazati alapoktatás követelményeinek digitáliskompetencia vizsgálata valósult meg munkacsoportok keretében. A munkacsoportok az ágazati alapoktatáshoz tartozó szakmák számától függően 2–6 főből álltak. Ezt követően történt a szakmai oktatás (szakmák, szakmairányok közös és a szakmairányok) követelményeinek digitáliskompetencia vizsgálata.

## 5. A szakértői munka folyamata, a fejlesztés eredményei

A KKK-ban a szakmai követelmények tanulási eredményekként kerültek megfogalmazásra a közismert négy deskriptor - Készségek, képességek; Ismeretek; Elvart viselkedésmódok, attitűdök; Önállóság és felelősség mértéke - mentén. A fejlesztésben résztvevő szakértők a tanulási eredményeket vizsgálva, a digitáliskompetencia szempontjából releváns tanulási eredményekhez megfogalmazták a szakmai digitális tevékenységeket. A szakértők a szakmai digitális tevékenységek meghatározásához TEA (tanulási eredmény alapú) módszertant alkalmazva aktív cselekvő igék használatával kontextusba helyezett tevékenységeket fogalmaztak meg.

A szakértőknek lehetőségük volt arra, hogy amennyiben szükségesnek látták, olyan új szakmai digitális tevékenységeket is definiáljanak, amelyek fontosak a szakma szempontjából, de a szakma KKK-ja jelenleg nem tartalmaz azokhoz kapcsolható tanulási eredményeket. Az ágazati alapoktatások és a szakirányú oktatások követelményeinek vizsgálatánál is egy-egy esetben került erre sor. Az ágazati alapoktatás esetében munkavállalói ismeretekkel, a szakirányú oktatásnál digitális önképzéssel kapcsolatos szakmai digitális tevékenységet fogalmaztak meg a szakértők, amelyek más szakmánál a tanulási eredményekhez köthetően szerepeltek. A szakmai digitális tevékenységhez a tanulási eredmény megírása nem volt feladata a szakértőknek.

A szakértők a DigKomp Állampolgári Digitáliskompetencia-keret kompetenciaelemeinek figyelembevételével fogalmazták meg a szakmai digitális tevékenységeket. Megvizsgálták tanulási eredményenként, hogy a 21 kompetenciaelem közül melyek relevánsak az adott tanulási eredményhez. A szakértők minden egyes, általuk meghatározott szakmai digitális tevékenységgel kapcsolatosan megállapították, hogy a DigKomp Állampolgári Digitáliskompetencia-keret releváns kompetenciaelemei mely jártassági szintnek felelnek meg.

KKK tanulási eredmények					Információk és adatok kezelése, használata			
Készségek, képességek	Ismeretek	Elvárt viselkedésmódok, attitűdök	Önállóság és felelősség mértéke	A tanulási eredményekben megjelentendő szakmai digitális tevékenységek	Információk és adatok kezelése, használata			
					Adatok, információk és digitális tartalmak böngészése, keresése és szűrése	Adatok, információk és digitális tartalmak kiértékelése	Adatok, információk és digitális tartalmak kezelése	Kompetencia terület legmagasabb DigKomp jártassági szintje
Megszervezi és lebonyolítja a különböző rendezvényeket a megrendelő igényei alapján.	Ismeri a rendezvényszervezés alapjait, a beállítottal való kapcsolattartás protokollját.	Munkája során törekszik a tökéletes kivitelezésre, rugalmasan és legjobb tudása szerint kezeli a felmerülő problémás szituációkat.	Kollégáival szoros együttműködésben tervezi meg és bonyolítja le a rendezvényeket. A rá bízott feladatokat teljes felelősséggel végzi el a rendezvény sikeressége céljából.	A rendezvény előkészítése során táblázatkezelő szoftver segítségével előkalkulációt, eszközlistát, munkaidőbeosztást, nyersanyagösszesítést készít. A lebonyolítást követően elszámolást készít. Szövegszerkesztő használatával, ajánlatot, visszaigazolókat készít, e-mailt, és továbbít megrendelője részére. A lebonyolításhoz szükséges menükártyákat, irányító és információs táblákat egyaránt elkészíti szövegszerkesztő használatával. Kiadványszerkesztővel meghívót, menükártyát, belépőkártyát készít és elküld nyomdai kivitelezéshez. A rendezvény lebonyolítása után digitális számlát állít ki és szükség esetén továbbít Megrendelője részére.	4	3	4	4
Irányítja a beosztott munkatársait, megtervezi a munkafolyamatokat a megfelelő munkavégzés érdekében, az üzlet napi feladatainak megvalósítása során.	Ismeri a beosztással járó munkaköri feladatokat, a vállalkozásnál működő hierarchiát, az alá- és fölé rendeltségi viszonyokat, felelősségi köröket és hatásköröket.	Törekszik az igényes munkavégzésre, munkatársaitól is maximális precizitást vár el. Munkaköri feladatát legjobb tudása szerint teljes körűen ellátja. A pozitív munkahelyi hangulat megteremtésére és megőrzésére nagy hangsúlyt fektet.	Önállóan irányítja munkatársai tevékenységét, azt rendszeresen ellenőrzi, a belső szabályzatokat betartja és betartatja. Szükség esetén gyors döntéseket hoz. Egyes problémás esetekben a vállalat vezetésével egyeztet.	Munkatársai részére munkaidő beosztást készít táblázatkezelő szoftver segítségével. A munkaidőnyilvántartást, valamint az ütköztség elszámolásához szükséges dokumentumokat szövegszerkesztő és táblázatkezelő segítségével készíti. Munkatársait e-mail úton vagy közös levelezőrendszeren keresztül digitális formában folyamatosan értesíti a változásokról.	4	3	4	4

2. ábra: Digitáliskompetencia jártassági szint beazonosítása (részlet)<sup>3</sup>

A DigKomp Állampolgári Digitáliskompetencia-keret releváns kompetenciaelemei jártassági szintjeinek beazonosítását követően a szakértők javaslatot tettek a DigKomp alapú digitális profil megalkotásához az 5 kompetenciaterület jártassági szintjére az ágazati alapoktatás és a szakma (szakirányú oktatás) követelményeinek tekintetében. Ezáltal meghatározásra került a 23 ágazati alapoktatás és a 175 szakma javasolt digitáliskompetencia-profilja.

Digitáliskompetencia-profil a DigKomp Állampolgári Digitáliskompetencia-keret alábbi öt területéhez rendelve a jártassági szintet jelölő szám:

1. Információk és adatok kezelése, használata
2. Kommunikáció és együttműködés
3. Digitális tartalmak létrehozása
4. Biztonság
5. Különböző problémák kezelése

Az öt DigKomp kompetenciaterület esetében az egyes ágazati alapoktatások és szakirányú oktatások (szakmák, szakmairányok) követelményeiben a digitáliskompetencia különböző szintű felkészültséget igényelhet, különböző jártassági szinten jelenhet meg. A digitáliskompetencia-profilban található öt db számérték tehát a fent felsorolt öt digitáliskompetencia területhez tartozó jártassági szinteket jelöli, melyek különbözőek is lehetnek.

Példaként az alábbi 3. ábrán az érettségi végzettséghez nem kötött, szakképző iskolában nappali rendszerű szakmai oktatásban 3 éves képzési idővel oktatható Pincér-vendégtéri szakember szakma esetében az ágazati alapoktatás követelményeinek digitáliskompetencia jártassági szintjét mind az öt DigKomp kompetenciaterületen 3-as középszintre, míg a szakirányú oktatás követelményeinek digitáliskompetencia jártassági szintjét három DigKomp kompetenciaterülethez 4-es (közép) szintre és két területhez 5-ös (haladó) jártassági szintre javasolták a szakértők.

3 Pincér-vendégtéri szakember szakma projekt eredménytermék részlet

A szakma alapadatai	
Az ágazat megnevezése:	Turizmus - vendéglátás
A szakma megnevezése:	Pincér-vendégtéri szakember
A szakma azonosító száma:	4 1013 23 04
A szakma szakmairányai:	
Ágazati alapoktatás megnevezése:	Turizmus-vendéglátás ágazati alapoktatás
Szakmajegyzékben a Digitális Kompetencia Keretrendszer szintje:	5
Ágazati alapoktatás javasolt digitáliskompetencia jártassági szintje:	3 3 3 3 3
<b>Szakma javasolt digitáliskompetencia jártassági szintje:</b>	<b>4 4 4 5 5</b>

Követelmény	Információk és adatok kezelése, használata	Kommunikáció és együttműködés	Digitális tartalmak létrehozása	Biztonság	Különféle problémák kezelése
Szakma/szakmairányok közös követelményei	4,0	4,0	4,0	5,0	5,0
Szakmairányok	0				
	0				
	0				
	0				
	0				
Szakma digitáliskompetencia jártassági szintje (átlag):	4,0	4,0	4,0	5,0	5,0

3. ábra: Digitáliskompetencia-profil javaslat<sup>4</sup>

Az egyes ágazati alapoktatások és szakirányú oktatások digitáliskompetencia-készleteinek kialakításához első lépésként a szakértők meghatározták az egyes ágazati alapoktatások és szakirányú oktatások követelményei alapján a szakmai kompetenciaterületeket. A szakmai kompetenciaterületeket a szakértők nem az adott ágazati alapoktatáshoz vagy szakirányú oktatáshoz tartozó tanulási területek vagy tantárgyak alapján, hanem az ágazati alapoktatás vagy szakirányú oktatás tevékenységi egységei, területei alapján határozták meg. A szakmai kompetenciaterületek azokat a kulcstevékenységeket jelentik, melyek az ágazati alapoktatás vagy szakirányú oktatás követelményei alapján megállapíthatóak, és azokra jellemzőek.

Egy-egy ágazati alapoktatás, illetve szakirányú oktatás esetében maximum tíz szakmai kompetenciaterület meghatározását javasoltuk a szakértőknek.

Ezt követően a szakértők az egyes ágazati alapoktatáshoz és az egyes szakirányú oktatáshoz megfogalmazott szakmai digitális tevékenységeket hozzárendelték a szakmai kompetenciaterületekhez, kidolgozva a 23 ágazati alapoktatás és a 175 szakma digitáliskompetencia-készletét.

<sup>4</sup> Pincér-vendégtéri szakember szakma projekt eredménytermék részlet

Szakmai digitális tevékenységek	Szakmai kompetenciatérség				
	Üzleti kommunikáció	Vendégteri feladatok	Rendezvényszervezés	Vendéglátó üzlet üzemeltetése	Adminisztráció
A vendéglátó üzlet digitális gépeinek működését felméri, hibáit ellenőrzi. Digitális pénztárgépet az üzlet nyitásakor megnyitja, működését folyamatosan ellenőrzi.		1	1	1	1
Az üzlet digitális választéktábláit elkészíti, feltölti a napi ajánlatokkal, kínálattal. A rendelésvételhez használt digitális eszközöket ellenőrzi. Az internetes közösségi felületein közzéteszi az aktuális ajánlatokat.		1	1	1	
Az áru érkezése során a beszállítótól kapott áruátvevő ívet készletnyilvántartó programban rögzíti, adatait kezeli, szükség esetén frissíti. Az áru érkezését követően a beszállító digitálisan megküldött számláit lementi, adatbázisában rögzíti, pénzügyekért felelős személy részére e-mail útján továbbítja. Az áruátvétel során digitális mérleggel ellenőrzi a mennyiségeket.		1	1	1	1

4. ábra: Digitáliskompetencia-készlet (részlet)<sup>5</sup>

A fejlesztés eredményeként kidolgozott digitáliskompetencia-készletek segíthetik a hatékony oktatást, értékelést, a minőségi tananyagok, taneszközök elkészítését a Szakképzés 4.0 Stratégia céljainak megfelelően.

## Irodalomjegyzék

- [1] A Szakképzés 4.0 - a szakképzés és felnőttképzés megújításának középtávú szakmapolitikai stratégiája, a szakképzési rendszer válasza a negyedik ipari forradalom kihívásaira című stratégia elfogadásáról és a végrehajtása érdekében szükséges intézkedésekről szóló 1168/2019. (III. 28.) Korm. határozat
- [2] Szakképzés 4.0 ITM, 2019. hozzáférés: 2022. 05. 30.  
[https://www.nive.hu/index.php?option=com\\_content&view=article&id=1024](https://www.nive.hu/index.php?option=com_content&view=article&id=1024)
- [3] [A szakképzésről szóló 2019. évi LXXX. törvény
- [4] A szakképzésről szóló törvény végrehajtásáról szóló 12/2020 (II.18) Korm. rendelet
- [5] [DigComp 2.1: Állampolgári digitáliskompetencia-keret /magyarul/, 2019. hozzáférés: 2022. 05. 30.  
<https://dpmk.hu/2019/07/25/a-digitalis-kompetencia-unios-referenciakerete-magyarul/>
- [6] A Digitális Kompetencia Keretrendszer fejlesztéséről és bevezetésének lépéseiről szóló 1341/2019. (VI. 11.) Korm. határozat

<sup>5</sup> Pincér-vendégteri szakember szakma projekt eredménytermék részlet

Ne csak útra csomagoljunk!  
Miért fontos a csomagolás a digitális megőrzésben?

Kalcsó Gyula  
Országos Széchényi Könyvtár Digitális Bölcsészeti Központ (Budapest)  
[kalcsso.gyula@oszk.hu](mailto:kalcsso.gyula@oszk.hu)

The Hungarian version of the Open Archival Information System (OAIS) reference model standard (MSZ 14721) has been published. Recently, more and more is discussed in Hungary about the packaging procedure recommended by the reference model for long-term digital preservation, and this paper aims to shed light on its essence and usefulness. It deals with what content is required by the OAIS exactly and with the packaging recommendations that have been developed as a result of the OAIS (e.g. the packages recommended by the E-Ark project), and why these content elements are relevant for long-term digital preservation. It examines which established practices are based on the use of other standards and recommendations (METS, PREMIS, etc.). It will show how packages can be used in different digital content management processes (both digitised and born digital preservation): their role in ingest, preservation, providing access.

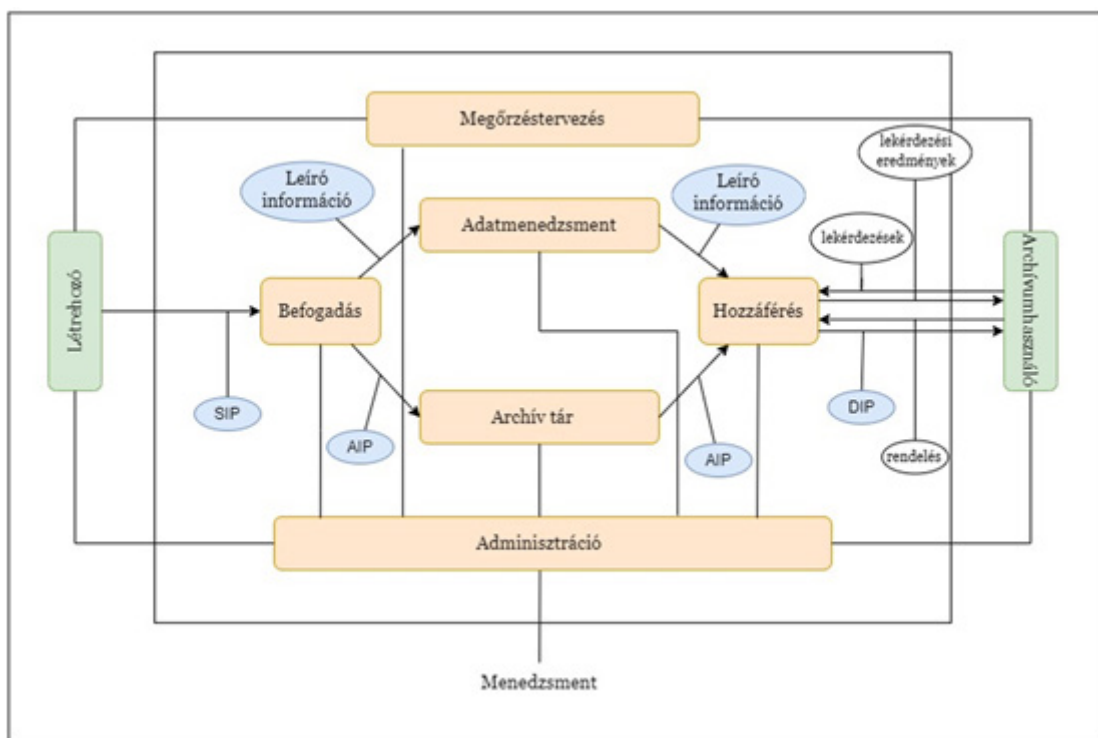
**Keywords:** long-term digital preservation, digital object packaging, OAIS, BagIt, digital repositories, digital object metadata

1. A hosszú távú digitális megőrzés a digitális objektumok egyre növekvő száma miatt szinte minden archiválással foglalkozó intézmény napi problémájává vált. A kialakult gyakorlatok legnagyobb része azonban a digitalizált objektumok megőrzésén alapul, és ezek zöme voltaképpen az egy-egy analóg objektumnak megfelelő egy-egy fájl repozitóriumbeli tárolásában ki is merül. A megőrzendő born digital objektumok számának növekedésével azonban az archívumok újabb problémákkal néztek szembe: a digitális megőrzés általánosabb problémái (a nagy mennyiség, a heterogenitás, az elavulás, az értelmezhetőség fenntartása) mellett a born digital objektumok összetettsége egészen új megközelítésmódot kíván. Ezek az objektumok sokszor nem egy-egy fájlból, hanem egymással strukturált összefüggésben álló fájlhalmazokból állnak, azaz a fájlankénti tárolásra épülő (de akár még a tárolt fájlok együttes kezelését lehetővé tevő) repozitóriumok számára is legfeljebb nehézkesen kezelhetők. Ilyen pl. egy archivált weboldal, egy e-mail a mellékleteivel, vagy egy írói hagyatékban gyűjteménybe kerülő számítógép (akár csak részleges) tartalma, pl. az író teljes elektronikus levelezése. A hosszú távú digitális megőrzés követelményei (Bánki-Kómár 2019: 139–158.) közül a born digital objektumok esetében különösen fontos a megfelelő, lehetőleg platformfüggetlen, időtálló formátum kiválasztása és az értelmezhetőség fenntartásának a biztosítása. A digitális (nem csak a born digital!) objektumok csomagban kezelése az egyik lehetséges megoldás ezekre a problémákra, mivel a nagy mennyiségű, struktúrát alkotó fájlhalmazokat egyben tudja kezelni. A csomagolásnak azonban ennél jóval sokrétűbb funkciói vannak (a digitális megőrzés fentebb említett bizonyos problémáira is megoldást jelenthetnek). Jelen cikk a digitális megőrzésben alkalmazott csomagolóeljárások lényegét próbálja összefoglalni.



2. A digitális objektumok csomagban kezelése az Open Archival Information System (OAIS) referenciamodellre, 2012 óta ISO-szabványra vezethető vissza. A modellről az utóbbi időben már magyarul is születtek publikációk (Bánki-Kómár 2019; Dancs 2016, 2017a-b; Kalcsó 2021; Szatucsek 2021), valamint megjelent a szabvány magyar változata<sup>1</sup>. Korábban a Magyar Nemzeti Levéltár az e-levéltári rendszerét<sup>2</sup> OAIS-alapokra építette<sup>3</sup>, valamint a kezdeményezésükre megjelent egy rendelet az elektronikus iratkezelésről<sup>4</sup>, amelyben az OAIS fogalmi rendszeréből a SIP-csomag (lásd lentebb) mint a beadott elektronikus iratcsomag formátuma szerepel. Ugyanakkor nem mondható el, hogy széles körben használatos lenne akár az OAIS referenciamodell, akár a csomagolási eljárás a magyar közgyűjtemények digitális megőrzési gyakorlatában (az e-levéltári iratbeszolgáltatásban használt csomagokon túl). Pedig az OAIS-csomagok a hosszú távú digitális megőrzés de facto szabványának tekinthetők, és a követelményeinek a legmesszemenőbbekig eleget tesznek.

A modell három csomagot különböztet meg: a magyar szabvány honosított kifejezéseivel élve az átadási, a megőrzési és a disszeminációs információs csomagot (az angol betűszavakkal SIP-, AIP- és DIP-csomag: Submission Information Package, Archival Information Package és Dissemination Information Package; a magyar szabvány az elterjedt betűszavakat a jobb azonosíthatóság miatt meghagyta angol eredetiben).



1. ábra: Az OAIS-referenciamodell.

Forrás: <https://prod.mszt.hu/hu-hu/szabvanyositas/hirek/2022/03/nyilt-archivumi-informacios-rendszer-oais>, elérés 2020. március 15.

- <https://ugyintezes.mszt.hu/Publications/Details/178190>, <https://prod.mszt.hu/hu-hu/szabvanyositas/hirek/2022/03/nyilt-archivumi-informacios-rendszer-oais>, elérés 2020. március 15.
- [https://mnl.gov.hu/mnl/szkk/elektronikus\\_leveltar\\_projekt](https://mnl.gov.hu/mnl/szkk/elektronikus_leveltar_projekt), elérés 2022. március 15.
- <https://mnl.gov.hu/mnl/szkk/szabvanyok>, elérés 2022. március 15.
- Magyar Közlöny, „34/2016. (XI. 30.) EMMI rendelet az elektronikus formában tárolt iratok közlevéltári átvételének eljárásrendjéről és műszaki követelményeiről”, elérés 2022. március 15., <https://magyarkozlony.hu/dokumentumok/2469dad42e27b2e67d808c7fdbea7612bbc2249b/megtekintes>.



Az OAIS egy-egy csomagja – a megőrzendő digitális objektumon túl – szabványos formátumban tárolt metaadatokat tartalmaz. A megőrzési információs csomag (AIP) pl. a tartalomra vonatkozó és a megőrzést leíró információs egységből áll. Az elsőben benne van maga a megőrzendő digitális objektum, valamint az ún. reprezentációs információi (pl. a struktúrájának a leírása, ill. a struktúrát magyarázó szemantikai információk, amelyek az objektum értelmezéséhez szükségesek), a másodikban pedig a megőrzési információk (azonosító, proveniencia, fixity<sup>5</sup>, kontextuális információk, hozzáférési jogokra vonatkozó információk). A megőrzési csomagok az átadási csomagokból állnak elő, valamint alapját képezik az igény szerint létrejövő disszeminációs csomagoknak.

Más okból és más célból jött létre a Library of Congress által kifejlesztett Baglt csomagolóeljárás<sup>6</sup>. A sokszor nagy fájlmenyiséget képviselő digitális objektumok együttes kezelése (pl. a hálózati továbbítás megkönnyítése), valamint a fájlok integritásának ellenőrizhetősége volt a legfontosabb szempont, ennek megfelelően jóval egyszerűbb csomagszerkezetet ír le. A Baglt-csomag felépítése a következő:

```

<gyökérfkönyvtár>/
|
+-- bagit.txt
|
+-- manifest-<algoritmusnév>.txt
|
+-- [címkefájlok]
|
+-- data/
|   |
|   +-- [megőrzendő/továbbítandó objektum]
|
+-- [címkekönyvtárak]/
|
+-- [címkefájlok]

```

2. ábra: A Library of Congress által ajánlott Baglt-csomagszerkezet.

Forrás: <https://datatracker.ietf.org/doc/html/rfc8493>, elérés: 2022. március 10.

A félkövérrel kiemelt elemek a kötelező tartozékai a csomagnak. A <gyökérfkönyvtár> a csomag tetszőleges nevű mappája. Ebben három elemnek kell lennie: a bagit.txt fájlban a Baglt-csomagolóeljárásra vonatkozó információ van (verziószám), a manifest-<algoritmusnév>.txt fájl tartalmazza a csomagban őrzött digitális objektum fájljainak az ellenőrzőösszegeit (a fájlnevben fel kell tüntetni a checksum számításához használt algoritmus nevét, pl. SHA-256), valamint a kötelezően data-nak nevezendő mappa, amely a megőrzendő/továbbítandó digitális objektumot tartalmazza tetszőleges struktúrában. Az eljárás megenged továbbá más címkefájlokat a gyökérfkönyvtárban, valamint további címkekönyvtárakat további címkefájlokkal (de ezek már nem kötelezőek). Mivel a Baglt ajánlás teljesen kompatibilis az OAIS-sal, ezért gyakori, hogy a digitális megőrzésben Baglt-csomagokat használnak az OAIS-ra épülő archívumokban.

5 A fájl bitszintű integritásának ellenőrzéséhez szükséges információk (pl. a fájl teljes bitfolyamát azonosító ún. ellenőrzőösszegek) összessége.

6 <https://datatracker.ietf.org/doc/html/rfc8493>, elérés: 2022. március 10.

### 3. Mire jók a csomagok?

3.1. Nagy fájlmenyiség esetén könnyebben kezelhetővé teszi a digitális objektumot, valamint a bonyolult struktúrájú, sok fájlból álló entitásokat egyben tartja. A digitális objektumok között egyre több olyan van, amely igen nagy számú fájlból és/vagy mappából álló könyvtárszerkezetet alkot. Ezek fájlankénti kezelése sokszor lehetetlen, illetőleg nem is kívánatos. A metaadatok külön fájlokban és könyvtárakban, de mégis a megőrzendő digitális objektummal együtt tárolhatók.

3.2. Mivel a csomagban minden szükséges metaadat benne van, ezért nincs szükség speciális szoftverekre az értelmezhetőséghez. A csomagolóeljárás egyik nagy előnye, hogy a becsomagolt és megfelelően metaadatosított digitális objektum felhasználhatósága nem függ adatbázistól vagy más feldolgozószoftvertől (pl. repozitóriumtól), a csomag önmagában is értelmezhető (természetesen csak akkor, ha időtálló, szabványos formátumban tároljuk a metaadatokat, l. 3.3.).

3.3. Ha a csomag minden eleme szabványos, biztosítja a hosszú távú megőrzést. A nyílt szabványok használata az értelmezhetőség fenntartásának az egyik kulcsa. Ez mind a megőrzendő objektumok, mind a metaadatok esetében elengedhetetlen. A fájlformátumokra vonatkozóan érdemes valamilyen mértékadó szakmai ajánlást követni<sup>7</sup>, a metaadatok tárolása pedig Unicode kódolású szabványos szöveges formátumokban (TXT, XML) kívánatos.

3.4. A csomag szabványos formátumban tárolt metaadatai meglévő rendszerek által is hasznosíthatók. Bár az OAIS-kompatibilis csomagok kezelésére sokféle repozitóriumszoftver áll már rendelkezésre (vö. 5.), arról sem kell lemondanunk, hogy meglévő integrált rendszerekkel összekapcsoljuk őket. Ehhez természetesen meg kell teremteni a megfelelő interfészt, de a szabványosság garantálhatja az átjárhatóságot, a metaadatok a csomagból exportálhatók, más rendszerekbe importálhatók.

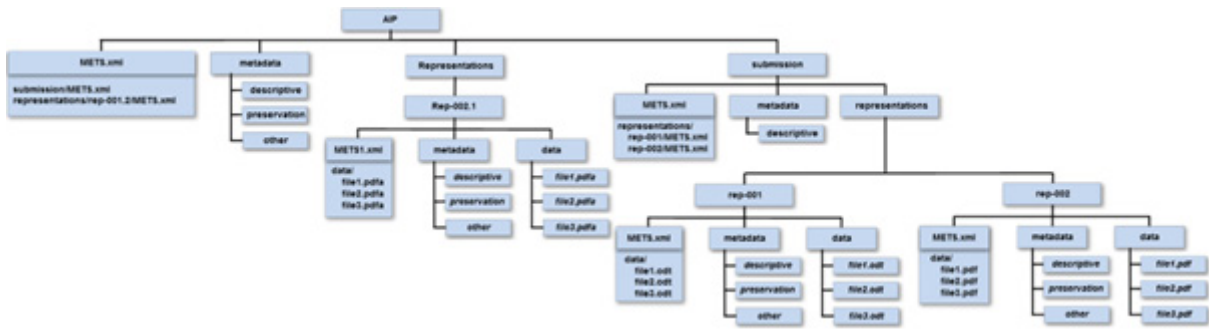
4. A csomagok tartalmának a 3. pontban említettél közelebbi meghatározása az adott digitális megőrzési feladattól függ. Általános alapelvek természetesen itt is megfogalmazhatók. A következő kérdésekre érdemes megadni a választ egy-egy csomagolással tervezett digitális megőrzési projektben: milyen metaadatokat, milyen formátumban szeretnénk a csomagba helyezni? A különböző csomagfajták (SIP, AIP, DIP) esetében milyen tartalmat, milyen struktúrában kívánunk előírni? Illusztrációképpen álljon itt három csomagszerkezeti ajánlás: egy nemzetközi (EU-s) együttműködés (E-Ark) keretében kialakított levéltári csomag, egy az általános célú digitális megőrzésben széles körben használatos szoftver (Archivematica), és egy e-mailek archiválására kidolgozott módszer (Mailbag) csomagszerkezete és -tartalma.

4.1. Az E-Ark (European Archival Records and Knowledge Preservation) projekt 2014 elején vette kezdetét, és az volt a kiindulópontja, hogy Európán belül nem alakult ki egyértelmű, széles körben elfogadott megoldás az elektronikus iratok megőrzésére, sőt az egyes európai levéltárak eltérő megközelítést alkalmaztak még ugyanolyan típusú iratok kezelésére is. A konzorciális keretek között<sup>8</sup> fejlesztett egységes eljárás- és eszközrendszer az OAIS-on alapult. Eredményeképpen ajánlás született az E-Ark-OAIS-csomagok tartalmára vonatkozóan is. Az E-Ark pl. az AIP-csomag esetében a következő felépítést és tartalmat javasolja:

---

7 Ilyen pl. a Library of Congress hosszú távú megőrzésre ajánlott formátumajánlása: <https://www.loc.gov/preservation/resources/rfs/>, elérés: 2022. március 22.

8 A projektben részt vett öt, a legmodernebb technológiákat a gyakorlatban is használó európai nemzeti levéltár (Észtország, Dánia, Norvégia, Szlovénia és Magyarország nemzeti levéltára), négy kutatóintézet, három fejlesztő cég és két kormányzati szereplő.



3. ábra: Az AIP-csomag az E-Ark ajánlásában.

Forrás: Specification for Archival Information Packages. <https://earkaip.dilcis.eu/>, elérés: 2022. március 10.

Látható, hogy az E-Ark AIP-jában megtalálható a teljes átadási csomag (submission), valamint a megőrzendő digitális objektumokat ún. reprezentációkban kezeli. Elsősorban a strukturális metaadatok tárolása miatt előírja a METS használatát<sup>9</sup>, a leíró, a megőrzési és egyéb metaadatoknak pedig külön könyvtárat tart fenn.

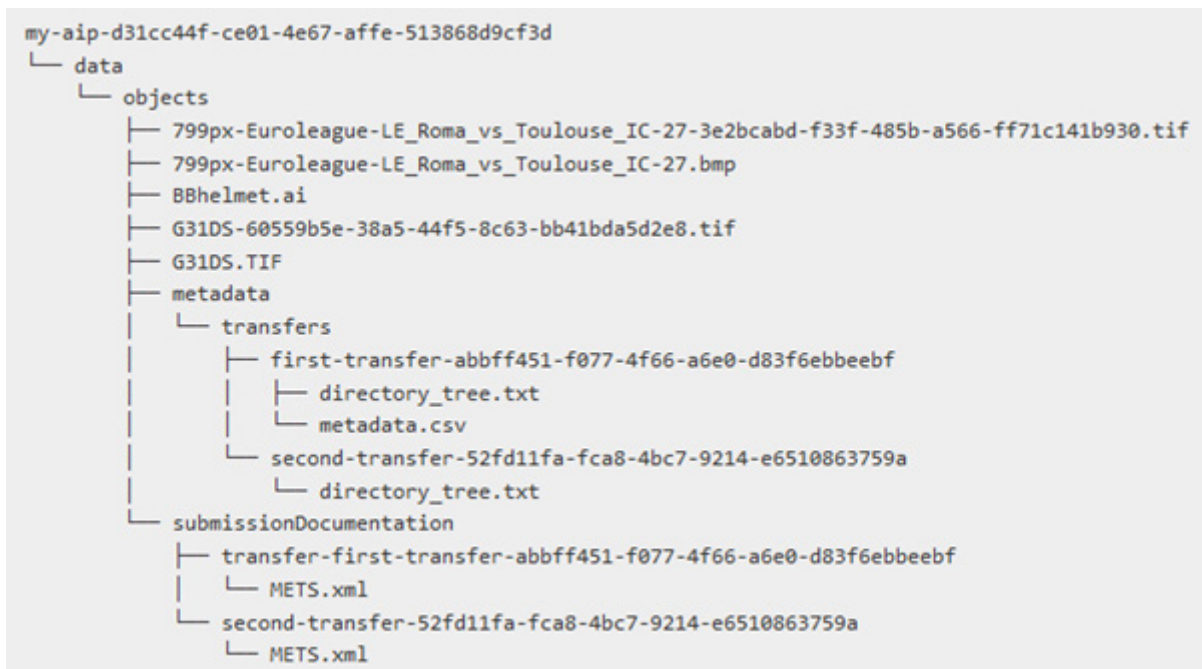
4.2. Az Archivematica a kanadai Artefactual cég által fejlesztett nyílt forráskódú szoftvereszközök integrált csomagja, amely lehetővé teszi a felhasználók számára, hogy a digitális objektumokat az OAIS-nak megfelelően dolgozzák fel a befogadástól a hozzáférésig. A metaadat- és egyéb szabványok közül támogatja a METS, a PREMIS<sup>10</sup>, a Dublin Core<sup>11</sup>, a Bagit használatát. Az Archivematica AIP-csomagja az alábbi felépítésű:

```
[1] my-aip-d31cc44f-ce01-4e67-affe-513868d9cf3d
[2] |─ bag-info.txt
[3] |─ bagit.txt
[4] |─ manifest-sha512.txt
[5] |─ tagmanifest-md5.txt
[6] |─ data
[7]   |─ logs
[8]   |─ objects
[9]   |─ thumbnails
[10] |─ METS.d31cc44f-ce01-4e67-affe-513868d9cf3d.xml
[11] |─ README.html
```

9 <https://www.loc.gov/standards/mets/>, elérés: 2022. március 13.

10 <https://www.loc.gov/standards/premis/>, elérés: 2022. március 13.

11 <https://www.dublincore.org/specifications/dublin-core/>, elérés: 2022. március 13.



4. ábra: Az Archivematica szoftver AIP-csomagja, valamint a data/objects mappa belső felépítése. Forrás: <https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/archival-storage/aip-structure/>, elérés: 2022. március 10.

Az Archivematica AIP-csomagja egy Baglt-csomag, amelynek data mappájában nem csak az archiválandó digitális objektum található, hanem egyéb komponensek is, pl. egy README.html fájl, amely az OAIS-modell reprezentációs információt leíró elemének felel meg. Az objects mappában található a megőrzendő objektum, valamint a hosszú távú megőrzésre és a szolgáltatásra konvertált reprezentációi és további metaadatokat tartalmazó mappák. A strukturális metaadatok METS-fájlokban vannak.

4.3. A Mailbag<sup>12</sup> a New York-i állami egyetem (University at Albany) M. E. Grenander Department of Special Collections & Archives (M. E. Grenander Speciális Gyűjtemények és Archívumok Osztálya) kezdeményezése az e-mailek hosszú távú megőrzésének Baglten alapuló szabályozására. A projekt 2021 márciusában indult. A specifikáció<sup>13</sup> részletesen leírja, hogy a Bagltnak megfelelő elektronikuslevél-archiváló csomagnak mit érdemes tartalmaznia. Az ajánláshoz szoftvert is fejlesztenek<sup>14</sup>, amely alkalmas lesz a Mailbag-csomagok létrehozására. Az alábbi általános csomagszerkezetet javasolják:

<sup>12</sup> <https://archives.albany.edu/mailbag/>, elérés: 2022. március 10.

<sup>13</sup> <https://archives.albany.edu/mailbag/spec/>, elérés: 2022. március 10.

<sup>14</sup> <https://github.com/UAlbanyArchives/mailbag>, elérés: 2022. március 10.



```

|gyökérvkönyvtár>/
|
|-- bagit.txt
|-- bag-info.txt
|-- mailbag.csv
|-- manifest-<algoritmusnév>.txt
|-- tagmanifest-<algoritmusnév>.txt
|
|-- data/
|   |
|   |-- mbox/
|   |   |-- [megörzendő objektum]
|   |-- pdf/
|   |   |-- [megörzendő objektum]
|   |-- warc/
|   |   |-- [megörzendő objektum]
|   |-- attachments/
|   |   |-- [Mailbag-Message-ID]/
|   |       |-- [megörzendő objektum]
|   |   |-- [Mailbag-Message-ID]/
|   |       |-- [megörzendő objektum]
|   ...

```

5. ábra: A Mailbag ajánlás által előírt csomagszerkezet. Forrás: Mailbag Specification 0.3 (Release Candidate).

<https://docs.google.com/document/d/1BZHkIc6MKktXJBPcvFvLxLRoX8ICidFemflppqpUQ7s>,

elérés: 2022. március 10.

A Mailbag a Bagit-csomagok kötelező elemein túl előírja a bag-info.txt címkefájl használatát előre definiált metaadatmezőkkel. Szerepel még egy mailbag.csv fájl is, amelyben ugyancsak előre definiált, az e-mailekre vonatkozó metaadatok vannak. A data mappa is speciális felépítésű: a Mailbag legalább egyféle formátum szerepeltetését írja elő az alábbiak közül: MBOX, PST, MSG, EML, PDF, WARC egy ennek megfelelően elnevezett mappában. Amennyiben bármelyik megörzendő e-mailnek van legalább egy csatolmánya, kötelező az attachments mappa használata is, amelyben egyedi azonosítóval elnevezett mappában (a mailbag.csv-ben megadott Mailbag-Message-ID) található a csatolmány.

5. Hogyan illeszthetők be a csomagok a digitális tartalmak menedzsmentjébe? A teljes digitális gyűjteményi életciklus folyamatait segíthetik a gondosan megtervezett, szabványos formátumokat használó csomagok. Csak három fő mozzanatra térünk ki: az átadásra-befogadásra, az archiválásra és a hozzáférés biztosítására.

5.1. A gyűjteményi átadást-befogadást (ingest) jelentősen megkönnyíti, ha a folyamathoz részletesen kidolgozott csomagformátum áll rendelkezésre. A Magyar Nemzeti Levéltár által kidolgozott iratátadási-átvételi workflow (amely minisztériumi rendelet formájában is manifesztálódott, l. fentebb) pl. meghatározza az átadási (SIP-) csomag pontos tartalmi jellemzőit az előírt formátumokkal együtt (l. Szatucsek 2021). Ily módon biztosítható, hogy az archívum a megfelelően előkészített befogadó eszköze gyűjteményi segítségével automatizáltan, dokumentáltan, ellenőrzött módon, nagy mennyiségben tudjon átvenni megőrzésre szánt digitális objektumokat. A csomagban tárolt ellenőrzőösszegek biztosítják az adatok (fájlok) integritását, a szabványos formátumokban tárolt metaadatok a gyűjteményi rendszerekbe való automatikus feltöltést. Az átadás körülményeire vonatkozó dokumentáció is helyet kaphat a SIP-csomagban, ami megkönnyíti a későbbi ellenőrzést, hibakeresést, -javítást.



5.2. A befogadási folyamat végén a SIP-csomagból előáll az AIP-csomag<sup>15</sup>. Ennek pontos tartalmát, formátumait ugyancsak szabályozni kell. A legfontosabb célok (az időtállóság és az értelmezhetőség fenntartása) határozzák meg a döntéseket. Az archivált csomagok nem tekinthetők véglegesen lezártak, az archívum állandó feladata az említett céloknak való megfelelés fenntartása, azaz szükség esetén be kell avatkozni, módosítani kell az AIP-csomag tartalmát (pl. az elavult megőrzési formátumokat le kell cserélnie). Az AIP-csomag képezi a disszeminációs csomag alapját is, annak tartalma ebből áll elő. Minden műveletet (az AIP-csomag létrejöttétől annak bármilyen módosításán keresztül a belőle készült DIP-csomagok összeállításáig) szabványos formában dokumentálni kell<sup>16</sup>.

5.3. Az ún. disszeminációs csomagok, amelyek a szolgáltatáshoz szükségesek, ugyancsak az AIP-csomagokból állnak elő. A csomag tartalma az aktuális felhasználás céljaitól függ. Lehetnek előre összeállított DIP-csomagok, de készülhet ad hoc, egy konkrét igény felmerülése esetén összeállított csomag is. A szolgáltatás céljának és körülményeinek a dokumentációja is bekerülhet a DIP-be (pl. egy megrendelt szolgáltatás esetén akár a felhasználási szerződés stb.).

## 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](https://ommik.hu/media/attachments/2019/12/09/fehr_knyv.pdf).
- Dancs Szabolcs. 2016. „A hosszú távú megőrzés szabványos útja(és a Rosetta rendszer példája)”. *Tudományos és Műszaki Tájékoztató* 63 (1): 27–37.
- Dancs Szabolcs. 2017a. „A hiteles digitális másolatról és a másolati példányok bibliográfiai kezeléséről – 1. rész: Hitelesség szabványos alapokon”. *Könyv, Könyvtár, Könyvtáros* 26 (10): 7–15.
- Dancs Szabolcs. 2017b. „A hiteles digitális másolatról és a másolati példányok bibliográfiai kezeléséről – 2. rész: Metaadatok”. *Könyv, Könyvtár, Könyvtáros* 26 (12): 7–14.
- Kalcsó Gyula. 2021. „Born digital workflow tervezése a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjában”. In *Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6–9. Eötvös Loránd Tudományegyetem*, 195–203. HUNGARNET Egyesület. <https://doi.org/10.31915/NWS.2021.20>.
- Langley, Somaya. “Digital Preservation Should Be More Holistic: A Digital Stewardship Approach”. In Myntti, Jeremy. American Library Association, *Digital Preservation in Libraries: Preparing for a Sustainable Future*. Apollo – University of Cambridge Repository, 2018. <https://doi.org/10.17863/CAM.34317>.
- Szatucsek Zoltán. 2021. *Az elektronikus iratok kezelése*. Budapest: Nemzeti Közszerzői Egyetem Közigazgatási Továbbképzési Intézet. <https://kti.uni-nke.hu/document/vtkk-uni-nke-hu/Az%20elektronikus%20iratok%20kezel%C3%A9se.pdf>.

---

15 Előfordulnak olyan esetek, amikor egy SIP-csomagból több AIP-csomag jön létre, az egyszerűség kedvéért azonban maradunk az egy az egyhez megfelelésnél.

16 Erre a digitális archívumokban, repozitóriumokban széles körben használt PREMIS (l. fentebb) a legalkalmasabb.

## A CIRCLE felhő elmúlt évtizede

Karsa Zoltán István, Szeberényi Imre  
BME Irányítástechnika és Informatika Tanszék  
[{karsa|szebi}@cloud.bme.hu](mailto:{karsa|szebi}@cloud.bme.hu)

### Abstract

The last decade of the CIRCLE cloud

Nowadays, we can confidently say cloud computing belongs to the group of mature, but rapidly evolving technologies. However, there is no widely used cloud manager specifically designed for education and science used in the academic domain. In this article, we would like to summarise the experience gained during the development of the CIRCLE system, developed at the Budapest University of Technology: what changes have been induced by the university environment from the first version to the present day, what lessons can be learned. Henceforth, what future changes do we see necessary to further develop the system.

**Keywords:** virtualization, cloud, Infrastructure-as-a-service, education, libvirt, python, Django

### Kivonat

Ma már bátran kijelenthetjük, hogy a számítási felhők a kiforrott, de mégis gyorsan fejlődő technológiák csoportjába tartozik. Nincs azonban olyan széles körben használt felhőkezelő, amelyet kifejezetten oktatási és tudományos területek számára fejlesztettek ki. Cikkünkben a Budapesti Műszaki Egyetemen kifejlesztett CIRCLE rendszer tapasztalatait szeretnénk bemutatni: milyen változásokat indukált az első verziótól a mai napig az egyetemi környezet, milyen tanulságokat tudunk levonni. Továbbá milyen jövőbeni terveket látunk szükségesnek a rendszer továbbfejlesztésekor.

**Kulcsszavak:** virtualizáció, felhő, infrastruktúra mint szolgáltatás, oktatás, libvirt, python, Django

## 1. Bevezetés

A gyakran változó szoftverspecifikációk, illetve a takarékoság előtérbe helyezték a felhő alapú erőforrás elosztást az oktatási környezetekben is [1]. Bár a cikk elsősorban az IaaS szolgáltatásimodellre követő rendszerekről szól, a többi modellhez tartozó szolgáltatások hamar betörték és nagy népszerűségnek örvendek mind a mai napig. Gondoljunk csak a Google által nyújtott elsősorban SaaS<sup>1</sup> környezetekre: Google Docs, Slides stb. Nemcsak oktatói nyomásra, de a hallgatók is a könnyű elérhetőségnek és a testre szabható megosztási beállításokkal hamar rátértek ezek használatára.

A CIRCLE [2] felhőmenedzser elődjét 2012-ben egy hallgatói csapat kezdte el fejleszteni, miután más rendszerek akkor még gyerekcipőben jártak: az első próbálkozások egy OpenNebula [3] alapú rendszerrel kezdődtek. Az OpenNebula bár tartalmazott felhasználói felületet az adminisztráláshoz, de elsősorban informatikában jártas oktatók tudták azt hatékonyan használni. Ezért a hallgatói csapat egy Django<sup>2</sup> alapú webes portált fejlesztett, ami egyszerű felhasználói felületen elérhetővé tette a különböző virtuális gépek létrehozását,

1 Szoftver, mint szolgáltatás (Software as a Service)

2 Python alapú web fejlesztői keretrendszer: <https://www.djangoproject.com/start/overview/>

indítását, megszüntetését, felhasználók jogosultságkezelését, stb. (IK-Cloud). Az OpenNebula akkori gyerekbetegségei, illetve egy átgondolt API<sup>3</sup> hiánya miatt később egy saját menedzser fejlesztésébe kezdett a csapat<sup>4</sup>. Felmerült ugyan az OpenStack Grizzly [4] használata is, de az első telepítési nehézségek miatt akkor az sem tűnt biztos iránynak, ezért ezt elvetettük. Mai szemmel lehet, hogy ez nem a legjobb döntés volt. Mindenesetre a szinte nulláról megtervezett és felépített opensource rendszer sok szép kihívást és jelentős eredményeket hozott.

## 2. CIRCLE rendszer fejlesztése

A 2014-re éles üzemben működő rendszert kb. 2018-ig fejlesztették, bővítették intenzíven. A cikk írásáig (2022) körülbelül 20-25 hallgató vett részt a rendszer fejlesztésében önálló laboratóriumi projekt, szakdolgozat és/vagy diplomaterv formájában. Az első éles rendszer a BME IIT és BME IK oktatási feladatait segítette, illetve backoffice erőforrásokat biztosított a két szervezet számára, de Győrben Miskolcon és még Glasgow-ban is üzemelt egy-egy kísérleti rendszer. 2018-ban a rendszert kiszolgáló hardver infrastruktúra bővítése lehetővé tette, hogy a Villamosmérnöki és Informatikai Kar (VIK) összes tanszéke és hallgatója használhassa a rendszert. A tervezés és fejlesztés során elsődleges szempont volt az oktatói és kutatói környezet támogatása, és egy olyan felület elkészítése, amin informatikai ismeretekkel nem rendelkező felhasználók is könnyen eligazodnak.

Az elmúlt évtized során nagyon sok bővítés készült el, ilyen például az OCCI<sup>5</sup> interfész támogatása, VXLAN<sup>6</sup> modul, valamint az Microsoft Azure vagy OpenStack alapú VM-kezelés. Ezek nagy része sajnos csak szakdolgozat/diplomaterv szintjén készült el, éles üzembe nem kerültek. A legígéretesebb testvérprojektet pedig a Covid szétverte. Ennek keretében a teljes rendszert újraterveztük úgy, hogy a menedzser rész teljesen leválasztható legyen a hagyományos IaaS rétegről, ugyanakkor biztosítson minden kényelmi szolgáltatást, amit a CIRCLE rendszer nyújt. Ezzel lehetővé vált volna, hogy legalul egy OpenStack, vagy Azure fusson. A teljesen üzemkész, biztos változat sajnos már nem készült el. [5].

## 3. CIRCLE Funkciók

A tervezésnél a legfontosabb cél az volt, hogy megvalósuljanak az alapvető IaaS<sup>7</sup> funkciók (a virtuális gépek készítése, indítása, leállítása, törlése, stb.), valamint az üzemeltetéshez szükséges adminisztrátori feladatokat támogató funkciók (fizikai gépek felvétele, törlése, VM<sup>8</sup>-ek migrálása, monitorozása, stb.). Ráadásul ezeket egy egységes letisztult grafikus felületről lehessen elérni.

A rendszerben új virtuális gépet többnyire az előre elkészített sablonokból lehet létrehozni. A sablonok szerepe hasonló az OpenStack flavor [6] szerepéhez. Sablonokat a CIRCLE rendszerben nem csak adminisztrátor tud létrehozni, ezen felül a virtuális gépek erőforrásai (RAM, vCPU, diszk) a sablontól függetlenül bármikor változtathatók a felhasználók kérésére.

---

3 Application Programming Interface

4 A hallgatói csapat oszlopos tagjai: Bach Dániel, Dudás Ádám, Guba Sándor, Kálmán Viktor, Óry Máté voltak. Konzulensük Szeberényi Imre volt.

5 Open Cloud Computing Interface ([https://en.wikipedia.org/wiki/Open\\_Cloud\\_Computing\\_Interface](https://en.wikipedia.org/wiki/Open_Cloud_Computing_Interface))

6 Virtual Extensible LAN ([https://en.wikipedia.org/wiki/Virtual\\_Extensible\\_LAN](https://en.wikipedia.org/wiki/Virtual_Extensible_LAN))

7 Infrastruktúra, mint szolgáltatás (Infrastructure as a Service)

8 Virtuális gép (Virtual Machine)

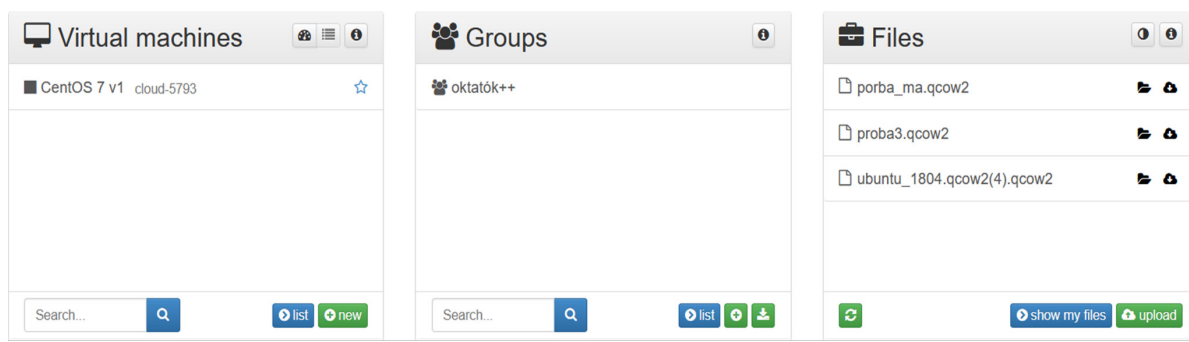


Egy sablon egy már elkészült virtuális gép pillanatképét tárolja. Ebből a képből pedig más virtuális gépek indíthatók, a kiinduló virtuális gépre előkészített szoftverkönyezettel és konfigurációval. Így az oktatók a speciális szoftvereket könnyen meg tudják osztani a hallgatókkal, illetve órákon használt különböző szoftvercsomagok telepítését (ami akár félévről félévre változhat) nem kell elvégezni a fizikai gépen, így csökken a rendszergazdák, operátorok terhelése is.

A virtuális gépek külső elérését IPv4 felett portforward technikával<sup>9</sup> oldjuk meg korlátozott számú IP címek miatt. IPv6 felett viszont minden gépnek egyedi címe van. Az elérést tűzfalszabályok szabályozzák, melyet a felhasználók a **kontrollpanelről** pár kattintással tudnak konfigurálni. A rendszerüzemeltetői feladatokat szintén grafikus felületű kontrollpanel segíti, amelyről a hálózati és tárolási funkciókhoz tartozó összes feladat elvégezhető (VLAN, DNS, FW, Jogosultságok, stb).

A felhasználók jogaiknak kezelése alapvetően ACL<sup>10</sup> alapú. A felhasználókhöz, illetve azok csoportjaihoz három különböző hozzáférési jogosultság társítható: tulajdonos, operátor és felhasználó. Mivel elsősorban oktatási környezethez készült rendszerről van szó, a felhasználók a használat után nem fizetnek, mint például egy üzleti felhős szolgáltatásnál. Ugyanakkor szükséges valamilyen módon rávenni a felhasználókat, hogy indokolatlanul ne használják az erőforrásokat, hiszen az erőforrások sajnos nem végtelenek.

Ezért egy ún. életciklus vagy bérleti modell alapján kezeljük a virtuális gépeket. Ez azt jelenti, hogy minden virtuális gépnek van egy felfüggesztési (alvó állapot) és egy törlési ideje. Ezt alapvetően a sablon tulajdonosa (pl. oktató) állítja be, de a felhasználó (pl. hallgató) korlátozás nélkül megújíthatja. A 2x45 perces egyetemi laborokhoz használt tipikus életciklus a 2 óra + 1 hét, ami azt jelenti, hogy a gép felfüggesztésre kerül 2 óra múlva, ill. törlésre kerül 1 hét múlva, ha a felhasználó időközben nem újítja azt meg. Így a hallgató 1 héten belül bármikor ellenőrizheti a laborban végzett munkáját, de korlátlanul meg is újíthatja a bérletet. Ahogy az üzleti célú szolgáltatásoknál is, úgy a CIRCLE esetén is a felhasznált memória, vCPU<sup>11</sup>, háttértár, egyszerre indítható gépek száma is korlátozható akár felhasználónként is.



1. ábra: Áttekintő kép

A CIRCLE rendszer autentikációs modulja az egyetemi SSO<sup>12</sup> (eduID) rendszert használja, ami extra adatokat is képes szolgáltatni a tanulmányi rendszerből. Így a felhasználó sikeres azonosítása után a rendszer képes megkülönböztetni az oktatókat és a hallgatókat ill. olyan

<sup>9</sup> Lehetővé teszi, hogy azonos IP cím mögött több belső hálózaton levő gép is elérhető legyen.

<sup>10</sup> Access-control List

<sup>11</sup> Virtual CPU

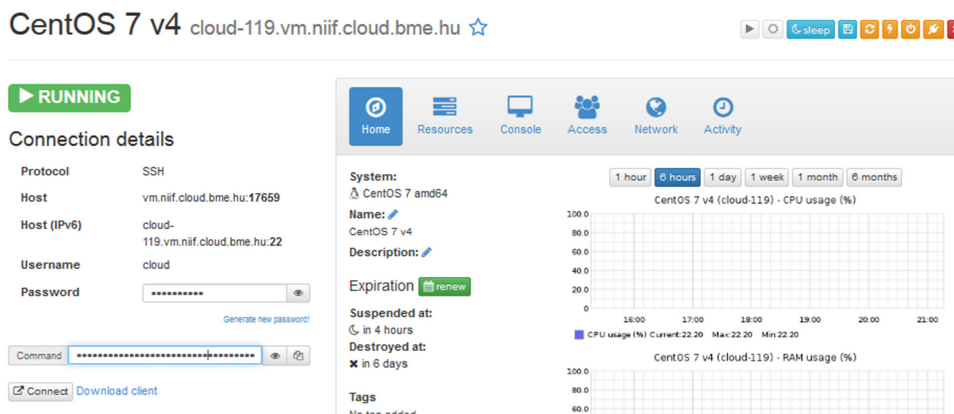
<sup>12</sup> SSO: Single Sign On, eduID: <https://eduid.hu/hu>

adat is rendelkezésre áll, hogy az adott hallgató milyen tantárgyakat hallgat az adott félévben. Ennek ismeretében egy adott tárgy oktatója nagyon egyszerűen be tudja állítani, hogy az általa létrehozott sablont azok használhassák, akik a tantárgyat **hallgatják**. Nagy létszámú (4-500 fő) alaptárgyak esetében ez a beállítás hallgatónként munkaigényes lenne.

#### 4. Felhasználói felület

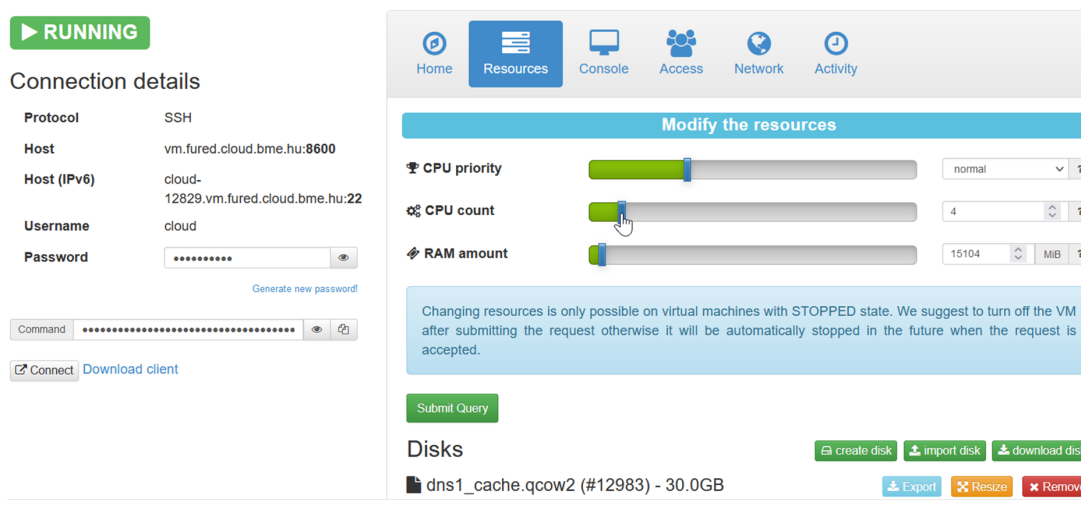
Azonosítás után a felhasználót egy áttekintő kép várja (1. ábra), melyen megjelennek az adott felhasználó virtuális gépei, valamint a perzisztens tároló tartalma. Oktató számára az általa kezelt csoportok és sablonok is megjelennek. Ezen a felületen lehet új gépet indítani, vagy egy konkrét virtuális gép részletet megtekinteni

A részletes információk között látható a gép állapota (2. ábra), erőforrásai, életciklusa, hozzáférési adatok, valamint a géppel kapcsolatos tevékenységek (altatás, leállítás, törlés, stb.) kezelőgombjai. A részletes információkat adó képről lehet elérni a virtuális gép konzolját is.



2. ábra: Virtuális gép részletei

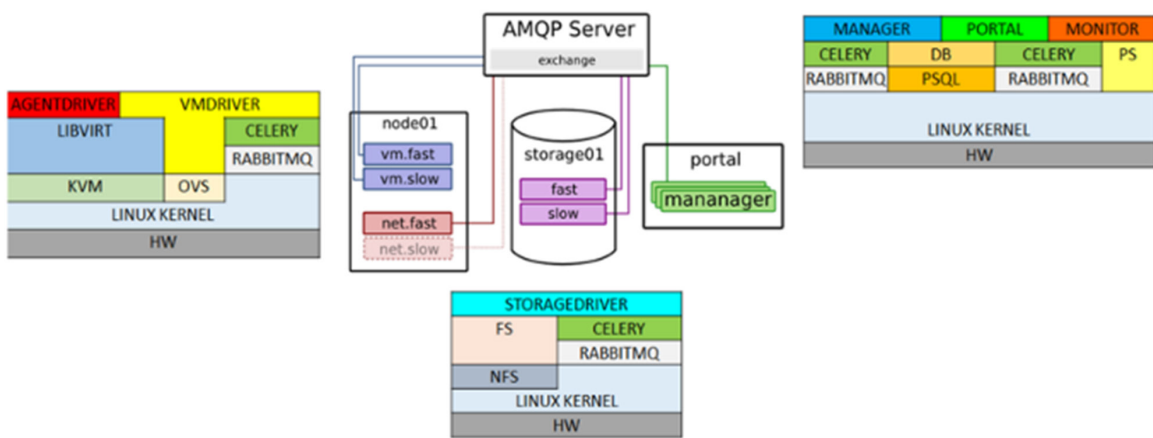
Az erőforrások módosítására vonatkozó kérvény az erőforrásinformáció fülön kapott helyet (3. ábra). A kérvényt kitöltéséről az üzemeltetés azonnal értesítést kap. Amennyiben a kérés jóváhagyásra kerül, az abban kért erőforrások automatikusan lefoglalódnak és hozzárendelődnek az adott virtuális géphez.



3. ábra: Erőforrások szerkesztése

## 5. Felépítés

A rendszer főbb komponenseit és kapcsolatait a 4. ábra mutatja be. A főbb modulok Python nyelven készültek, kivéve néhány opensource komponenst, mint pl. az AMQP<sup>13</sup> szerver. A webes felhasználói felületet a Django keretrendszer, a virtualizációt pedig a QEMU/KVM<sup>14</sup> hypervisor biztosítja, amit libvirt<sup>15</sup> API-n keresztül használunk. A különböző modulok irányítására és vezérlésére üzenet-alapú kommunikációval (AMQP) történik, melyet a RabbitMQ szerver továbbít és oszt el. Az üzeneteket a fogadó oldalon Celery<sup>16</sup> taszkok dolgozzák fel. Ezek az üzenet típusától függően végzik a feladatukat pl. virtuális gépet indítanak, migrálnak, megszüntetnek, stb. A virtuális gépek alapszintű konfigurációját (hálózati címek beállítása, jelszó és kulcs beállítás, stb.) egyedi megoldás, az ún. agent program végzi, ami alapfeladata mellett a felhasználónak érkező rendszerüzeneteket is fogadja és továbbítja az adott virtuális gépen éppen aktív felhasználó felé.



4. ábra: CIRCLE rendszer felépítése

A rendszerben létrejövő virtuális hálózatot Open vSwitch<sup>17</sup> segítségével építjük fel, melynek konfigurációját folyamatosan frissíti a menedzsermodul, ahogyan a virtuális gépek létrejönnek, mozognak ill. megszűnnek. Fontos kiemelni, hogy a virtuális gépek között egyedi VLAN-ok is kialakíthatók hallgatói mérések számára, vagy éppen egy virtuális klaszter számára.

A virtuális gépek és a felhasználók lokális gépei közötti fájlcsere érdekében a remote desktop adta lehetőségek segítik, illetve minden felhasználónak rendelkezésére áll egy korlátozott méretű perzisztens tároló, ami egyszerűen csatolható a felhasználó virtuális gépéhez, illetve drag-and-drop módszerrel kezelhető felhasználó lokális gépéről. [7]

## 6. Tapasztalatok, felhasználás

Jelenleg a VIK hallgatói és oktatói számára elérhető CIRCLE felhőszolgáltatást 4 adatközpont szolgálja ki, melyekben eltérő hardver konfigurációk vannak, de eltérő koruk miatt nagyjából azonos teljesítményűek. A 4 adatközpont közül 3 a BME telephelyein működik, a negyedik pedig a KIFÜ-NIIF-től bérelt infrastruktúrában.

13 Advanced Message Queuing Protocol ([https://en.wikipedia.org/wiki/Advanced\\_Message\\_Queueing\\_Protocol](https://en.wikipedia.org/wiki/Advanced_Message_Queueing_Protocol))

14 QEMU: Nyílt forráskódú emulátor (<https://en.wikipedia.org/wiki/QEMU>), KVM: Kernelbe épített virtualizáció ([https://en.wikipedia.org/wiki/Kernel-based\\_Virtual\\_Machine](https://en.wikipedia.org/wiki/Kernel-based_Virtual_Machine))

15 Virtualizációs API (<https://libvirt.org/>)

16 Elosztott üzenet-alapú feladatvégrehajtó (<https://docs.celeryq.dev/en/stable/>)

17 Nyílt forráskódú virtuális hálózati kapcsoló (<https://www.openvswitch.org/>)



A CIRCLE felhőmenedzsert elsősorban oktatási céllal használjuk. Számos hallgatói laboratórium és vizsga háttérinfrastruktúráját adja. Különösen nagy jelentőséget kapott Covid-19 által kiváltott távoktatási időszakban. Az oktatási feladatok mellett kutatás feladatokra is eredményesen használjuk. Ha speciális szoftverkörnyezetre van szükség, vagy egy új szoftver kipróbálására, akkor ma már a legtöbb kutatónak a karon a CIRCLE felhő jut eszébe. Továbbá olyan termekben is tudunk erőforrásigényesebb laborokat tartani, ahol csak vékonykliens érhető el.

A nagy számításigényű kutatási feladatok támogatásához kialakítottunk egy virtuális HPC klasztert [8] a felhőben. A klaszter jelenleg 10 virtuális munkagépből (worker node) és egy fejgéből áll. A munkagépek mindegyike 10vCPU-t és 40GB RAM-ot tartalmaz. 5 gép pedig dedikált GPGPU<sup>18</sup> (Nvidia TESLA V100) kártyával is rendelkezik. A klaszteren Slurm [9] ütemezőt használunk a feladatok ütemezésére és az erőforrások lefoglalására. A Sarus [10] rendszer segítségével pedig docker konténereket is tudunk fogadni.

HOME Token Admin karsa Logout

## Server Options

Választott konfiguráció: [node] [-gpu/mps-] [RAM] [időkeret]

Beállítási lehetőségek:

- node:
  - debug (csak demo célokra)
  - htp (csak CPU intenzív taszkokra)
  - gpu (Nvidia Tesla kártyával)
- GPU mód (csak gpu-s node):
  - gpu: a teljes gpu-s kártya lefoglalásra kerül, így más nem is használhatja
  - mps: egy kártyát több felhasználó is megkaphat

Select a job profile:

GPU gpu 20GB 10vCPU 2h

Start

5. ábra: JupyterHub felület jupyter allokáció

Elsősorban a mesterséges intelligencia és Big-data<sup>19</sup> kutatásokhoz Jupyter notebook<sup>20</sup> környezetet is biztosítunk, ami meglehetősen kedvelt a gyors próbákhoz. A rendszerhez elérhető egy olyan JupyterHub-os webes interfész is (5. ábra), ami együttműködik a SLURM ütemezővel, segítségével egyszerűen egy webböngészőből tudunk erőforrást allokálni, és a feladatokat lefuttatni a klaszteren. Ennek a funkciónak a bevezetését azért tartottuk fontosnak, mert nem minden felhasználó van hozzászokva a parancssoros felülethez.

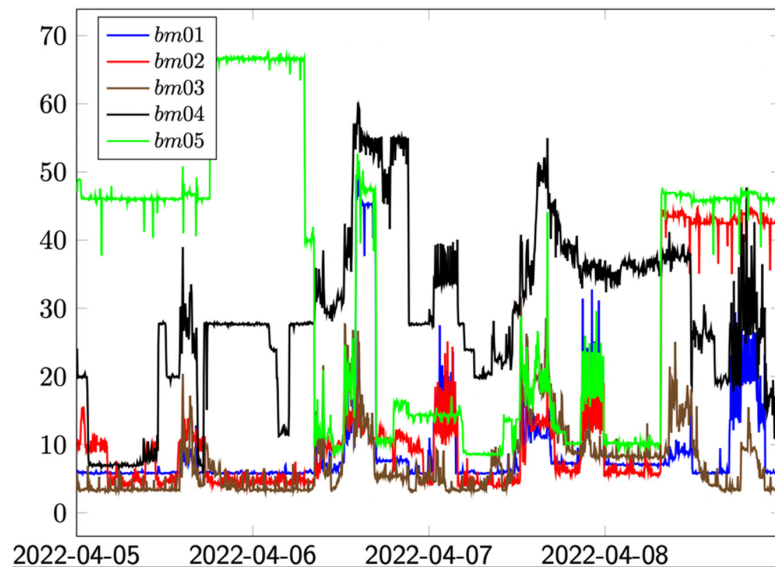
A 6. ábra a CPU kihasználtságot mutatja gépenkénti bontásban a balatonfüredi adatközpontban (5 fizikai gép, 2TB RAM, 200 CPU). Megfigyelhető a szorgalmi időszak napi lüktetése. Látható, hogy a rendszer terhelése nem kiegyensúlyozott, ez az oktatási környezet miatt van: ha egy felhős labor kezdődik, akkor a terhelés hirtelen megugrik, ha

18 General-purpose computing on graphics processing units

19 Nagyméretű adathalmazok komplex feldolgozását igénylő feladatok és technológiák gyűjtőneve ([https://hu.wikipedia.org/wiki/Big\\_data](https://hu.wikipedia.org/wiki/Big_data))

20 Webes interaktív programozói felület különböző programozási nyelvekhez (pl. Python, C++, Java, ...) (<https://jupyter.org/>)

vége akkor pedig értelemszerűen csökken. Az is megfigyelhető, hogy egy-egy fizikai gépen néha majdnem 2-szer annyi virtuális gép fut, mint amennyi CPU fizikailag rendelkezésre áll (40 CPU mag van egy fizikai gépben).



6. ábra: CPU kihasználtság (Füred, április 4–8.)

Adatközpontjainkban megfigyeltük, hogy az aktív virtuális gépek maximális száma 100–120, attól függően milyen operációs rendszert használunk. Azt találtuk, hogy a korlát elsősorban a központi tároló IO<sup>21</sup> kapacitásával van összefüggésben, ezért hasonló rendszerek tervezésénél erre külön figyelmet kell fordítani.

## 7. Tervek

Jelenleg folyamatban van a Python3-as verzióra való átállás (CIRCLE3), mivel a 2-es verzió támogatása megszűnt. Az átállás sajnos nem egyszerű, mivel több, korábban támogatott csomag, modul már nem elérhető, így vagy saját kezűleg kell ezeket továbbfejleszteni, vagy alternatív megoldásokat kell keresni

Egy igazán teljesítőképes felhő rendszer egyik fontos paramétere a fájlrendszer IO áteresztőképessége. A produktív környezetekben végzett mérések alapján a központi tároló és az NFS<sup>22</sup> jelenti a szűk keresztmetszetet. Mivel a hardverek cseréje már kevésbé éri meg, nincs összhangban az ár/érték arány, ezért szoftveres úton NFS helyett GlusterFS<sup>23</sup> vagy Ceph<sup>24</sup> alapú tárolást szeretnénk bevezetni.

Szükségesnek látjuk a már „szabványosnak” minősülő cloud-init mechanizmus megvalósítását, amivel egyszerűbbé válik új operációs rendszer változatok telepítése. A manapság egyre népszerűbb big-data témakörben is szeretnénk, elsősorban a tárkezelő modult bővíteni, hogy hatékonyabban tegye lehetővé a kiszolgálást.

<sup>21</sup> Input/output

<sup>22</sup> Network File System ([https://hu.wikipedia.org/wiki/Network\\_File\\_System](https://hu.wikipedia.org/wiki/Network_File_System))

<sup>23</sup> Nyílt forráskódú elosztott fájlrendszer (<https://www.gluster.org/>)

<sup>24</sup> Nyílt forráskódú elosztott fájlrendszer (<https://ceph.io/en/>)

## 8. Irodalomjegyzék

- [1] GUPTA, Awaneesh, et al., „Role of cloud computing in management and education,” in *Materials Today: Proceedings*, 2021. Oldalak: 1, 3-4.  
<https://doi.org/10.1016/j.matpr.2021.07.370>
- [2] BME IK, „CIRCLE Cloud honlap,” Budapesti Műszaki és Gazdaságtudományi Egyetem Informatikai Központ, 2022. [Online]. Available: <https://circlecloud.org/>. [Hozzáférés dátuma: 24 augusztus 2022].
- [3] OpenNebula, „OpenNebula: The Open Source Cloud & Edge Computing Platform for the Enterprise,” [Online]. Available: <https://opennebula.io/>. [Hozzáférés dátuma: 24 augusztus 2022].
- [4] Wikipedia, „OpenStack,” 2010. [Online]. Available: <https://www.openstack.org/>. [Hozzáférés dátuma: 24 augusztus 2022].
- [5] C. G. Belákovics Ádám, „A RECIRLCE Felhőmenedzser,” in *Networkshop Konferencia*, 2020.
- [6] OpenStack, „OpenStack Docs,” [Online]. Available: <https://docs.openstack.org/nova/rocky/user/flavors.html>. [Hozzáférés dátuma: 24 augusztus 2022].
- [7] G. Sándor, Oktatási felhő kialakítása (Diplomaterv), BME Diplomaterv portál, 2014.
- [8] BME IK, „Virtuális klaszter HPC feladatokhoz,” 2022. [Online]. Available: <https://git.ik.bme.hu/joker/joker/wikis/home>. [Hozzáférés dátuma: 24 augusztus 2022].
- [9] SchedMD, „Slurm workload manager,” 2022. [Online]. Available: <https://slurm.schedmd.com/documentation.html>. [Hozzáférés dátuma: 24 augusztus 2022].
- [10] ETH-CSCS, „Sarus - An OCI-compatible container engine for HPC,” 2018-2022. [Online]. Available: <https://sarus.readthedocs.io/en/stable/>. [Hozzáférés dátuma: 24 augusztus 2022].

Az MI lehetőségei a kora újkori filológiában: Johannes Michael Brutus  
*Rerum Ungaricarum libri* kéziratának digitális kiadása<sup>1</sup>  
(esettanulmány)

Bobák Barbara  
Magyar Tudományos Akadémia Bölcsészettudományi Kutatóközpont  
Irodalomtudományi Intézet, DigiPhil  
[bobak.barbara@abtk.hu](mailto:bobak.barbara@abtk.hu)

Kasza Péter  
SZTE BTK Klasszika-Filológia és Neolatin Tanszék  
[kasza.peter@oszk.hu](mailto:kasza.peter@oszk.hu)

2020 pandémia sújtotta nyarán Petneházi Gábor kollégával rendkívüli felfedezést tettünk az olaszországi Trento érseki könyvtárának (*Vigilianum*) kéziratai között: megtaláltuk Báthory István erdélyi fejedelem (1571–1586) és lengyel király (1576–1586) udvari történetírójának, az olasz Gian Michele Brutónak (latinos alakban használt nevén: Brutus) egy addig ismeretlen, magyar történeti tárgyú kéziratát. A felfedezés nemcsak szűkebb szakmai berkekben keltett komoly visszhangot, de a hír a nyomtatott és elektronikus sajtó ingerküszöbét is átlépte, és ennek nyomán szinte az összes komolyabb országos és vidéki lap beszámolt a Brutus-kézirat előkerüléséről.

Az alábbi közleményben egyrészt arról számolunk be, miért volt olyan jelentős ez a felfedezés, illetve milyen kihívások elé állította a szöveg feldolgozására vállalkozó filológusokat, másrészt arról, hogy a modern digitális technológia hogyan segített megfelelni ezeknek a kihívásoknak.

### Egy európai rangú humanista és műve

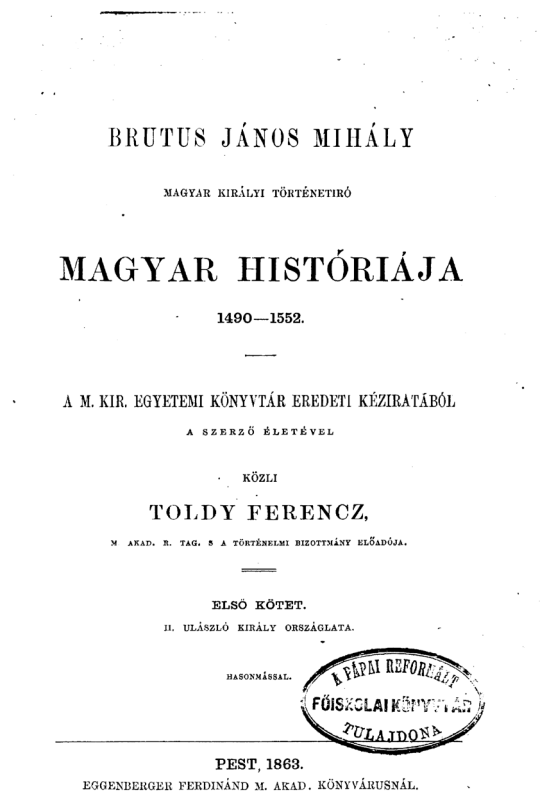
Gian Michele Bruto 1517-ben született Velencében. Ágostonos szerzetes volt, aki utóbb, a protestáns eszmék hatására kilépett a rendből. Képzett humanista volt, aki tudását igyekezett különböző fejedelmi udvarokban értékesíteni és így megélhetéshez jutni. Ennek folyamánként szinte egész Európát beutazta és jelentős hírnévre tett szert, mint jó tollú és szakavatott történetíró.

Magyarországon a 16. század a történetírás virágkora. Ennek hátterében a politikai helyzet keresendő. A Mohács után (1526) előbb két, majd Buda elestével (1541) három részre szakadó ország humanistái igyekeztek választ találni a kínzó kérdésre, mi okozhatta a 15. század végén még közép-európai nagyhatalomnak számító Magyar Királyság váratlan és szédítő gyorsaságú összeomlását. Az okok és felelősök keresése vezetett 1526 után a kortörténetírás felvirágzásához. Erősítette a tendenciát, hogy az országban két hatalmi központ jött létre: a Bécsből irányított Habsburg-ország rész állt szemben a Gyulafehérvár centrummal előbb a Szapolyai, majd a Báthory család uralma alatt fokozatosan formát öltő Erdélyi Fejedelemséggel. A két versengő államalakulat nemcsak a csatamezőn, de a propaganda mezején is háborúban állt egymással, így mindkét udvar humanisták sorát alkalmazta, hogy megírják a számukra kedvező, az ő legitimitációjukat erősítő kortörténeti narratívát. Brutus meghívása ebben a kontextusban merült föl még 1571 körül, majd a nagy tekintélyű történetíró ténylegesen 1574-ben érkezett meg Erdélybe és állt Báthory István fejedelem szolgálatába. Feladata a magyar történelem Mátyás halála és Báthory trónra lépte (1490–1571) közti évtizedeinek megírása volt. Korabeli levelezésekből származó adatok szerint a *Rerum Ungaricarum Libri*

1 Jelen tanulmány a Digitális Örökség Nemzeti Laboratórium és az OSZK-SZTE Kulcsár Péter Historiográfiai Kutatócsoport támogatásával készült.

című, monumentális latin nyelvű munka húsz könyvre tagolva az 1582–83 évekre már készen állt, de Báthory haláláig (1586) nem jelent meg. Brutus ekkor előbb váratlan húzással Habsburg-szolgálatba lépett, és megpróbálta némi átdolgozás után Bécsben kiadatni élete fő művét, de nem járt sikerrel. A kutatás jelenlegi állása szerint 1591-ben azért tért vissza Erdélybe a *Rerum Ungaricarum Libri* legutolsó változatával aládájában, hogy ott próbálja kiadni a művet, de halála (1592) meggátolta ebben.

A kiadásra csak évszázadokkal később, a 19. század második felében kerülhetett sor. De mire a Magyar Tudományos Akadémia megbízásából Toldy Ferenc hozzáfogott, hogy sajtó alá rendezze Brutus-művét, már csak két, erősen csonkult kézirat állt a rendelkezésére. Noha Brutus levelezése alapján tudni lehetett, hogy a *Rerum Ungaricarum Libri* húsz könyvet tartalmazott, ezek tetemes része a 19. századra elveszett. A bécsi *Nationalbibliothek*ban levő példány csak az 1–3, illetve 6–8. könyveket őrizte meg, tehát összesen hat könyvet, az eredeti mennyiség alig harmadát. A pesti Egyetemi Könyvtár két kötetes példánya ennél jóval bővebb szöveganyagot tartalmaz: hézagokkal ugyan, de megvan benne a *Rerum Ungaricarum Libri* 1–13. könyve, illetve a 16. egy kisebb részlete, vagyis hozzávetőlegesen a teljes mű kétharmada. Árnyalja ugyanakkor a képet, hogy míg a pesti példány 1–8. könyve gondos és megbízható másolat, addig a maradék könyvek nem csupán töredékesek, de számos kéztől származó, hevenyészett, sokszor értelmezhetetlenül romlott szöveget őriztek csak meg. Toldynak tehát ebből kellett gazdálkodnia és ezekre alapozta a három kötetben megjelent szövegkiadást.



1. kép: Toldy Ferenc által sajtó alá rendezett Brutus-mű főcímlapja

A trentói kézirat jelentőségét viszont az adja, hogy lényegében teljesnek tekinthető. Brutus autográf javításait tartalmazza, így hitelessége megkérdőjelezhetetlen. Két kötetben, összesen 2200 oldalon hozza a *Rerum Ungaricarum Libri* teljes húsz könyvét. Az utolsó könyv ugyan

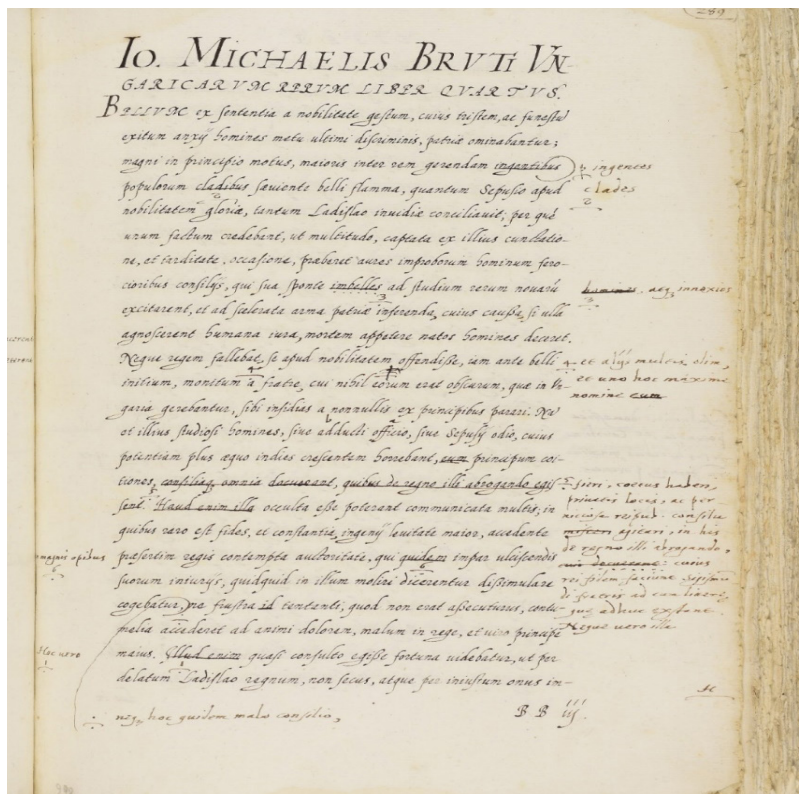


mondat közben megszakad, de minden jel arra mutat, hogy nem a kézirat csonkult, hanem maga Brutus nem írta tovább a szöveget. A felfedezés így lehetőséget kínál arra, hogy teljességében megismerjük a 16. század egyik legjelentősebb történetírójának, mindaddig éppen a hibás kéziratokra alapozott, nehezen érthető és csonka kiadás miatt elhanyagolt fő művét.

Ilyen terjedelmű és jelentőségű kézirat lényegében évszázados távlatban nem került elő. Így jogos elvárás, hogy legalább a latin szöveget mihamarabb hozzáférhetővé tegyék a szakmai közösség számára, még ha ez nem is jelent rögtön a kéziratok összevetésén alapuló, tudományos igényű kritikai kiadást, pláne nem a nagyközönség számára is hozzáférést biztosító magyar fordítást. De a mielőbbi közzététel kívánalmain túl a szöveget feldolgozó kutatók számára is létkérdés, hogy a hatalmas kéziratot és annak óriási szöveganyagát valahogy kezelhetővé, kereshetővé és így megközelíthetővé, feldolgozhatóvá tegyék. Ez praktikus azt jelenti, hogy a kéziratot számítógépre kell vinni. Nem elég a digitális fotókról beszélni, hiszen azok csak azt biztosítják, hogy a fizikailag Trentóban levő kéziratot megbízható kópia nyomán bárhol tanulmányozni lehessen. A *Vigilianum* jóvoltából és a római Magyar Nagykövetség közbenjárásának köszönhetően a kézirat teljes digitalizált változata már a birtokunkban van, de ez még nem oldja meg a szöveg kezelhetővé tételének problémáját.

Még egy évtizede sem nagyon lett volna erre más megoldás, mint hogy a kéziratot sorról sorra begépeljük. Csakhogy 2200 oldalnyi sűrű szövegről van szó. Ennek hatékony átgépelése, még paleográfiai gyakorlattal rendelkező kutató számára is komoly terhelést jelent. Egy nap átlagosan mindössze 4-5 oldal átírásának elkészítése lehetséges, így a 2200 oldal átírása egy ember közel másfél-két évi munkáját jelentené. Nem beszélve arról, hogy a szövegben óhatatlanul számos hiba, elütés marad. Ha az ellenőrzést is hozzávesszük, lényegében a pusztán és valamennyire megbízható közlés közel három évet vesz igénybe.

Számunkra azonban kivételes lehetőséget biztosított az a tény, hogy bár hatalmas mennyiségű szövegről van szó, a 2200 oldalból 2000 teljesen egységes kézírással íródott.



2. kép: A Brutus-kézirat egy oldala



Nem Brutus-autográf ugyan a szöveg, de Brutus írnokának levelekből ismert kézírását tartalmazza. Az egységes kézírás pedig lehetőséget adott arra, hogy az MI-alapú Transkribus-szolgáltatást<sup>2</sup> alkalmazva hidaljuk át a problémát. Petneházi Gábor kollégával ezért felvettük a kapcsolatot a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetében működő DigiPhil munkacsoporttal, és velük összefogva kezdtük meg a Brutus-szöveg feldolgozását.

Előbb az SZTE Klasszika-Filológia és Neolatin Tanszékén működő Neolatin Doktori programnak a latinban és a kéziratolvasásban egyaránt jártas doktorandusz hallgatónak (Csapó Fanni, Kintli Dóra, Majoros Máté) bevonásával átírtunk 200 kéziratoldali szöveget, majd Bobák Barbara bevonásával megkezdődött a projekt Transkribus-szakasza.

## Automatikus kézírásfelismerés

A DigiPhil munkacsoport 2018-ban kezdett el a Transkribus átíróprogrammal dolgozni, Arany János akadémiai hivatali levelein.<sup>3</sup> Ez a projekt volt az első, amely magyar nyelvű kézirat átíratását célozta meg a Transkribus segítségével és ebből a munkából született meg az első – és azóta többször fejlesztett – magyar nyelvű modell is, amely a kézírásfelismeréshez elengedhetetlen.<sup>4</sup> Az évek során a Transkribus jelentős népszerűsége tette szert – dacára annak, hogy ma már nem ingyenes a szolgáltatás<sup>56</sup> –, amit mi sem mutat jobban, mint hogy egyre többféle nyelvre egyre több modell áll rendelkezésre. Népszerűsége többek között annak köszönhető, hogy a kézírásfelismerés teljes folyamatát magába foglalja, amit egy felhasználóbarát grafikus felülettel tesz elérhetővé. A Transkribus szolgáltatásban a kéziratok átíratása csupán néhány egyszerűbb lépésből áll. A beszkenelt kéziratokról készült képfájlok feltöltését követően a képfájlok szegmentálást végzünk, amely folyamatot layout analízis segít, illetve ennek során a program beazonosítja azokat a területeket, ahol szöveg található. A teljes szöveget és szövegegységeket úgynevezett dobozokba teszi és sorokra, valamint szavakra szegmentálja azt. A szegmentálást követően nyelvi modellt kell választanunk, amely az adott kézirat nyelvének és szerzőjének karakterkészletét sajátosságait foglalja magába, azok felismertetésére szolgál. A Transkribusban számos nyilvános modell<sup>7</sup> közül választhatunk, így a latin nyelvű Brutus kéziratot a kutatócsoport először tesztelte a meglévő latin nyelvű modelleket. Mivel egyik sem hozott eredményt, a Brutus kéziratához saját modell létrehozására volt szükség, melynek tanítóanyagként az említett kézzel átírt 200 oldal szolgált. Az átíró program dokumentációja<sup>8</sup> a tanítóanyag méretét 5 000 és 15 000 szó között határozza meg, amely mennyiség az adott nyelv komplexitásától, illetve a kézirat sokszínűségétől függően változhat. A 200 oldal annak érdekében határoztuk meg, hogy hatékonyabb legyen a nyelvi modell, amely így egészen pontosan 54 344 szón tanulhatott. Bár a modell egyelőre csak a kutatócsoport számára érhető el, a közeljövőben bekerül a nyilvános modellek közé.

---

2 A Transkribus online szolgáltatás, melynek segítségével nyomtatott szövegek és kéziratok felismertetését és átíratását végezhetjük. Az alábbi weboldaltól tölthető le ingyenesen, regisztrációt követően: <https://readcoop.eu/transkribus/download/>

3 Bobák, Barbara és Gábori Kovács, József (2019) Kézírásfelismerés Arany János levelein. In: Networkshop 2019. HUNGARNET Egyesület, Budapest, pp. 38 – 44.

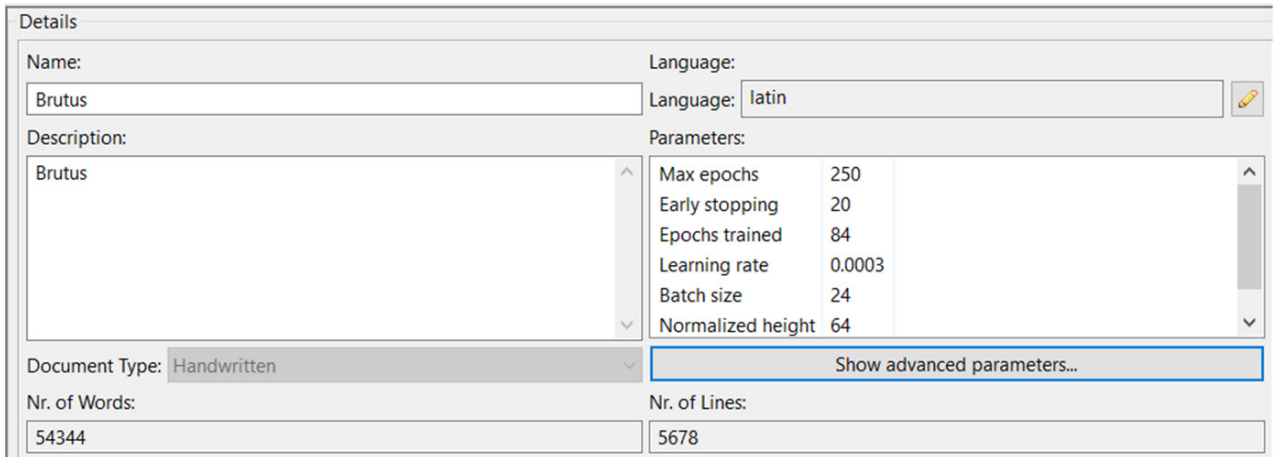
4 A magyar nyelvre létrehozott modellek köre azóta bővült egy 19. századi, Kiss József levelezését feldolgozó projektnek köszönhetően: Mihály, Eszter (2021) A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben. In: Online térben az online térért, Budapest, pp. 91–109.

5 A munkafolyamatok zavartalan működése érdekében a BTK Irodalomtudományi Intézet krediteket vásárolt a Transkribus szolgáltatáshoz.

6 <https://readcoop.eu/transkribus/credits/>

7 <https://readcoop.eu/transkribus/public-models/>

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



Details													
Name:	Brutus												
Language:	latin												
Description:	Brutus												
Parameters:	<table border="1"> <tr><td>Max epochs</td><td>250</td></tr> <tr><td>Early stopping</td><td>20</td></tr> <tr><td>Epochs trained</td><td>84</td></tr> <tr><td>Learning rate</td><td>0.0003</td></tr> <tr><td>Batch size</td><td>24</td></tr> <tr><td>Normalized height</td><td>64</td></tr> </table>	Max epochs	250	Early stopping	20	Epochs trained	84	Learning rate	0.0003	Batch size	24	Normalized height	64
Max epochs	250												
Early stopping	20												
Epochs trained	84												
Learning rate	0.0003												
Batch size	24												
Normalized height	64												
Document Type:	Handwritten												
Nr. of Words:	54344												
Nr. of Lines:	5678												

3. kép: a Brutus nyelvi modell metaadatai

A Transkribus szolgáltatás számos előnye közé sorolható, hogy felhasználói felülete – amelynek megléte sem triviális az átíró szolgáltatások között<sup>9</sup> – könnyen kezelhető, kifejezetten felhasználóbarát. Az átírás minden szakasza, maga a modellépítés is néhány kattintással elvégezhető, kedvezve mindazon felhasználóknak, akik kevésbé járatosak a különböző online szolgáltatások, programok világában. Az átíratás folyamatának következő lépése, hogy a létrejött Brutus modell segítségével lefuttatjuk a HTR<sup>10</sup> modult, amely magát a kézírásfelismerést végzi a modellben rögzült nyelvi és karakterkészletű sajátosságok alapján. A Transkribusban lehetőség van arra, hogy az átírás egyes lépéseinek végeredményét kézzel korrigáljuk. A szöveg szegmentálását, illetve az átíratott szöveget is tudjuk javítani, kiegészíteni, törölni vagy pótolni.

### A Brutus nyelvi modell és hatékonysága

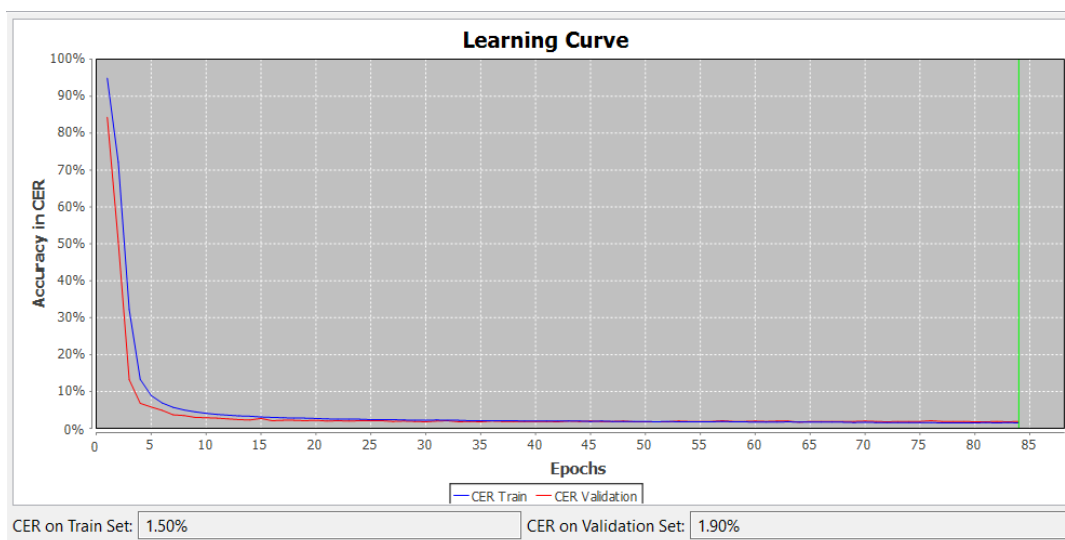
A Brutus-kézirat gépi átíratása rendkívül eredményesnek tekinthető. A Transkribus dokumentációja kitér arra, hogy a nyelvi modellek hatékonyságát mutató karakterhiba-arány<sup>11</sup> hogyan értelmezhető.<sup>12</sup> Az automatizált átírásnál a 10% alatti hibaarányt eredményező nyelvi modell magas hatékonyságúnak tekinthető. Esetünkben a Brutus modell karakterhiba-aránya 1,90%-ot ért el, azaz az átírt szöveg kevesebb, mint 2%-át rontotta csak el. Az alábbi koordináta-rendszer a Brutus nyelvi modell tanulási görbét (Learning Curve) mutatja:

<sup>9</sup> A Google Tesseract például kizárólag parancssoros módon használható.

<sup>10</sup> A HTR feloldása Handwritten text recognition, azaz kézírásfelismerés, képfájlok átalakítása feldolgozásra alkalmas szövegfájl formátummá.

<sup>11</sup> Character error rate (CER): a hibásan átírt karakterek aránya százalékos formában.

<sup>12</sup> <https://readcoop.eu/transkribus/howto/how-to-train-a-handwritten-text-recognition-model-in-transkribus/>



4. kép: a Brutus nyelvi modell tanulási görbéje

Az ábrán látható koordináta-rendszer y-tengelyén (Accuracy in CER) a karakterhiba-arány pontossága látható, százalékos formában. Ezen a tengelyen a görbe mindig 100%-tól indul, hiszen a nyelvi modell a tanulási folyamat elején nem rendelkezik információkkal az adott kézírásról. A tanulás során a modell fejlődik, melynek mértékét a görbe esése mutatja.

Az x-tengelyen (Epochs) azon szakaszok arányos leosztása látható, ahol a tanulási folyamat megáll, és a Transkribus értékeli a tanulás eredményességét. A nyelvi modell létrehozásakor megadható ezen szakaszok mennyisége – esetünkben 85 szakasz –, mellyel tehát azt határozhatjuk meg, hányszor kerüljön kiértékelésre a betanítási adat. Az ábra alján látható még a konkrét tanítóanyag<sup>13</sup> (1,50%), illetve a tesztoldalakon<sup>14</sup> (1,90%) ejtett karakterhibák százalékos aránya. A modellépítésnél a tanítóanyag 5–10%-át tesztoldalként különíti el a program, de kézzel is válogatható vagy bővíthető a tesztanyag.

A Transkribusban tehát a nyelvi modelleknek nemcsak a metaadatait láthatjuk, hanem a tanulási görbéjét is, amely megmutatja a modell tanulási folyamatát (kék vonal), illetve eredményességét (piros vonal). Az átíró program működésének minőségén túl a gyorsasága is számottevő. Bár a modellépítéshez szükséges tanítóanyag átírása elkerülhetetlen, ez általában a teljes anyag maximum 20%-át jelenti. A Brutus kézirat több mint 80%-nak átírása a program segítségével egy napba sem telt, az igen alacsony hibaaránynak köszönhetően pedig az átírat ellenőrzése sem túl időigényes. A megspórolt munkaidőn és energián túl mind a feldolgozandó kézirat mennyisége, mind az egységessége – amely a Brutus nyelvi modell magas hatékonyságát is eredményezte – alátámasztotta a Transkribus melletti döntést.

<sup>13</sup> Train Set

<sup>14</sup> Validation Set



## Bibliográfia

Bobák, Barbara és Gábori Kovács, József (2019) Kézírásfelismerés Arany János levelein. In: Networkshop 2019. HUNGARNET Egyesület, Budapest, pp. 38 – 44.

<http://ocs.mtak.hu/index.php/nws/2019/paper/view/23/40>

Mihály, Eszter (2021) A digitális szövegkiadások nehézségei és lehetőségei a közgyűjteményekben. In: Online térben az online térért, Budapest, pp. 91–109.

<http://ocs.mtak.hu/index.php/nws/2021/paper/view/69/94>

Transkribus – Public models. Hozzáférés: 2022.06.16.

<https://readcoop.eu/transkribus/public-models/>

Transkribus – How to guides. Hozzáférés: 2022.06.16.

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

Transkribus – Model training. Hozzáférés: 2022.06.16.

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

**Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok**

Egyed-Gergely Júlia, Vajda Róza, Gárdos Judit, Horváth Anna, Meiszterics Enikő  
*Társadalomtudományi Kutatóközpont (TK KDK)*  
[egyed-gergely.julia@tk.hu](mailto:egyed-gergely.julia@tk.hu)

Micsik András, Martin Dániel, Marx Attila, Pataki Balázs, Siket Melinda  
*Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI DSD)*  
[micsik@sztaki.hu](mailto:micsik@sztaki.hu)

**Abstract**

Social science, research data and artificial intelligence: options and lessons learned

Our aim was to facilitate the research of collected interview texts at the Research Documentation Centre of the Centre for Social Sciences. Currently it is difficult to find the parts related to specific topics or obtain a thematic mapping of lengthy interviews and collections. To this end, we identified and tested the most promising NLP tools supporting the Hungarian language. Furthermore, a suitable and domain-oriented taxonomy had to be created for the classification of available texts. Our work and experiences are described in this paper, including the adaptation of the European Language Social Science Thesaurus, the building of a gold standard for classification, the comparison of various automated indexing methods, and finally the implementation of a researcher tool supporting custom visualizations and faceted search.

**Keywords:** natural language processing (NLP), named entity recognition, extreme classification, exploratory UI, text visualization

**Bevezetés**

A Társadalomtudományi Kutatóközpont Kutatási Dokumentációs Központja (TK KDK) a TK négy intézetében (Jogtudományi Intézet, Kisebbségkutató Intézet, Politikatudományi Intézet, Szociológiai Intézet) folytatott kutatások során keletkezett dokumentumokat (interjúkat, kérdőíveket, kutatási terveket stb.) gyűjti össze, rendszerezi és teszi elérhetővé online felületén az anyagok másodfelhasználásában, illetve tudománytörténeti és/vagy módszertani vizsgálatában érdekelt kutatók számára. 2019 óta a KDK az 1960 és 2010 közötti hazai kvalitatív szociológiai kutatások szétszórtan heverő anyagainak (elsősorban interjúleíratoknak, emellett fotóknak, rajzoknak, videóinterjúknak, jegyzeteknek, tanulmánytöredékeknek stb.) gyűjtésével, digitalizálásával és kutatási célokra való feldolgozásával foglalkozó 20. Század Hangja Archívum és Kutatóműhelyt is magába foglalja. A két online dokumentumtár több tízezer digitális fájlja ingyenesen hozzáférhető, de az anyagok egy része regisztrációköteles.

Az állomány folyamatos bővülése és az átláthatóság kívánalma egyre nagyobb kihívások elé állítja az archívumokat üzemeltető KDK munkatársait. A dokumentumok alapadatait rögzítő metaadat-struktúra ugyanis csak korlátozottan teszi lehetővé az anyagok tematikus áttekintését, kereshetőségét, rendszerezését, ráadásul a két archívum metaadatrendszere különbözik egymástól. Mindkét archívum alkalmaz ugyan tárgyszavakat, azonban nem szöveg- illetve dokumentumszinten, illetve nem egységes szókészletből, és nem azonos módszerekkel történik a hozzárendelés. Az eddigi általános gyakorlat szerint kutatásszinten

(egy kutatáshoz rendszerint több dokumentum is tartozik) szerepel néhány – a 20. Század Hangja Archívumban 3-3, a KDK anyagai esetében sem sokkal több – esetlegesen meghatározott tárgyszó. Mindezek miatt az archívumokat látogató kutatóknak invenciózusan saját módszerekhez kell folyamodniuk, hogy a céljaiknak megfelelően feltérképezhessék a szövegállományt és kiválaszthassák a munkájukhoz szükséges dokumentumokat.

A tartalomban történő böngészést szövegszerű keresési funkció szolgálja, a két archívumban külön-külön. Ez persze nagyon hasznos, ha egészen specifikus fogalom vagy helyszín érdekli a kutatót. Amint azonban általánosabb jelenségre, tématerületre kíváncsi, kénytelen először “lefordítani” azt különféle széles körben használt kifejezésekre, amelyekről feltételezi, hogy szó szerint szerepelhetnek a szövegekben. Márpedig javarészt strukturálatlan interjúkból álló, meglehetősen heterogén szövegállomány esetében nem kis munka és fejtörés különféle fogalmakkal „körüllőni” egy-egy témát. Például, ha a társadalmi mobilitás kérdése foglalkoztatja az illetőt, akkor kezdheti olyan szavakkal, mint „szegénység”, „pénz”, „oktatás”, és folytathatja még hosszan a sort. Azonban nem remélheti, hogy vaktában eltalálva az összes vonatkozó kifejezést rálel valamennyi, számára releváns dokumentumra. Ugyanígy, ha mondjuk a főváros története érdekli, nem elégedhet meg pusztán a „Budapest” szóra kereséssel, hanem végig kellene próbálnia valamennyi helynevet a Margit hídtól Ferencvárosig, ami kész lehetetlenség. De jogos kutatói igény lehet az is, hogy valaki egyszerre lássa át, melyek a leggyakrabban említett helyek, szereplők, intézmények, időszakok, vagy tudja meg, melyik kutató mikor, mivel foglalkozott.

Az anyagok „feltárása”, az állomány áttekinthetővé tétele mint sürgető feladat és a feladat megoldásával kecsegtető technológiai újítások elérhetővé válása együtt adta a projektünk alap gondolatát. Kézi tárgyszavazás ilyen nagy szövegtömegre óriási mértékű humán erőforrást igényelne, ezért a lehetőségeink ismeretében ez fel sem merült. Mesterséges intelligenciára épülő algoritmusokat ezidáig ugyan más, egyszerűbb problémák megoldására használtak a társadalomtudományokban, mégis úgy gondoltuk, ez az eszköz alkalmas lehet a céljaink eléréséhez.

### A tárgyszókészlet és a tanítóhalmaz létrehozása

A két archívum kutatási anyagainak gépi tárgyszavazását többfázisos manuális előkészítés előzte meg. A mesterséges intelligenciát is használó algoritmusok betanításához, illetve a kapott eredmények validálásához tanítóanyagra volt szükség. Ehhez létre kellett hozni egy egységes tárgyszókészletet, meg kellett határozni a tanulókorpuszt, tesztelni rajta a tárgyszókészletet, majd előállítani a gold standardot. E lépések több körben, egymást javítva követték egymást.

### Tárgyszókészlet

Az archívumok kutatási anyagainak tárgyszavazásához strukturált tárgyszókészletre van szükség (Molnár, 2016). Magyar nyelvű társadalomtudományos tárgyszókészlet a munka kezdetén nem állt rendelkezésre – a speciális feladat miatt ráadásul olyan készletre volt szükség, amelyik elég részletes ahhoz, hogy lefedje a két archívum anyagainak témáit, viszont kellően szűk ahhoz, hogy gépi módszerekkel dolgozni lehessen vele, mindemellett később a két archívum közös keresőjének alapjául is szolgálhat. E feltételek egymással ellentétes irányba hatnak: az első és a harmadik minél részletesebb szókészletet, míg a második inkább átfogó szakszavakat igényel. Az egyensúly megtalálása az egyik oldalról kompromisszumokkal jár, míg a másik oldalról kemény korlátokba ütközik. Fel kellett tehát adni az aprólékos, részletekbe menő tárgyszó hozzárendelést, és alkalmazkodni a kódolás korlátos lehetőségeihez.



Az általános magyar tezausz (OSZK Köztaurusz), különböző szakkönyvek tárgyszavait és különböző nemzetközi társadalomtudományos szókészleteket átnézve kiindulásul a CESSDA<sup>1</sup> (Consortium of European Social Science Data Archives) ELSST<sup>2</sup> (European Language Social Science Thesaurus) angol nyelvű tezauszára esett a választás. Emögött az a megfontolás állt, hogy a kutatási területen bevett nemzetközi szókészlet választása egyfelől elősegíti egy hiánypótló magyar társadalomtudományos tárgyszókészlet előállítását, másfelől lehetőséget biztosít nemzetközi projektekbe való bekapcsolódásra, valamint a két archívum nemzetközi láthatóságának növelésére. A munka során maga a fordítás önálló projektté vált, a TK és a CESSDA együttműködése eredményeként a magyar változat 2022 szeptemberében kerül az eddig 14 nyelven elérhető szókészlet mellé és lesz hozzáférhető az ELSST honlapján.

A saját tárgyszókészlet előállításához először a valamivel több mint 3300 kifejezésből álló tezausz magyarra fordítására került sor. Az eredeti angol nyelvből a SZTAKI kutatói által létrehozott SZTAKI-s és más online szótárak segítségével gépi, majd ezt követően manuális módszerrel ültették át magyar nyelvbe a kifejezéseket. Ezután a fordítás többszempontú (jogi és nyelvészeti) lektorálása következett, majd az eredményt a projekthez igazítottuk (Albaugh et al, 2013, Albaugh et al, 2014). A fordítás olyan nem várt dilemmákat vetett fel, mint például az angol nyelv alapvetően többesszámot használó formájának átvétele vagy a kifejezések egyesszámúsítása, a magyar nyelvben nem használt (például speciális jogi) kifejezések megfelelő fordításának megtalálása, vagy a nem társadalomtudományos kifejezések problémája. A munka során újragondoltuk a nemzetközi szókészlet bizonyos elemeit, amivel támogattuk annak megújulását – a konzorcium folyamatosan karban tartja a tárgyszókészletet, az éves frissítés részben a magyar fordítás tapasztalatai alapján történik.

A szókészlet testreszabása során többkörös szűkítéssel, tömbösítéssel, pontosítással és kiegészítéssel elkészült egy 220 elemű, a gépi tárgyszavazáshoz alkalmasnak tűnő tárgyszólista. A lista kialakításánál szempont volt a relevancia (társadalomtudományi fogalmak kiválasztása a magyar szociológia témáinak, alapfogalmainak figyelembevételével), a lefedettség (minden fontos társadalomtudományi tématerület képviselése), az arányosság (az archívumok súlypontjaihoz való igazodás), a diszjunktivitás (jól elkülönülő fogalmak kiválasztása) és a mennyiség (gépi algoritmusok számára befogadható számú fogalom). Az eredeti lista drasztikus szűkítése mellett szükség volt az új kifejezések felvételére is, elsősorban a magyar szociológia és a magyar történelem témáinak megragadhatósága végett. A tárgyszólistát háromszintű hierarchikus rendszerbe rendeztük – támogatandó a manuális és gépi kódolást, valamint a KDK keresőjének fejlesztését.

## Tanítóhalmaz

A tanítóhalmaz létrehozásához a tanulókorpuszt a két archívum interjúkat tartalmazó kutatási anyagaiból kiválasztott 21 (735 oldalnyi) interjú adta. A kiválasztásnál szempont volt, hogy különböző témájú (a szociológiai szakma nagyjaival készült élettörténeti interjúk, börtönviseltekkel folytatott beszélgetések, magukat romáknak vallókkal készített interjúk, a Kádár-korszak emlékezetével kapcsolatos kutatás során készült anyagok, stb), különböző hosszúságú, eltérő módszerrel készült (narratív interjú, félig strukturált interjú, fókuszcsoportos beszélgetés), más-más gyűjteményekből származó és mindkét archívumot megjelenítő anyagok kerüljenek a korpuszba.

A kiválasztott interjúk manuális annotálását összetett feladatként határoztuk meg: a tanítóhalmaz kialakításához az annotátoroknak kulcs- és tárgyszavakat<sup>3</sup> kellett rendelniük

---

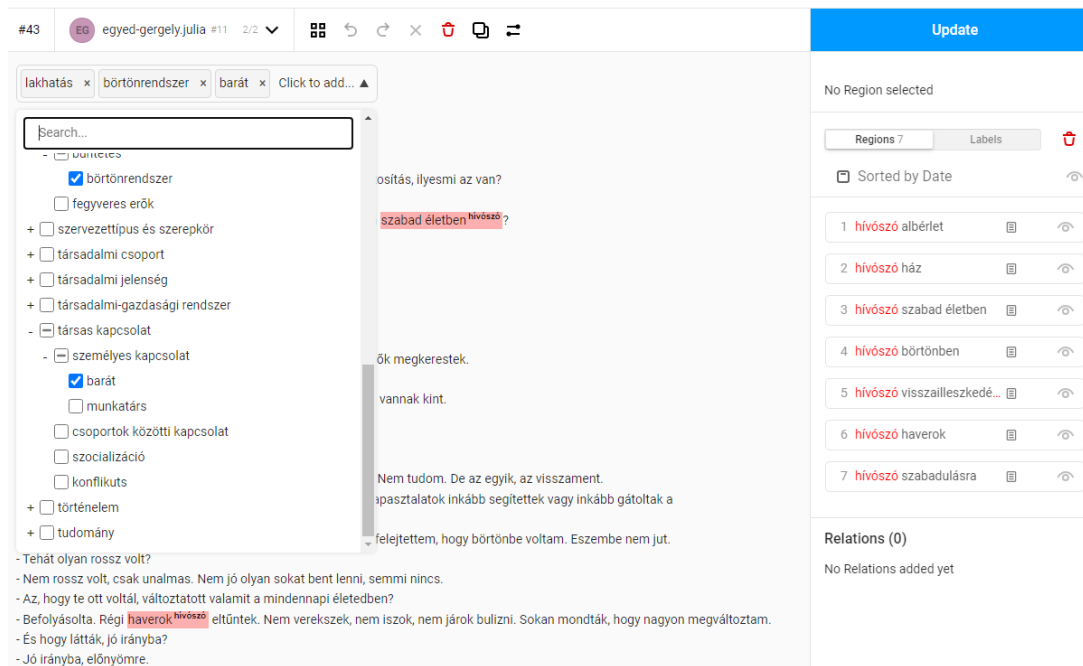
1 <https://www.cessda.eu/>

2 <https://thesauri.cessda.eu/elsst/en/index>

3 Míg a kulcsszavak a gépi annotálást hivatottak segíteni, addig a tárgyszavak feladata az interjú metaadatainak bővítése.

a szakaszokra bontott interjúszövegekhez. A hozzárendelés célja kettős volt: a tanítóanyag elkészülése mellett a saját tárgyszókészlet validálására, korrigálására is szolgált.

Az annotálást szakértők végezték, előzetesen kialakított irányelvek alapján (Molnár, 2016). A nagy mennyiségű tárgyszó és az alapvetően strukturálatlan szövegek miatt komoly kihívást jelentett az egységesség (Balázs-Sebők, 2016). Az annotálás Label Studio<sup>4</sup> felületen zajlott (1. ábra), amelyet a saját tárgyszókészlet betöltésével és az interjúfeldolgozáshoz alkalmas struktúra és megjelenítés kialakításával a SZTAKI kutatói a projekthez igazítottak. A szakaszokra bontott interjúkat először két-két annotátor egymástól függetlenül kódolta, majd az egyezés mértékétől függően vagy az eredeti kettő, vagy harmadik, független annotátor határozta meg a gold standardot.



1. ábra. Annotálás közben a Label Studio felülete és a hierarchiába rendezett tárgyszókészlet

Az egyezés meghatározását a projekt jellegéhez igazodva, saját elvek alapján alakítottuk ki. Egyezőnek tekintettük a tárgyszó-hozzárendelést, ha a tárgyszókészlet hierarchiájában a hozzáadott tárgyszóhoz tartozó, a legfelső szinten lévő kifejezés azonos volt. Azaz például, ha egy szakaszhoz az egyik annotátor a „szokások és hagyományok” tárgyszót, míg a másik a „kulturális esemény” tárgyszót adta, akkor ezt egyezésként fogadtuk el, ugyanis a hierarchiában mindkét kifejezés a „kultúra” tárgyszó alatt helyezkedik el. A két annotátor, bár más tárgyszót választott, ugyanazt a nagyobb területet jelölte velük.

Az egyezés kívánt mértékét szintén a feladat sajátosságai szerint határoztuk meg. 30% alatti egyezésnél harmadik annotátor, 30% feletti egyezésnél az eredeti két annotátor döntött a gold standardról. A szakirodalom ennél szigorúbb egyezést szokott elvárni, a speciális alkalmazás azonban speciális feltételek kialakítását igényelte és tette lehetővé. A „hagyományos” annotálási feladatokhoz képest több szempontból is különlegesnek ítéltük meg a helyzetet. Az interjúkat nem bonthattuk mondatokra vagy nagyon rövid (és ennek megfelelően várhatóan egyértelmű és könnyebben megfogható tartalmú) egységekre, ebben az esetben túlságosan sok, és gyakran irreleváns tárgyszót kaptunk volna. Egy-egy szövegegység általában több témát ölelt fel, az annotátoroknak ezek közül kellett megjelölnie a relevánsabbakat, a tárgy megjelölésén túl így a kiválasztás is feladat volt. Egy adott szakaszhoz nem egy (a legrelevánsabb) tárgyszót, hanem többet is lehetett rendelni, a hangsúlyosan

4 <https://labelstud.io/>

megjelenő témák számától függően. Ez nagyban nehezítette az egységes eredmény elérését, komoly kihívás elé állítva a munkafolyamat kialakítóit és az annotátorokat (Albaugh et al, 2014). A munka jellegéből adódóan a tárgyszavak egymáshoz képest nem kizáróak, nem is lehetnek azok (mint például a pozitív/negatív/semleges kódolásánál), így két kódoló által adott különböző tárgyszavak egyformán relevánsak lehetnek akár még akkor is, ha még csak nem is ugyanahhoz a nagyobb tématerülethez tartoznak. Az egy szakaszban megjelenő különböző témák közül a relevánsak kiválasztásának módja nehezen egységesíthető. Egy szakaszban megjelenhet egymás mellett például a „térbeli elhelyezkedés”, a „jövőkép” és a „fogyasztás” mint fontosabb téma, ezek közül bármelyik és bármelyek kombinációjának kiválasztása releváns döntés lehet a kódoló részéről. Végül a korpusz jellege (gyakran leginkább a köznyelvhez hasonló interjúszövegek) is indokolta a speciális – a szokásosnál megengedőbb – irányelvek kialakítását.

A gold standard és ezzel a tanítóhalmaz előállítása után a KDK és a 20. Század Hangja Archivum nagy mennyiségű interjúanyagának gépi annotálása következett, 368 interjú bevonásával, majd az eredmények manuális validálása véletlenszerűen kiválasztott interjúk ellenőrzésével.

### Automatikus tárgyszavazás

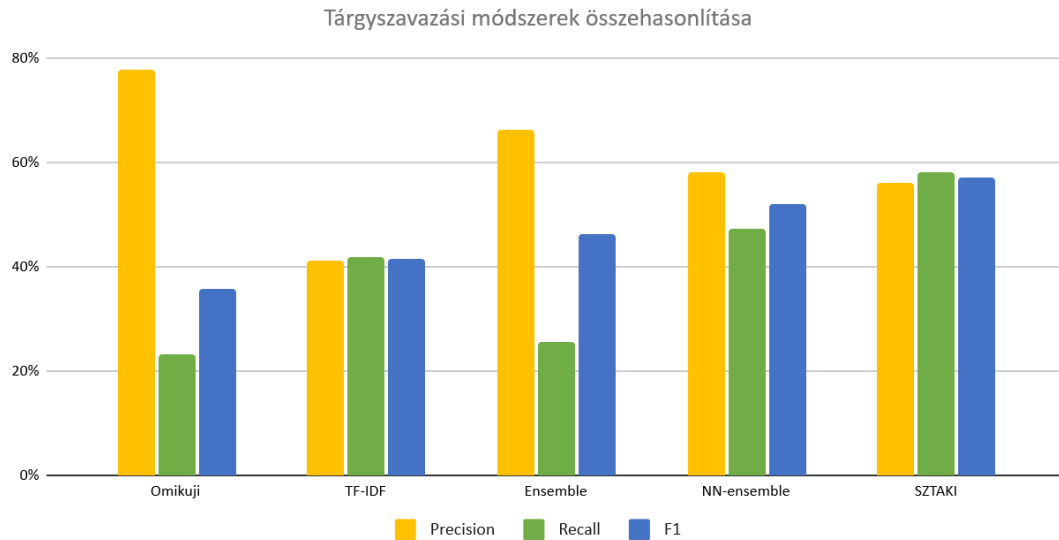
Az automatikus tárgyszavazási feladat főbb informatikai jellemzői: 2-5000 karakternyi szövegrészekhez (szakaszokhoz) kell néhány jellemző tárgyszót kiválasztani egy háromszintű tárgyszó taxonómiából. Ez egy elég speciális feladat, amelyhez a hagyományos klasszifikációs MI algoritmusok nem használhatók a nagy taxonómia miatt, a topic modelling megoldások pedig azért, mert előre adott a taxonómia. Szerencsére kifejezetten erre a célra jött létre az annif szoftvereszköz (Suominen et al. 2022), amellyel tízféle módszer közül lehet választani ilyen jellegű tárgyszavazáshoz. A módszerek között van tanuló és statisztikus, illetve kombinálni is lehet a módszereket. Az annif wiki<sup>5</sup> oldalán megtalálható az egyes módszerek rövid leírása és a megalapozó tudományos cikkek hivatkozásai. Az összes módszert kipróbáltuk a gold standard korpuszunkkal, és a legjobban teljesítők eredményét tüntettük fel a 2. ábrán. A vizsgált módszerek közül a legjobb eredményt elért NN-ensemble az Omikuji és a TF-IDF tárgyszavait 3:1 arányban kombinálja.

Kicsivel ennél is jobb eredményt ért el saját fejlesztésű módszerünk, amely a tanítóhalmaz készítése során a tárgyszavakhoz gyűjtött kulcsszavak használatán alapul. Itt a textacy<sup>6</sup> és a legfrissebb huspacy (Orosz et al. 2022) segítségével kulcsszavakat gyűjtöttünk a szövegrészhez, és a kulcsszavakhoz tartozó tárgyszavak statisztikája alapján alakítottuk ki a végső tárgyszólistát. A textacy előnye, hogy többféle kulcsszavazási módszer is választható. Ezek közül számunkra a yake adta a legjobb javaslatokat, ezt szubjektív vizsgálati módszerekkel sem volt nehéz eldönteni.

---

5 <https://github.com/NatLibFi/Annif/wiki>

6 <https://github.com/chartbeat-labs/textacy>



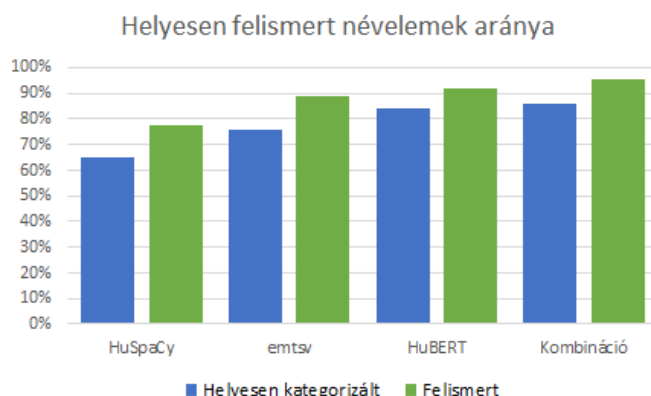
2. ábra. A különböző tárgyszavazási módszerek eredményeinek összehasonlítása

Az első és második helyezett módszer fenntarthatóságát érdemes megvizsgálni. Például, mi történik, ha meg kell változtatni a tárgyszórendszert? Az első helyezett módszer esetében a megváltozott tárgyszavaknál át kell alakítani a társított kulcsszólistát, és validálni kell egy kialakított kisszámú mintán. Ezután újra kell futtatni a tárgyszavazást az interjúkon (vagy egy részükön), amely időigényes lehet, mivel a szöveget nyelvileg elemezni kell (szótövezés, szófaj meghatározás stb.).

A második helyezett módszer esetében kissé bonyolultabb a helyzet: a módszert újra be kell tanítani, ehhez pedig ki kell bővíteni a tanulóhalmazt a változott tárgyszavak elegendő számú előfordulásával. Ez egy iteratív folyamat, amelynek során, ha az eredmények már elég jók, akkor lehet újrafuttatni a tárgyszavazást az interjúkon.

### Névelemek felismerése, összekapcsolása

Másik fontos célunk volt kinyerni az interjú szövegekből az említett földrajzi, személy- és egyéb tulajdonneveket. A névelem-felismerés vizsgálatához kialakítottunk egy több mint 3500 névelemet tartalmazó minikorpuszt az interjúkból, és azon futtattunk három névelem-felismerőt: a már említett huspacy-t, a korábban készült emtsv emBert modulját (Nemeskey 2020), valamint a huBERT-et (Nemeskey 2021). Utóbbit az időközben elkészült NYTK-NerKor korpuszon (Simon 2021) tovább tanítottuk, és mivel ez a legbővebb és legfrissebb névelem-felismerési korpusz, talán ennek tudható be, hogy ezzel értük el a legjobb eredményt. Még ez sem volt azonban elég jó, a névelemek majdnem tizede kimaradt. Amikor a három módszer különbözőképp ítelt meg egy névelemet, egyszerű szabályok alapján (pl. találtunk-e hozzá Geonames azonosítót) kombinált eredményt alakítottunk ki, amelyhez már a wikifikálás kimenetét is felhasználtuk. Ezen a területen is van még fejlődési lehetőség, mivel a névelemek 5%-a rejtve maradt, 9%-a pedig rossz kategóriába került.



3. ábra. A különböző nyelvi elemzők névelem felismerési eredményei

A felismert névelemekre rákerestünk a Wikidata-ban azért, hogy a következő fejezetben bemutatott interaktív interjú olvasóban linkeket tudjunk felajánlani az egyes nevek említéseihez. A Wikidata aggregálja sok más regiszter azonosítóját, így a hivatkozások nagy részét a Wikidata-ból meg lehet szerezni (pl. PIM, VIAF, Geonames). Viszont az egyértelműsítés (*disambiguation*) terén sok munka lenne még: a szövegekben előforduló közgazdászok, írók, tudósok helyett rendszeresen azonos nevű focisták és más celebritások linkjeit kaptuk első helyen a Wikidata-tól. Pár, a tudásgráfból lekérhető alapvető jellemzőt megpróbáltunk figyelembe venni a választáskor, mint például az entitás típusa vagy a személy születési éve, de ez sem volt elegendő az eredmények jelentős javulásához.

## Kutatói munkafelület

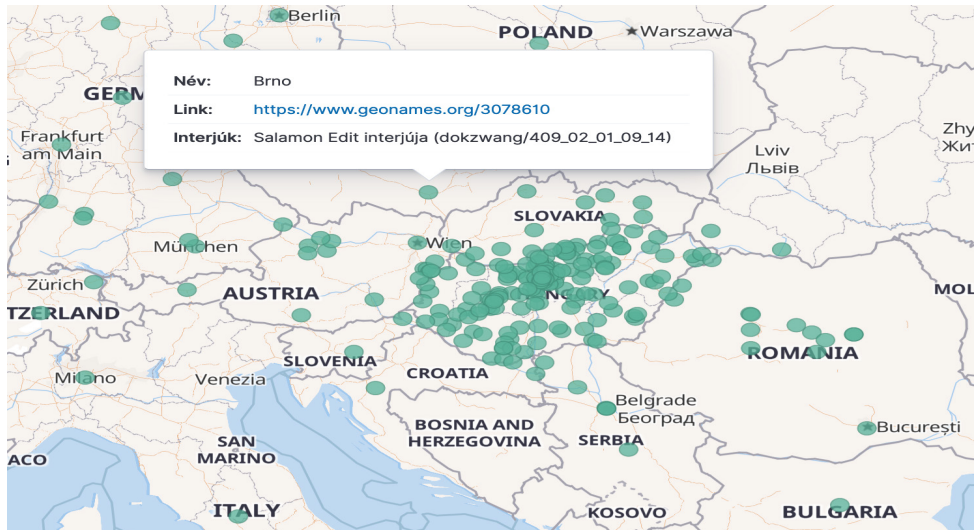
Az összes interjú nyelvi elemzésével előállt egy adatbázis, amelyet a kutatók számára felfedezhetővé akartunk tenni. Ezért nyílt forráskódú komponensekből összeállítottunk egy fazettás keresőt, amelyben a teljes szöveges keresés mellett változatos szűrési lehetőségek is vannak (4. ábra). Az ábrán középen a kék keretes dobozokban látható a két aktív szűrőfeltétel, illetve a szűrt elemek piros háttérrel is meg vannak jelölve.

The screenshot shows a search interface with the following elements:

- Collection:** Romainterjúk (32), **Börtöninterjúk** (21), MABISZ (14), Szociológiatörténet (14), Foglalkozási diszkrimin... (6), A rendszerváltás demo... (5), Jobboldali radikalizmus (3), Kádár-korszak (2).
- Search filters:** Search bar, Topic: technológia és innováció, Collection: Börtöninterjúk, Reset Search.
- Results:** 21 Results. Document path dropdown.
- Result details for 'Börtöninterjú #101':**
  - URL: <https://openarchive.tk.mta.hu/185/>
  - Document path: bortoninterjuk/101
  - Date: 2013
  - Interviewees: első alkalommal végrehajtandó szabadságvesztésre ítélt, 30 éven aluli, rövid büntetési tételű (3 évig terjedő) férfi
  - Interviewers: F. A. (férfi)
  - Topics: mindennapi élet, származás, szegénység, **technológia és innováció**
  - Yake keywords: gyerek, segítség, buli, tető, börtön, alkalmazott, értelem, esély, pártfogó, család, interjú, program, autó, vállalkozó, sulí, hazajárás, folyamat, jogsi, későbbi, após, haverik
  - Wikifier keywords: Társadalombiztosítás, Társadalmi szervezet, Társadalom

4. ábra. Fazettás keresés az interjú-gyűjteményekben





5. ábra. Az interjúkban említett földrajzi nevek térképes megjelenítése

A találati listában az interjúk főbb metaadatai, valamint az automatikus elemzés interjóra összesített leggyakoribb tárgyszavai, kulcsszavai és Wikidata oldalai szerepelnek. Az interjú teljes szövegét is meg lehet itt nyitni olvasásra, ahol a szövegben azonosított névelemek mellett kis információs panel jelenik meg a megfelelő Wikidata, Geonames, VIAF stb. oldalakra mutató linkekkel. Az egyedi vizualizációk készítéséhez egy Kibana<sup>7</sup> felületet is integráltunk a kereső mellé, amellyel számtalan grafikontípusban jeleníthetjük meg a kiválasztott adattartalmat. Az 5. ábrán példaként egy térképes vizualizációt mutatunk be, amely az interjúkban említett földrajzi helyeket prezentálja, illetve megmutatja azt is, hogy mely interjúkban történik az említés.

## Összefoglalás

A rendelkezésre álló nyílt elérésű szoftverekkel viszonylag könnyen lehet gazdag, felfedező, áttekintő jellegű (exploratív) kutatói felületeket létrehozni. Ez azért is szükséges, mert nagy mennyiségű értékes kutatási anyag létezik és keletkezik szöveges vagy hangfelvétel formájában, amit valamilyen előzetes feldolgozás nélkül nem lehet áttekinteni. A projekt során kipróbáltuk a legnépszerűbb és legfrissebb magyar és nyelvfüggetlen nyelvi feldolgozó eszközöket abból a szempontból, hogy mennyire használhatóak a konkrét kutatási célra a feldolgozás eredményei. Ehhez nélkülözhetetlen valamilyen viszonyítási alap, ám ennek előállítása meglehetősen erőforrásigényes. Meg kell jegyezni, hogy a szótövezés és névelem-felismerés magyar nyelven még mindig olyan mennyiségben hibázik, hogy az kutatói környezetben zavaró, és utólagos kézi javítást igényelhet. A másik kulcskérdés az, hogy milyen tárgyszórendszer a legalkalmasabb a tömeges, automatikus kategorizálásra. Ezen a téren egy európai szociológiai taxonómia hazai adaptációja mellett döntöttünk, bár valójában nem volt igazi alternatíva. Előremutató lenne a magyar nyelvű taxonómiákat is összegyűjteni és megfelelően formalizálni (pl. SKOS), hogy segítsünk a jövőben megbirkózni a hasonló kihívásokkal.

**Finanszírozás:** A publikációban bemutatott projektet, amelyet a TK és a SZTAKI valósított meg, az Innovációs és Technológiai Minisztérium és a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal támogatta a Mesterséges Intelligencia Nemzeti Laboratórium keretében.

<sup>7</sup> <https://www.elastic.co/kibana/>



## Bibliográfia

ELSST – European Language Social Science Thesaurus. <https://elsst.cessda.eu/>

Albaugh, Quinn - Sevenans, Julie - Soroka, Stuart - Loewen, Peter John (2013): *The Automated Coding of Policy Agendas: A Dictionary-Based Approach*. Paper presented at the 6th annual Comparative Agendas Project (CAP) conference, Antwerp, June 27–29.

Albaugh, Quinn - Soroka, Stuart - Joly, Jeroen - Loewen, Peter John - Sevenans, Julie - Walgrave, Stefaan (2014): *Comparing and Combining Machine Learning and Dictionary-Based Approaches to Topic Coding*. Paper presented at the 7th annual Comparative Agendas Project (CAP) conference, Konstanz, June 12–14.

Balázs Ágnes – Sebők Miklós (2016): Névelem-felismerés, In.: Sebők M. (szerk): *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*, L'Harmattan, 2016, pp.51–61.

Molnár Csaba (2016): Szövegkódolás a gyakorlatban: kézi, géppel támogatott és gépi megoldások, In.: Sebők M. (szerk): *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*, L'Harmattan, 2016, pp. 24-36.

Suominen, O., Inkinen, J., & Lehtinen, M. (2022). Annif and Finto AI: Developing and Implementing Automated Subject Indexing. *JLIS.It*, 13(1), 265–282. <https://doi.org/10.4403/jlis.it-12740>

Orosz György, Szántó Zsolt, Berkecz Péter, Szabó, Gergő, Farkas Richárd (2022). HuSpaCy: an industrial-strength Hungarian natural language processing toolkit. In XVIII. Magyar Számítógépes Nyelvészeti Konferencia.

Nemeskey Dávid Márk: Egy emBERT próbáló feladat. XVI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2020), Szeged, pp. 409–418.

Nemeskey Dávid Márk (2021): Introducing huBERT. XVII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2021). Szeged, pp. 3–14

Simon, Eszter; Vadász, Noémi. (2021) Introducing NYTK-NerKor, A Gold Standard Hungarian Named Entity Annotated Corpus. In: Ekštejn K., Pártl F., Konopík M. (eds) *Text, Speech, and Dialogue*. TSD 2021. Lecture Notes in Computer Science, vol 12848. Springer, Cham. [https://doi.org/10.1007/978-3-030-83527-9\\_19](https://doi.org/10.1007/978-3-030-83527-9_19)

## Az ELTE Drámakorpuszának létrehozása és lehetőségei

Szemes Botond, Bajzát Tímea, Fellegi Zsófia, Kundráth Péter,  
Horváth Péter, Indig Balázs, Dióssy Anna, Hegedüs Fanni, Pantyelejev Natali,  
Sziráki Sarolta, Vida Bence, Kalmár Balázs, Palkó Gábor  
ELTE BTK Digitális Bölcsészeti Központ  
[palko.gabor@btk.elte.hu](mailto:palko.gabor@btk.elte.hu)

### Absztrakt

Az ELTE DH Drámakorpusz egy folyamatosan bővülő adatbázis, amely a magyar drámairodalom szövegeit teszi elérhetővé és kereshetővé a felhasználók számára. A szövegeket TEI XML formátumba kódolva, részletes annotációval ellátva tesszük közzé, amely formátum nemcsak a szövegek felhasználóbarát megjelenítését teszi lehetővé, hanem egyben az ehhez tartozó keresőfelület létrehozásának is az alapja. Ez utóbbi segítségével a drámák nyelvi-grammatikai elemeinek eloszlását tudjuk kvantitatív alapon vizsgálni, ami a szövegek stilometriai elemzése számára jelent elengedhetetlen kiindulópontot. Hiszen így a drámák vagy akár a szereplők megszólalásainak nyelvi felépítése válik összehasonlíthatóvá mind stílárius, mind tematikus szempontból.

A Drámakorpusz ezen túl részét képezi a DraCor nemzetközi adatbázisnak is HunDraCor néven. Ennek köszönhetően a fenti jellemzőkön túl elérhetők a drámák karakterhálózatának, valamint a jelenetekben megszólalók arányának szintén az annotált változatokból automatikusan felrajzolt vizualizációi is. A karakterhálózatokban az egyes karakterek mint csomópontok, a közöttük lévő interakciók mint élek szerepelnek, ami által – a hálózatelmélet belátásait kamatoztatva – egy átfogó kép adható egy dráma szereplőinek egymáshoz fűződő kapcsolatáról. Ekkor a grammatikai elemek gyakoriságvizsgálatával szemben elsősorban nem a szereplők nyelvi kidolgozottságáról, hanem dramaturgiai funkciójáról tudhatunk meg többet.

### Abstract

The ELTE Drama Corpus is a continuously expanding database that makes texts from Hungarian drama history available and searchable for users. The texts are encoded in TEI XML format with detailed annotation, which not only allows for a user-friendly presentation of the texts, but also provides the basis of the searches. This allows us to analyze the distribution of linguistic/grammatical features quantitatively, which is an essential starting point for stylometric analysis of the plays, i.e. to compare the linguistic structure of the dramas or even of the characters' utterances from both stylistic and thematic point of view.

The Drama Corpus is also part of the DraCor international database. In addition to the above-mentioned features, here you can find visualizations of the character networks and other metrics regarding to the proportion of speakers in the scenes. In the character networks individual characters are represented as nodes and the interactions between them as edges, which provide a comprehensive picture of the interrelationships in a drama. In contrast to the analysis of the frequency of grammatical elements, with the help of these metrics and visualizations the dramaturgical structure of a play could be grasped computationally.

## 1. A korpusz létrehozásának szempontjai, általános jellemzők

Az ELTE *Drámakorpusz* az elmúlt években létrehozott *Verskorpusz* és *Regénykorpusz* mintájára egy folyamatosan bővülő adatbázist és egy hozzá tartozó keresőfelületet foglal magában.<sup>1</sup> Az adatbázis a magyar dráma történetének reprezentatív darabjait gyűjti egybe – egyelőre kizárólag a public domain körébe tartozó alkotásokat. A korpusz jelenlegi gyűjteménye két fázisban jött létre: első lépésként az online, a Magyar Elektronikus Könyvtár felületén is elérhető művekből hoztunk létre TEI XML kódolású fájlokat; második lépésként pedig online nem elérhető szövegek digitalizált és optikai karakterfelismeréssel (OCR) feldolgozott változataiból alkottuk meg a TEI XML fájlokat. Az első fázis során 58 drámát tettünk elérhetővé a *Drámakorpusz* felületén. A válogatást két szempont határozta meg: egyrészt a magyar drámairodalom kanonikus műveit kívántuk egybegyűjteni, másrészt törekedtünk arra, hogy egy szerzőtől több alkotást is tartalmazzon a korpusz, hogy ezáltal ne csak az egyes szövegek és műfajok, hanem drámaírói életművek is összehasonlíthatók legyenek.

A kanonikusság szempontját elsősorban színház- és irodalomtörténeti munkák alapján érvényesítettük, mindenekelőtt Kékesi Kun Árpád *Színházi kalauz* című munkájára és a Gintli Tibor főszerkesztésében 2010-ben megjelent *Magyar irodalom* című kötetre támaszkodva.<sup>2</sup> Elképzelésünk szerint azok a drámák, amelyek ezekben a munkákban megjelennek, kanonikusnak számítanak. Ezt a belátást némileg módosította az online elérhetőség kérdése, amely bizonyos értelemben szintén mint kanonizációs folyamat értelmezhető. Ebben az esetben a számunkra elérhető adatbázisok digitalizációs tevékenysége számít a kanonizáció aktusának, az adatbázis mögötti szempontrendszer pedig a kánon keretrendszerének. Ennek tanulságát fontos a saját projektünkre is levonnunk: a *Drámakorpusz* felülete szintén egyfajta kanonizációs gesztusként is érvényesíthető, amely különösen a korpusz bővítésének második fázisára igaz, amely során olyan szövegeket teszünk online elérhetővé, amelyek mindeddig nem rendelkeztek digitális kiadással. Meg kell továbbá említünk, hogy az online elérhetőség kérdése inkább bővítette és nem szűkítette a színház- és irodalomtörténeteken alapuló címlistát. A korpuszépítés során a Magyar Elektronikus Könyvtár mellett az MTA Irodalomtudományi Intézetének Szövegtárát használtuk a szövegek forrásaiként.<sup>3</sup>

Az első fázis 74 drámája 31 szerző között oszlik el; a művek keletkezési ideje a 16. századtól egészen a 20. századig terjed. A legtöbb, öt vagy hat szöveggel Csiky Gergely, Madách Imre és Jókai Mór, Hunyady László szerepel, míg Babits Mihály, Balázs Béla, Szigligeti Ede és Vörösmarty Mihály négy drámája érhető el a felületen. Mindez összesen 1 141 490 db

---

1 HORVÁTH Péter, „Az ELTE *Verskorpusz* automatikus annotációs eljárásai révén nyerhető kvantitatív adattípusok” in *Nyelvtan, diskurzus, megismerés*, szerk. SIMON Gábor, és TOLCSVAI NAGY Gábor, 313–331 (Budapest: Eötvös Kiadó, Budapest, 2020).

HORVÁTH Péter, KUNDRÁTH Péter, INDIG Balázs, FELLEGI Zsófia, SZLÁVICH Eszter, BAJZÁT Tímea Borbála, SÁRKÖZI-LINDNER Zsófia, VIDA Bence, KARABULUT Aslihan, TIMÁRI Mária, PALKÓ Gábor, „*ELTE Verskorpusz: a magyar kanonikus költészet gépileg annotált adatbázisa*”, in *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, 375–388 (Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2022).

BAJZÁT Tímea, SZEMES Botond, SZLÁVICH Eszter, „Az ELTE DH *Regénykorpusz* és lehetőségei”, in *Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6-9.*, szerk. TÍCK József, KOKAS Károly és HOLL András, 63–72 (Budapest: HUNGARNET Egyesület, 2021).

2 KÉKESI KUN ÁRPÁD, szerk., *Színházi kalauz* (Budapest: Saxum, 2008).

GINTLI TIBOR, szerk., *Magyar Irodalom*, (Budapest: Akadémiai Kiadó, 2010).

3 KECSKEMÉTI Gábor, MÉSZÁROS Tamás, BUCSICS Katalin, KISS Margit és MARKÓ Veronika, *Nemzeti klasszikusok kritikai kiadásai – A BTK Irodalomtudományi Intézet textológiai portálja*, v. 1.0 (2021. január 1.), szolgáltatja a BTK Irodalomtudományi Intézet, <https://szovegtar.iti.mta.hu/>. (Utolsó elérés: 2022. 06. 06.)



szóelőfordulást (token) jelent, valamint 871 542 lemmát.<sup>4</sup> A második fázis során az ELTE Egyetemi Könyvtárban elérhető kötetek közül válogattunk a fent bemutatott szempontok szerint, majd a Könyvtár munkatársai által készített szkennelt fájlok OCR-ezését és TEI XML kódolását végeztük el. Az OCR-ezést és a két fázis kódolási munkálatait az ELTE BTK hallgatói végezték el. A második fázis bővítése egyaránt vonatkozik drámaírók gyűjteményes kötetekre (ilyen például Bessenyei György és Hunyadi Sándor), valamint válogatáskötetekre, amelyekben több szerző művei találhatók.

Fontos továbbá kiemelni, hogy az ELTE BTK *Drámakorpusza* egyben a DraCor nemzetközi adatbázisának magyar nyelvű alkorpuszát is képezi HunDraCor néven. A DraCor felülete (<https://dracor.org/>) nem egyszerűen drámák gyűjteményét jelenti, hanem a szövegek elemzéséhez fontos eszközöket is kínál. Ezek közül is kiemelkedik a drámákhoz tartozó karakterhálózatok automatikus felrajzolása, amely hálózatokban az egy jelenetben szereplő karakterek mint csomópontok között húzódik él (ennek vastagsága a közös jelenetekben való előfordulások száma szerint alakul). Az ilyen karakterhálózatok az olvasás vagy az előadás nézésének szükségszerűen időben lezajló folyamatához képest szimultán teszik átláthatóvá a szereplők közötti viszonyrendszereket, az összetartozó csoportokat és az egyes figurák dramaturgiai funkcióit. Mindezen túl lehetőséget biztosítanak arra is, hogy hálózatelméleti mérőszámokkal írjuk le a felrajzolt ábrát (például milyen sűrű az adott hálózat), ami így a szövegek felépítésének számszerű összehasonlítására teremthet alapot.<sup>5</sup> A *Drámakorpusz* felületére készített TEI XML kódok megfelelnek a DraCor által érvényesített specifikációnak, így a két korpusz egységesítése probléma nélkül megtehető.

## 2. A drámaszövegek jelölőnyelvi kódolása

A korpuszba kerülő szövegek elsődleges forrásai tehát a Magyar Elektronikus Könyvtár adatbázisának szabadon hozzáférhető (public domain) magyar drámaszövegei voltak. Az elérhető fájlformátumok közül elsődlegesen az RTF-kiterjesztésű fájlokat preferáltuk a korpusz létrehozásakor, de azokban az esetekben, amikor ezek elérésére nem volt lehetőségünk, a HTML-formátumot választottuk, illetve némely mű esetében a kétrétegű PDF fájl volt az egyetlen elérhető forrás a jelölőnyelvi kódolást végző annotátorok számára.

A drámakorpusz szövegeinek a kódolási formátuma a TEI XML jelölőnyelv<sup>6</sup> szabványain alapuló és annak eleget tevő specifikáció, amely egyaránt megfelel a DraCor formai követelményének. A TEI XML egyik előnye, hogy eszköz- és rendszerfüggetlen, továbbá a jelölőnyelvi kódolással ellátott szövegek együttesen tárolhatók a metaadatokkal,<sup>7</sup> valamint lehetővé teszi a szövegek olyan strukturálását és annotációját, amelyek hozzájárulnak ahhoz, hogy a szövegek bizonyos nyelvi, tartalmi és metaadatszintű kereshetősége megvalósulhasson.

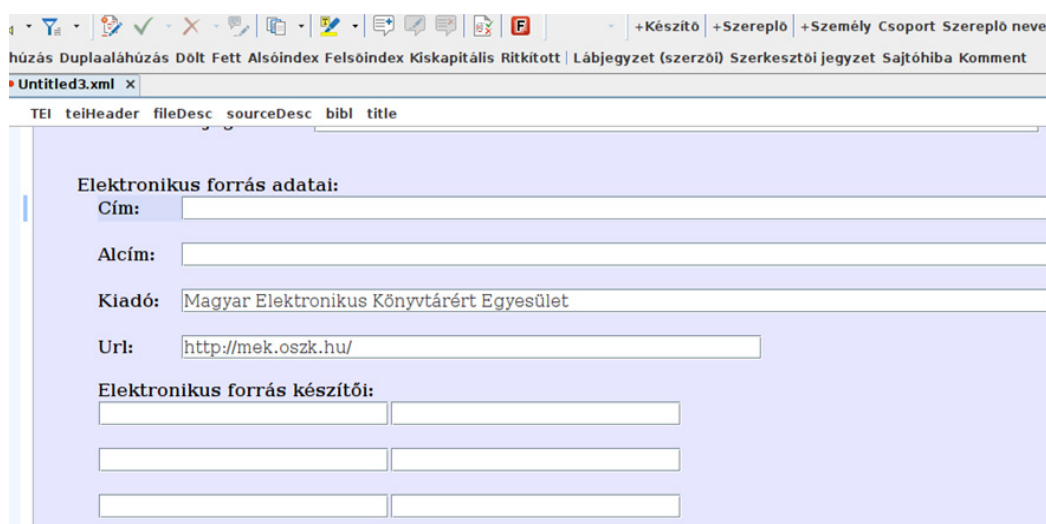
4 A tokenizálás a számítógépes szövegfeldolgozás fontos előkészítő lépése, amely során a szöveget a későbbi strukturálás alapjául vett tokenekre, egységekre bontjuk. A drámakorpusz esetében a token szóalakokat jelöl.

5 Frank FISCHER, Ingo BÖRNER, Mathias GÖBEL, ANGELIKA HECHTL, CHRISTOPHER KITTEL, CARSTEN MILLING, PEER TRILICKE, „Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama”, in *Proceedings of DH2019: „Complexities”*, 1–6 (Utrecht: Utrecht University, 2019) [doi:10.5281/zenodo.4284002](https://doi.org/10.5281/zenodo.4284002).

6 TEI Consortium, eds., *TEI P5: Guidelines for Electronic Text Encoding and Interchange* <http://www.tei-c.org/Guidelines/P5/> (Utolsó elérés: 2022. 06. 06.).

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

A jelölőnyelvi kódolás megkönnyítése, illetve a kódolás során potenciálisan fellépő szintaktikai és validációs hibák redukálása érdekében olyan digitális környezetet készítettünk, amely a kódolási munkafázis során vizualizációjában elrejtette az XML jelölőnyelvi struktúrát, és a szabályrendszernek megfelelő címkéket kattintással lehetett a megfelelő szöveghelyekhez rendelni. A <teiHeader> – a drámaszövegek metaadatait – tartalmazó fejlécként funkcionáló kódrészlet pedig adatlapszerűen jelent meg, ahol a megfelelő szövegdobozokat kitöltve lehetett felvinni a művek metaadatait. A kódolást az Oxygen XML Editor<sup>8</sup> programmal végeztük, a „kódmentes” munkakörnyezet elkészítésére is ez a program adott teret, Oxygen Framework formában. A framework egyszerre érvényesíti az általunk írt specifikációnak a validálási szabályait, továbbá felel a kódolási felület vizualizációjáért. A specifikációnak megfelelő validálást egy DTD-formátumú dokumentum végzi el, a fent részletezett elvek alapján. A címkékbe kerülő sztringek (például dátumok) jól formáltságát Schematron ellenőrzi, és hiba esetén üzenetet küld az annotátornak. A vizuális megjelenítés mögött egy CSS-fájl található. A címkék ikonszerű megjelenítését és a munkakörnyezet beállításait egy framework-kiterjesztésű dokumentum tárolja, és ennek szerkesztése ad lehetőséget a munkakörnyezet finomhangolására.



1. ábra. Részlet az Oxygen Framework által biztosított munkakörnyezetről.

Annak érdekében, hogy az annotált drámakorpuszon grammatikai tulajdonságok alapján keresést végezhessünk, szükséges volt a szöveg lemmatizálása, morfológiai, valamint szófaji elemzése. Ehhez az MTA Nyelvtudományi Intézete által fejlesztett e-magyar automatikus nyelvi elemzőrendszer<sup>9</sup> alkalmaztuk, úgy, ahogyan a bevezetőben említett *Verskorpusz* és *Regénykorpusz* esetében is: az e-magyart futtató szkript a drámák szerkezeti annotációit tartalmazó TEI XML fájlokból kieszte a szöveget, lefuttatta rajtuk az e-magyart, majd a grammatikai annotációkat belerakta a TEI XML fájlokba.

8 <https://www.oxygenxml.com> (Utolsó elérés: 2022. 06. 06.)

9 INDIG Balázs, SASS Bálint, SIMON Eszter, MITTELHOLCZ Iván, KUNDRÁTH Péter és VADÁSZ Noémi, „emtsv – egy formátum mind felett”, in *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. BEREND Gábor, GOSZTOLYA Gábor, VINCZE Veronika, 235–247 (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2019). MITTELHOLCZ Iván, „emToken: Unicode-képes tokenizáló magyar nyelvre”, in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. VINCZE Veronika, 61–69 (Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2017).



### 3. A lekérdezőfelület

Az ELTE drámakorpuszhoz elérhető egy nyilvános online lekérdezőfelület is, amely funkcióival és arculatával illeszkedik az ELTE Digitális Bölcsészet Tanszék többi szolgáltatása közé. A kereső részletes leírásáról a *Súgó* menüpontban lehet tájékozódni. A keresőfelület elérhető angol nyelven is.

A drámakorpusz esetében elérhető részletes keresőfelület funkciói lehetővé teszik, hogy létrehozzunk alkorpuszokat az adatbázisban található metaadatok alapján. Kereshetünk tokenekre és tokenkapcsolatokra, illetve megválaszthatjuk az adatok feldolgozási és megjelenítési módjait. A *Regénykorpusz*hoz és a *Verskorpusz*hoz képest a *Drámakorpusz* specifikuma az, hogy létrehozhatunk alkorpuszokat a mű szereplőinek megszólalásaiból, de akár a szereplők neve alapján is szűkíthetjük a találatokat.

Keresés ideje: 0,00s  
Összesen: 58 dráma (mutatva: 1-20 dráma)

2. ábra. Az ELTE drámakorpusz keresőfelülete

Ahogy az a lekérdezőfelületet bemutató 2. ábrából is látható, szerzők és műcímek megadásával szűkíthetjük a vizsgálati korpuszt. Továbbá megadhatunk időintervallumot a drámák keletkezési idejére vonatkozóan.

A tokenekre, illetve tokenkapcsolatokra való szűrés lehetővé teszi konkrét szóalakok vagy szótövek keresését. Nem csupán szóalakokra és szótövekre kereshetünk, hanem szófaji és morfológiai kategóriákra is, aminek révén konstrukciós mintázatok is adatolhatók. Több tokenre való keresés esetén megadható, hogy a tokenek milyen típusú szövegegységben belül, illetve egymástól maximálisan mekkora távolságra forduljanak elő.

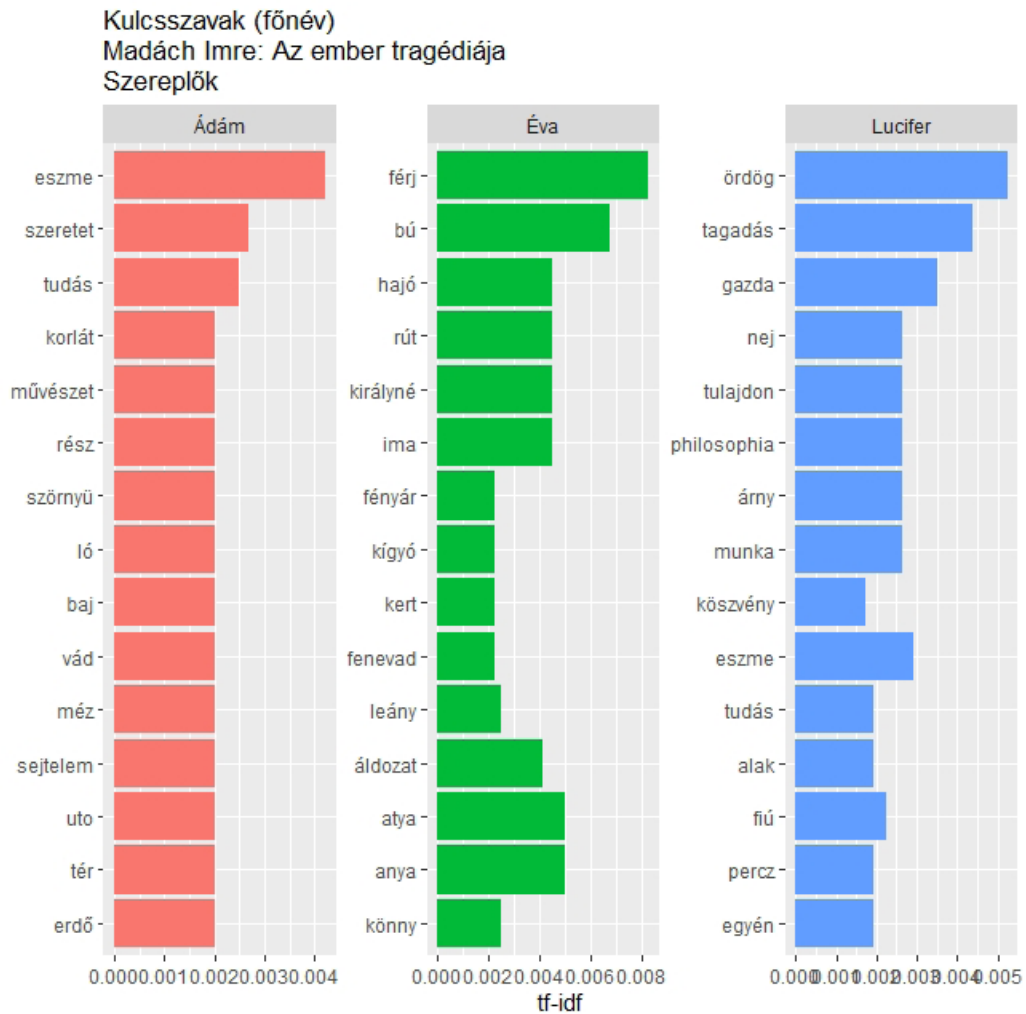
A találatok megjelenítése többféle módon lehetséges, érdemes azt az opciót kiválasztani, amely a leginkább garantálja az adatok előkészítését a további feldolgozásokhoz. Egyrészt megadható, hogy a találatokat mekkora kontextusban szeretnénk lekérni, illetve tetszőlegesen választható, hogy hány találatot 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 előállított listák TSV-formátumban tölthetők le.

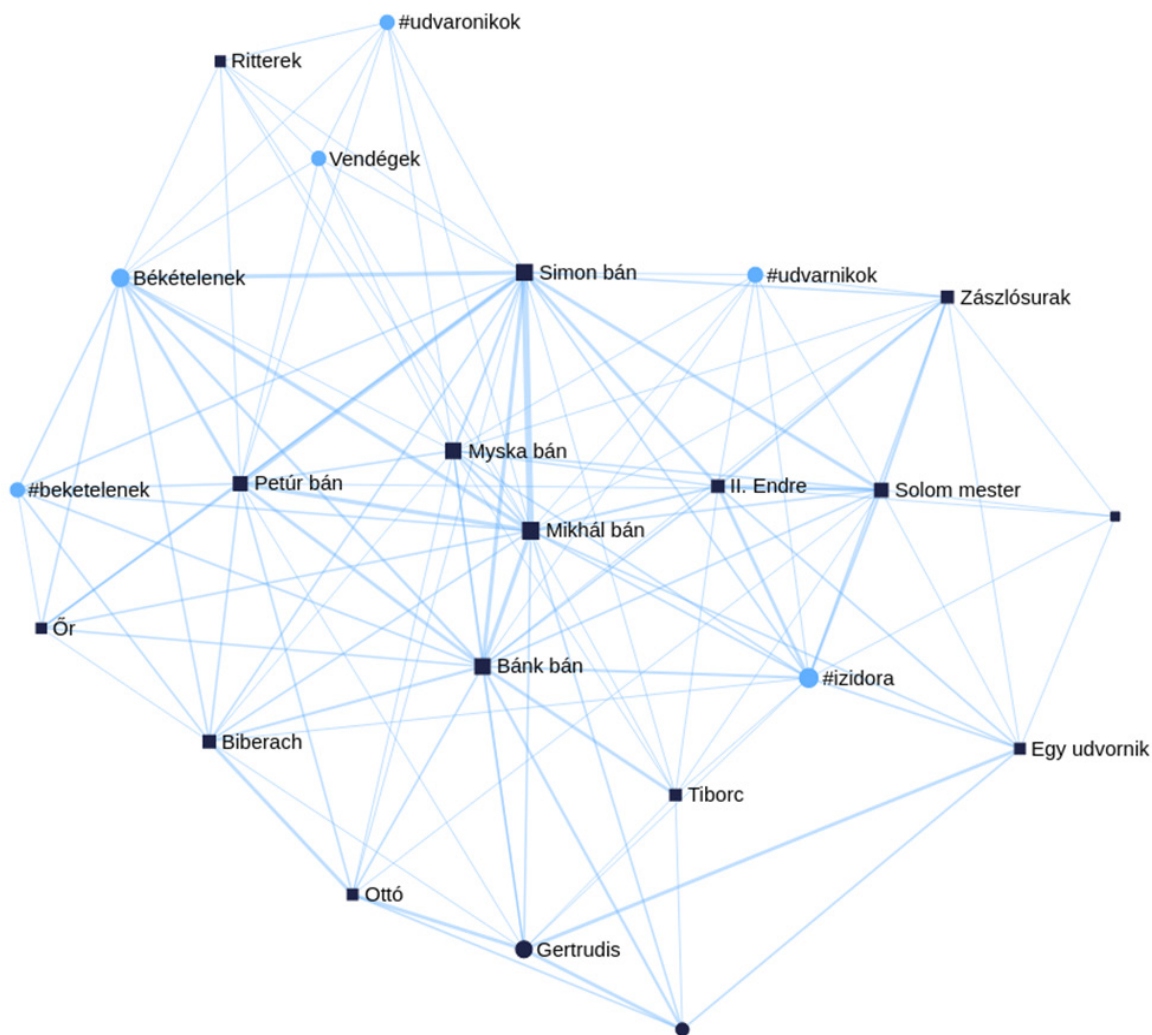
#### 4. Lehetséges felhasználás

Azáltal, hogy nem csupán a drámák szövegét, hanem azok nyelvi elemzővel annotált változatát tárolja az adatbázis, számos, a drámák szövegére vonatkozó keresést egyszerűbben végezhetünk el, mint korábban, illetve olyan módon is kereshetünk, amelyre nem lenne lehetőség abban az esetben, ha a korpusz csupán a drámaszövegeket tartalmazná. Például ha kíváncsiak vagyunk arra, hogy az „ördög” szó mely művekben, milyen szövegkörnyezetben fordul elő, akkor nem kell külön rákeresnünk a szó összes alakjára (*ördög*, *ördögök*, *ördögöt*, *ördögöket* stb.), hanem a szavak szótári alakjának az adatbázisban való szerepeltetése révén egy kereséssel listázhatjuk az összes ilyen szöveghelyet. Mivel a szavak szófaját és morfoszintaktikai jellemzőit is tartalmazza az adatbázis, nem csupán szavakra, hanem grammatikai jellemzőkre is kereshetünk. Például rákereshetünk az összes olyan szöveghelyre, amely a melléknév + *ördög* szerkezetet tartalmazza, de akár a középfokú melléknév + *ördög* szerkezeteket is listázhatjuk. Ezek a keresések elsősorban olyan gyakoriságvizsgálatok számára teremtenek alapot, amelyekeken keresztül az egyes művek tematikus és stíláriis szerveződésére is következtetni lehet (például a leggyakrabban használt főnevek listája), és ezáltal a drámák vagy a drámatörténet kvantitatív szempontú megértéséhez járulhatnak hozzá. A korábbi korpuszokhoz képest a *Drámakorpusz* sajátossága, hogy nemcsak teljes szövegek, hanem az egyes szereplők nyelvének összehasonlítását is lehetővé teszi. A 3. ábrán *Az ember tragédiája* három szereplőjének a főnevek közül kikerülő kulcsszavait mutatjuk be példaként.



3. ábra: Az ember tragédiája három szereplőjének a főnevek közül kikerülő kulcsszavai a tf-idf módszer alapján

A DraCor korábban említett adatbázisa mindezek mellett más típusú információkat is elérhetővé tesz. A DraCor használatával elsősorban ugyanis nem a szereplők nyelvi kidolgozottságáról, hanem dramaturgiai funkciójáról tudhatunk meg többet: mennyire tekinthető egy karakter központi szereplőnek, hány másik szereplővel tartja a kapcsolatot, mennyiben nélkülözhető a hálózat stabilitásának szempontjából stb. A jelenetekben megszólalók száma a dráma felépítéséről (például tömegjelenetek a dráma elején és/vagy végén) tanúskodik, illetve a monológok, dialógok és csoportos jelenetek arányát mutatja. A 4. ábra a Bánk bán karakterhálózatát mutatja, amelyet a Gephi vizualizációs szoftverben hoztunk létre a DraCor adatai alapján.



4. ábra: A Bánk bán karakterhálózata a DraCor felületén<sup>10</sup>

## 5. Függelék

Az ELTE Drámakorpusz lekérdezőfelülete elérhető: <https://dramakorpusz.elte-dh.hu/>

A drámák kódolt szövegei elérhetők: <https://github.com/ELTE-DH/drama-corpus>

## 6. Bibliográfia

BAJZÁT Tímea, SZEMES Botond és SZLÁVICH Eszter. „Az ELTE DH Regénykorpusz és lehetőségei”, In *Online térben az online térért: Networkshop 30: országos online konferencia. 2021. április 6-9.*, szerkesztette TICK József, KOKAS Károly és HOLL András, 63–72. Budapest: HUNGARNET Egyesület, 2021. <https://doi.org/10.31915/NWS.2021.7>

FISCHER, Frank, BÖRNER, Ingo, GÖBEL, Mathias, HECHTL, Angelika, KITTEL, Christopher, MILLING, Carsten and TRILCKE, Peer, „Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama”, In *Proceedings of DH2019: „Complexities”*, 1–6. Utrecht: Utrecht University, 2019.

<sup>10</sup> <https://dracor.org/hun/katona-bank-ban>



- HORVÁTH Péter, „Az ELTE Verskorpusz automatikus annotációs eljárásai révén nyerhető kvantitatív adattípusok”, In *Nyelvtan, diskurzus, megismerés*, szerkesztette SIMON Gábor és TOLCSVAI NAGY Gábor, 313–331. Budapest: Eötvös Kiadó, 2020.
- HORVÁTH Péter, KUNDRÁTH Péter, INDIG Balázs, FELLEGI Zsófia, SZLÁVICH Eszter, BAJZÁT Tímea Borbála, SÁRKÖZI-LINDNER Zsófia, VIDA Bence, KARABULUT Aslihan, TIMÁRI Mária és PALKÓ Gábor, „ELTE Verskorpusz: a magyar kanonikus költészet gépileg annotált adatbázisa”, In *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette BEREND Gábor, GOSZTOLYA Gábor és VINCZE Veronika, 375–388. Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2022.
- KECSKEMÉTI Gábor, MÉSZÁROS Tamás, BUCSICS Katalin, KISS Margit és MARKÓ Veronika, *Nemzeti klasszikusok kritikai kiadásai – A BTK Irodalomtudományi Intézet textológiai portálja*, v. 1.0 (2021. január 1.), szolgáltatja a BTK Irodalomtudományi Intézet, <https://szovegtar.iti.mta.hu/>. (Utolsó elérés: 2022. 06. 06.)
- INDIG Balázs, SASS Bálint, SIMON Eszter, MITTELHOLCZ Iván, KUNDRÁTH Péter és VADÁSZ Noémi, „emtsv – egy formátum mind felett”, In *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette BEREND Gábor, GOSZTOLYA Gábor és 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*, szerkesztette BODA István és MÓNOS Katalin, 65–72. Debrecen: MANYE, Debreceni Egyetem, 2011.
- GINTLI Tibor, szerk. *Magyar Irodalom*, Budapest: Akadémiai Kiadó, 2010.
- 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.
- KÉKESI KUN András, szerk. *Színházi kalauz*, Budapest: Saxum, 2008.
- TEI Consortium, *TEI P5*, eds. *Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (Utolsó elérés: 2022. 06. 06.)

## Az ELTEdata szemantikus adatbázis legújabb fejlesztései

Sebestyén Ádám

ELTE BTK TI Digitális Bölcsészeti Tanszék, Digitális Örökség Nemzeti Laboratórium  
[sebestyen.adam@btk.elte.hu](mailto:sebestyen.adam@btk.elte.hu)

A 2021-ben rendezett Networkshop-konferencián mutattam be először az ELTE Digitális Bölcsészeti Tanszéken fejlesztett, prozopográfiai, bibliográfiai és egyéb történeti tárgyú kutatások anyagait feldolgozó szemantikus adatbázist, az ELTEdatát.<sup>1</sup> Az akkori előadás jórészt a projekt ismertetésére, illetve az adatbázis felépítésének és működésének áttekintésére vállalkozott. Jelen tanulmány az újabb fejlesztések és az eddigi eredmények összegzésén túl azt vizsgálja, hogy a szemantikus adatfeldolgozás milyen egyedi mintázatok felismerését, milyen struktúrák és vizualizációk kialakítását teszi lehetővé.

A történelmi adatok tárolását és feldolgozását szolgáló FactGrid-platformhoz hasonlóan<sup>2</sup> az ELTEdata is wikibase-szoftveren alapul, és a Wikidata adatstruktúráját követi, azzal a fontos különbséggel, hogy az ELTEdata mind a szemantikus állítások, mind az entitások szintjén össze van kapcsolva a Wikidata megfelelő állításaival. Ennek módja egy külső azonosító adattípusú tulajdonság, mely az adott tételt (item) tetszőleges névterekhez és repozitóriumokhoz (mint például a Wikidata vagy az MTMT) köti. A szemantikus adatmodellezés nyelvét, az RDF-et követve, az adatbázis egyedi azonosítóval rendelkező elemekből épül fel, ahol minden szemantikus kijelentés tulajdonság (*property*) és érték (*value*) kettőséből tevődik össze. Fontos kiemelni, hogy a szolgáltatás nem korlátozódik prozopográfiai adatok tárolására és feldolgozására: lehetőség van többek közt entitásazonosításra a már digitalizált forrásszövegekben, de a bibliográfiaépítés ugyancsak megvalósítható.

Az ELTEdata jelenleg három kutatócsoport adatait tartalmazza: nagyságrendileg 10000 item (személy- és földrajzi nevek, intézmények, műcímek és egyéb bibliográfiai adatok), valamint 130 tulajdonság szerepel az adatbázisban.<sup>3</sup> A prozopográfiaiak mellett, mint azt már említettem, entitásazonosításon keresztül megoldható az adatbázis tételeinek hozzákapcsolása digitalizált forrásszövegekhez. A Tudásáramlás anyagaiban például nemcsak az útinaplók, illetve útleírások személy- és földrajzi neveit, hanem egy előre kialakított hierarchia alapján a fogalmakat is címkézni (tag-elni) lehet. Ezek a fogalmak többnyire olyan szövegrészeket jelölnek, melyek egy egyértelműen definiálható technológiai, politikai, történelmi vagy kulturális képzetet tárgyalnak. (Ígypéldául, ha a szöveg egy szakasza mezőgazdasági kérdéseket elemez, akkor a mezőgazdaság, illetve technológia fogalmakat lehet hozzákapcsolni.) Mivel a legjobban feldolgozott gyűjtemény a HECEdata, (melynek forrása a 3. lábjegyzetben említett humanista szerzői lexikon), a továbbiakban az adatleképezési és vizualizációs példákat is ebből a gyűjteményből idézem.

1 [https://eltedata.elte-dh.hu/wiki/Main\\_Page](https://eltedata.elte-dh.hu/wiki/Main_Page) (Hozzáférés: 2022. június 9.) A felület az utóbbi időszakban többször is megújult: részben a hibák kijavítása, részben új, kiegészítő modulok telepítése miatt.

2 [https://database.factgrid.de/wiki/Main\\_Page](https://database.factgrid.de/wiki/Main_Page) (Hozzáférés: 2022. június 9.)

3 A Humanizmus Kelet-Közép-Európában kutatócsoport (<https://hece.elte.hu/>) az 1420 és 1620 között a Magyar Királyság területén született irodalmi műveket és szerzőik értelmiségi karriermintázatait vizsgálja egy, a német *Verfasserlexikon* mintáját követő lexikon formájában. A második nagyobb forrásanyag a Tudásáramlás-projekt, (<http://tudasaramlas.btk.elte.hu/hu/>), mely a Kurucz György *Technológiai utazás a modern kor hajnalán* című forráskiadásában előforduló személy- és földrajzi neveket, illetve a digitalizált forrásszövegeket teszi közzé. A harmadik prozopográfia a két világháború közötti egyetemi tanárok biográfiai gyűjteménye, az ELTEdata (<https://tatk.elte.hu/kutatokozpontok/prozopografiai>). A projekt, lévén, hogy több generációról is közöl adatokat, kiválóan alkalmas többek közt a társadalmi mobilitás vizsgálatára.



A lexikonban a szócikkek adott szerző biográfiájából, műveik ismertetéséből (külön bekezdést szentelve azok kiadástörténetének) és egy bibliográfiai részből épülnek fel.<sup>4</sup> Mivel a biográfia kellően kidolgozott, lehetséges az alapadatokon túl komplexebb struktúrák szemantikus leképezése, úgymint pozíciók, azok megszerzésének ideje, a kinevező személy, patrónus neve, munkakörök helyszíne, vagy egy életút jellegzetes eseményei (ami lehet külföldi utazás, követjárás, peregrináció, zsinatokon, koronázásokon vagy háborúban való részvétel). Ezek között az elemek között hierarchiát is ki lehet alakítani, ami azt jelenti, hogy egy *subclass of* (P27) tulajdonság (property) segítségével érzékeltetni lehet az elemek közti alá-fölérendeltségi viszonyt. Mivel a HECE-lexikon adatai nem állnak rendelkezésre strukturált formában, manuális adatbevitelre volt szükség, ami némileg lassította a munkafolyamatot, még akkor is, ha ily módon bizonyos hibákat (például duplumok keletkezését) el lehetett kerülni. A félautomatikus adatbetöltésre irányuló kísérletek egyike volt a Tudásáramlás-projekt földrajzi neveinek bevitele egy Wikidata-eszköz, a QuickStatements segítségével.<sup>5</sup> Bár a betöltés hosszas kísérletezés után sikeresnek bizonyult, a formátum körülményes volta miatt egyszerűbbnek tűnik a későbbiekre nézve az API-n keresztül történő betöltés egy Pythonban írt kód segítségével. Az utóbbi időszak ennek a kódnak az elkészítéséről szolt, mely jelenleg tesztelés alatt áll. Segítségével a közeljövőben az elitkutatás Excel-táblázatban tárolt prozopográfiai anyagai kerülhetnek fel az ELTEdata felületére.

A félautomatikus betöltés a bibliográfiaépítésben is hasznosnak bizonyulhat, hiszen az olyan névterek, mint az MTMT vagy a PIM adatait importálni lehetne az ELTEdata-ra. A humanista szerzői lexikonból eddig felvitt bibliográfiai tételek külső azonosítók (mint az ISBN, ISSN vagy OCLC) segítségével lettek hozzákapcsolva névterekhez, könyvtári katalógusokhoz, illetve egyéb gyűjteményekhez, vagyis a bibliográfiaépítésnek is fontos része az adatok szemantikus leképezése, azok összekapcsolása más adatbázisokkal. Annak ellenére, hogy a szakirodalmi tételek felsorolására önálló property lett létrehozva, az is megoldható, hogy már az állításokban hivatkozzuk a forrásokat. A HECEdata esetében azt az elvet követtem, hogy az állításokban a nem kifejezetten a szekunder irodalomhoz sorolható tételeket tüntettem fel: egyetemi anyakönyvek, forráskiadások, lexikonok, enciklopédiák és repozitóriumok jelzik olyan adatok lelőhelyét, mint például egy szerző egyetemi tanulmányai.

educated at	University of Heidelberg	edit
start time	22 January 1597 <i>Gregorian</i>	
▼ 2 references		
stated in	Die Matrikel der Universität Heidelberg	
page(s)	187	
stated in	Hungarian Students on the German Universities and Academies between 1526 and 1700	
page(s)	123	
+ add reference		

1. ábra Hivatkozások az egyes állításokban

4 Egy mintaszócikkhez lásd az alábbi közleményt: MOLNÁR Dávid, „Szenci Molnár Albert. Mintaszócikk egy készülő enciklopédiából”, *Magyar Könyvszemle*, 135, 1. sz. (2019), 50-90.

5 Erről az eszközről lásd az alábbi útmutatót: <https://www.wikidata.org/wiki/Help:QuickStatements> (Hozzáférés: 2022. június 9.)



Mivel az ELTEdata komplex keresések hajthatók végre egy szemantikus adatbázisokra kidolgozott RDF lekérdező nyelv, a SPARQL segítségével, lehetőség van tetszőleges paraméterek beállításával keresni az adatkészletben, illetve vizualizálni az így kapott eredményeket.<sup>6</sup> Az ELTEdata SPARQL-végpontja megjelenését és funkcióit tekintve is nagyban hasonlít a Wikidata-nál használt Wikidata Query Service (WDQS) felületére. A SPARQL használatáról alapos útmutatók állnak rendelkezésre,<sup>7</sup> itt elegendő csupán annyit megjegyezni, hogy a keresés során megadandó paraméterek a wikibase-alapú szemantikus leképezésnek megfelelően a *triple*-formátum szintaxisát követik. A lekérdezés során nyert találatok különböző formátumokban letölthetőek (JSON, CSV, TSV), illetve beágyazhatóak. Az esetek nagy részében a lekérdezések a paraméterek alapján szűkítenek és válogatnak az adatkészletben, lehetőség van azonban egyetlen item valamennyi állításának a lekérdezésére is (lásd a 2. ábrát). A SPARQL alapértelmezett vizualizációi lehetővé teszik a különböző adattípusoknak megfelelő leképezést. Így például a gráf kapcsolati hálók feltüntetésére a legalkalmasabb, míg az idővonal egy személy életútját vagy irodalmi művek megjelenését kíséri végig, továbbá arra a kérdésre is választ ad, hogy mely nyomdákban és mikor jelentek meg a tárgyalt szövegek. Az olyan kvantitatív jellegű adatok, mint például hogy hányan tartoznak egy adott felekezetbe, jól szemléltethető buborékdiagramon.

```

1 SELECT ?propUrl ?propLabel ?valUrl ?valLabel
2 WHERE
3 {
4   hint:Query hint:optimizer 'None' .
5   { BIND(wd:Q443 AS ?valUrl) .
6     BIND("N/A" AS ?propUrl) .
7     BIND("identity"@en AS ?propLabel) .
8   }
9   UNION
10  { wd:Q443 ?propUrl ?valUrl .
11    ?property ?ref ?propUrl .
12    ?property rdf:type wikibase:Property .
13    ?property rdfs:label ?propLabel
14  }
15
16  ?valUrl rdfs:label ?valLabel
17  FILTER (LANG(?valLabel) = 'en') .
18  FILTER (lang(?propLabel) = 'en' )
19 }
20 ORDER BY ?propUrl ?valUrl
21
22
23

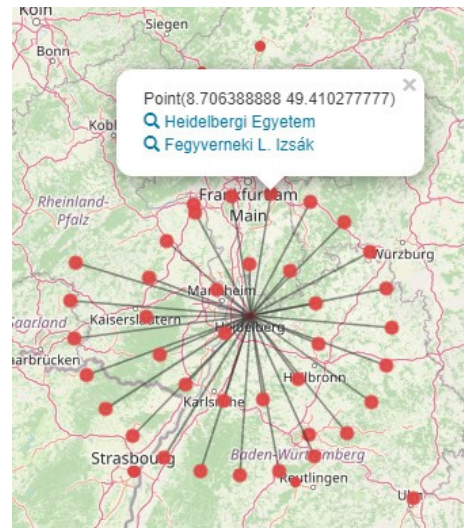
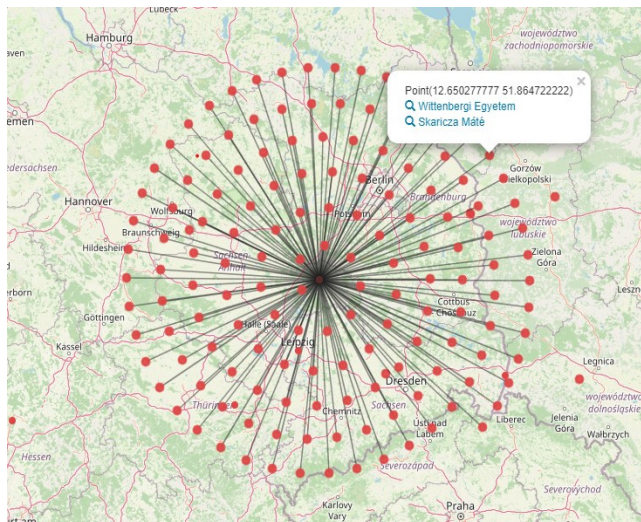
```

2. ábra. Egyetlen item állításainak lekérdezése

A földrajzi helyeket jelölő itemeknél lehetőség van megadni azok koordinátáit is, így bizonyos lekérdezések eredményei térképre vetítve is megjeleníthetőek. A 3. ábrán látható példák ennek segítségével képeznek le peregrinációs célpontokat, amiből kiderül, hogy mely oktatási intézmények voltak a legnépszerűbbek a késő 15. századtól az 1600-as évek első feléig tartó időintervallumban. (Az idézett ábrán a wittenbergi és a heidelbergi egyetemek látogatottsága szerepel.) Hasonlóképp lehet térképre vetíteni a születési helyeket, amiket összevetve adott szerző tanulmányainak, illetve későbbi működésének helyszíneivel, akár a társadalmi mobilitás meglétére is következtetni lehet (mindez az elitkutatás anyagain lesz igazán szemléletes).

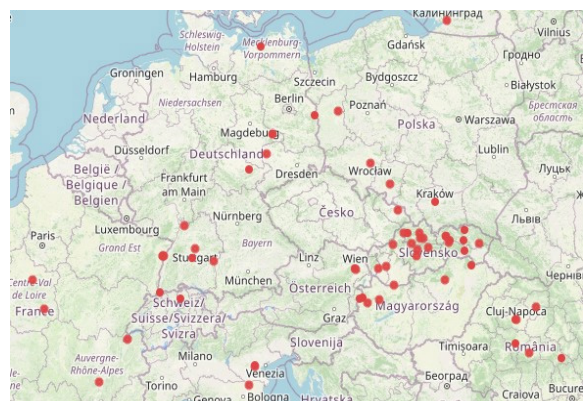
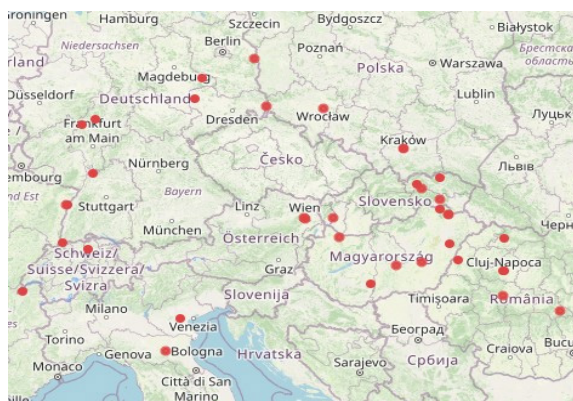
6 Az ELTEdata SPARQL-végpontja a következő: <https://query.elte-dh.hu/> (Hozzáférés: 2022. június 9.)

7 Mindenekelőtt ez a segédanyag: [https://www.wikidata.org/wiki/Wikidata:SPARQL\\_tutorial](https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial) (Hozzáférés: 2022. június 9.)



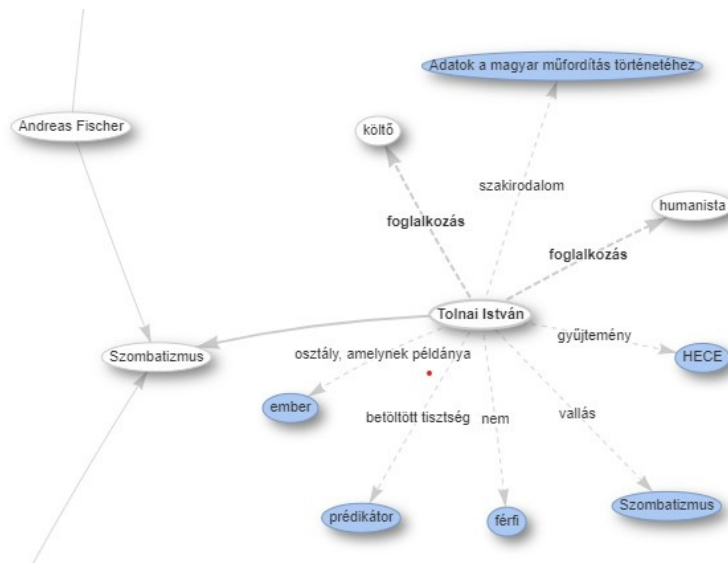
3. ábra. Peregrinációs célpontok

A vizualizációk nagyban hozzásegítenek ahhoz, hogy következtetéseket vonjunk le egy csoportra vonatkozóan azok életrajzi adataiból. A tanulmányok helyszíneinek felekezeti hovatartozás alapján történő megoszlása például kiválóan szemléltethető. A soron következő példákban négy felekezetre kérdeztem le az adatokat (katolikus, református, evangélikus és unitárius). A legnagyobb felekezetekre (katolikus és református) nagyságrendileg 200 találat jutott, míg evangélikusra 161, unitáriusra pedig 63. Fontos megjegyezni, hogy adott személyhez több felekezet is tartozhat, tekintetbe véve az áttéréseket, ám a személyek döntő többségénél csupán egy releváns felekezet szerepel. A katolikusoknál született találatok meglehetősen szórtaak, sűrűbb pontok leginkább az itáliai városoknál fordulnak elő. Az unitárius felekezetről már egyértelműen kirajzolódik, hogy népszerűek voltak körükben az erdélyi városok, ahol maga az antitrinitárius irányzat is maradandó felekezetté tudott szilárdulni. A reformátusok és evangélikusok eredményeit összevetve talán a legfeltűnőbb, hogy utóbbi felekezet tagjai nagy számban tanultak felső-magyarországi városokban, melyek jellemzően németajkú lakossága körében magas volt a lutheri tanok támogatottsága. Ez közvetett módon azt is mutatja, hogy az ország mely részén szilárdult meg a lutheri, illetve a kálvini reformáció. Ráközelítve a térképre jól kirajzolódnak a 16. században alakult népszerű református oktatási intézmények is, mindenekelőtt a debreceni, illetve a tolnai kollégiumok.



4. ábra. Református és evangélikus peregrinációs célpontok vizualizációja

A SPARQL beépített vizualizációi közül egy másik lehetőség a gráfok alkalmazása. Maradva a felekezeti megoszlásnál, nagyon jól lehet szemléltetni, hogy a lexikon szerzői mely vallási irányzatot követték. Önálló csoportban, protestánsként lehet feltüntetni azokat a személyeket, akiknél a források tanúsága alapján nem egyértelmű, hogy még a lutheri reformáció követői voltak, vagy már a helvét irányzaté. Az olyan kisebb irányzatok, mint a zwingliánus vagy a kriptokálvinista, jól láthatóan nem képeznek számottevő csoportot, de hasonlóan kevés taggal képviselteti magát a szombatos, illetve az anabaptista felekezet is. Ráközelítve a gráfra (ahogy az 5. ábrán megfigyelhető) kirajzolódnak az egyes személyeket takaró csomópontok, akiknek a nevére kattintva kibomlanak a leképezett adataik is: a foglalkozás, vallás és betöltött pozíció mellett az adott szerzőhöz tartozó esetleges szakirodalmi tételek is megjeleníthetők.



5. ábra Egy humanista adatai gráfon leképezve

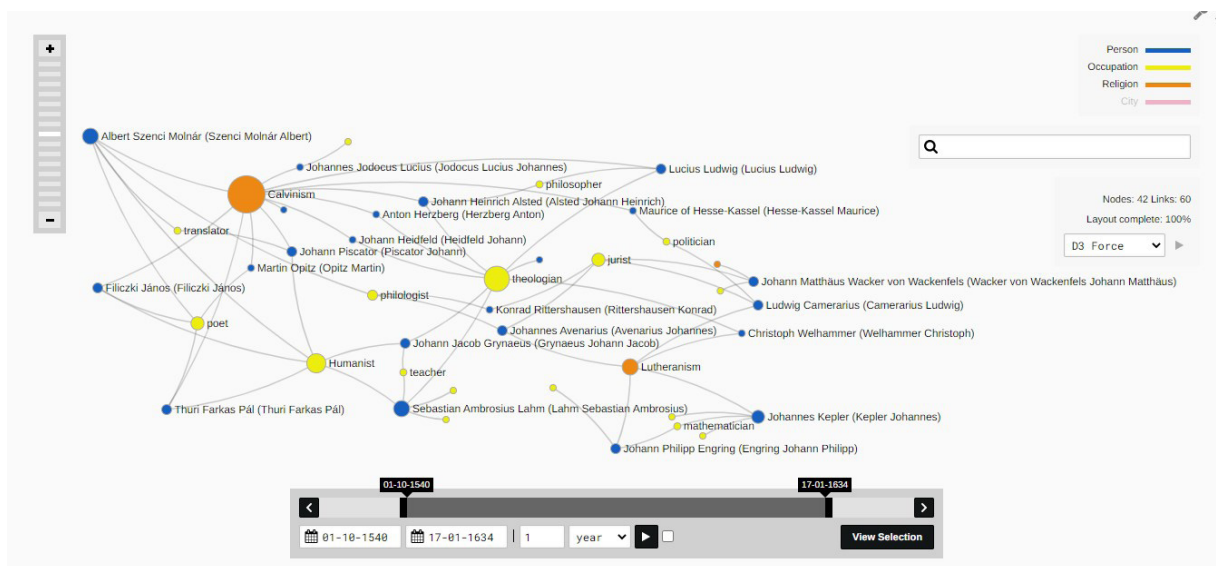
Fontos hangsúlyozni, hogy az ELTEdata adatvizualizációi nem korlátozódnak a SPARQL által kínált lehetőségekre. A következőkben ismertetendő vizualizációkhoz a *nodegoat* nevű platformot használtam, mely projekt adatkészlet-építéshez, adatmodellezéshez- és vizualizációhoz biztosít felületet.<sup>8</sup> Egy felhasználói fiók létrehozása után lehetőség nyílik saját adatmodell kialakítására: objektum-orientált megközelítésmódja révén a *nodegoat* jól alkalmazható az adatkészletben szereplő objektumok közötti komplex relációs hálózatok elemzésére. Az útmutató lépéseit követve kialakított adatmodellt utólag lehet „feltölteni” a kívánt adatokkal. Ily módon egy személynek megadható a foglalkozása és felekezeti hovatartozása, de lehetőség van külső azonosítók segítségével más adatbázisokhoz is hozzákötni (amennyiben a felhasználó az URI-t is megadja, közvetlen módon össze lesz kötve a kívánt névtérrel, például az ELTEdataival). A kézi adatbevitel mellett megoldható strukturált adatok feltöltése is, hiszen azok CSV-formátumban nagy mennyiségben importálhatóak, amennyiben a megfelelő adatmodell előzetesen ki lett alakítva. Ily módon tehát adott a lehetőség, hogy az ELTEdata SPARQL-végpontjáról letöltött CSV-t töltsük fel a *nodegoatra*. A bevitt adatokat ezután lehet vizualizálni, például térképen, ahol egy adott személy utazásai, lakó- és tartózkodási helyei ábrázolhatóak. A *nodegoat* egyik funkciója, hogy akár történelmi térképekre is rá lehet vetíteni a relációs adatokat, ami a HECEdatahoz hasonló történelmi prozopográfiaik esetében kifejezetten látványos. Földrajzi vizualizációk mellett lehetséges kapcsolati hálók ábrázolása is. Amennyiben például egy konkrét levelezésen keresztül akarok

<sup>8</sup> <https://nodegoat.net/> (Hozzáférés: 2022. június 9.) A felület használatához lásd az alábbi útmutatót: <https://nodegoat.net/guides>





egy kapcsolati hálót leképezni, egy úgynevezett levél-objektumtípust kell létrehozni, s ennek az adatmodelljét kialakítani. Amennyiben az adatmodell ki van dolgozva, az eredmény vizualizálható. A 6. ábra Szenci Molnár Albert kapcsolati hálójának egy részét mutatja, s mint az látható, megoldható az is, hogy az egyes személyek foglalkozása, vagy akár felekezeti hovatartozása is fel legyen tüntetve. Mi több, egy összeköttetés ugyancsak leképezhető azon foglalkozások között, melyekben mindkét személy tevékenykedett. Az ábrán jól látszik a csúcsok mérete közti különbség alapján, hogy ezek a csomópontok rendelkeznek a legtöbb éllel, vagyis közös pontok a legtöbb objektum között. A gráf tehát megmutatja, hogy Szenci Molnár ismerősei alapvetően református teológusok közül kerültek ki, de akadt közöttük jogász, humanista filológus, költő is, tehát mindenképpen a korabeli, döntően protestáns értelmiségi réteg tagjairól van szó. Érdekes összehasonlítási alapot jelenthet, hogyha egy szerényebb egyházi karriert befutott lelkész hálóját képezzük le hasonló módon, megfigyelve, hogy abban az esetben más foglalkozások (például lelkész vagy prédikátor) lesznek a gráf nagyobb csomópontjai.



6. ábra. Szenci Molnár Albert kapcsolatrendszere

A HECEdata adatiból még csak egy kisebb töredék lett importálva a nodegoat felületére, de a lehetőség adott, hogy komplexebb kapcsolati hálókat lehessen vizualizálni a későbbiekben.

Az ELTEdata egészével kapcsolatban a soron következő legfontosabb feladat egy olyan dokumentáció kidolgozása, mely mind az adatfelvitel kritériumait, mind az adatbázis használatát, funkcióinak bemutatását tartalmazza, így módon biztosítva felhasználói felületet azon kutatócsoportoknak, akik adatkészleteiket szemantikus alapon kívánják feldolgozni és közzétenni. Bár a kézi és a félautomatikus adatbetöltés még zajlik, elmondható, hogy már az eddig rendelkezésre álló adatok is eredményesen vizualizálhatóak, a szemantikus adatközpont pedig új lehetőségeket teremt a különféle kutatócsoportok és gyűjtemények számára.

## Új trendek a tanulási eredmények tanúsításában

Szlamka Erzsébet  
Digitális Jólét NKft.  
[szlamka.erzsebet@djnkft.hu](mailto:szlamka.erzsebet@djnkft.hu)

### Absztrakt

Három, az oktatási és képzési rendszeren belül tapasztalható, részben a digitális átállás részét képező változást mutat be a tanulmány:

1. az Egyéni Tanulási Számlát,
2. a mikrotanúsítványokat és
3. a tanúsítványokra vonatkozó digitalizációs törekvéseket.

A magasan képzett munkaerőt igénylő munkák 25%-kal emelkedtek az OECD országokban az elmúlt két évtizedben, és a dolgozók tovább maradnak a munkaerőpiacon; emiatt alapvető fontosságú az átképzés (reskilling) és a készségfejlesztés (upskilling). A munkaerőpiacon tapasztalható másik jelentős változás, hogy a korábbiakhoz képest fragmentáltabb az egyén karrierútja, egyre nagyobb arányú az atipikus foglalkoztatás. Ezen változásokat követnünk kellene a felnőttképzésben résztvevők arányának növelésével: reményeink szerint ebben is segíthet az Egyéni Tanulási Számla és a mikrotanúsítványok bevezetése, valamint a tanúsítványok digitalizációja.

A mikroképzések rövidebb idő alatt, könnyebben elvégezhetőek, így a felnőtt lakosság motiválhatóbb a tanulásban való részvételre. Ugyanakkor az ezek során szerzett tanulási eredményeket nem rögzítik megfelelően, az érték adható kreditek nem akumulálódnak, pedig ez fontos lenne a képzésben résztvevő, a munkaadó és a gazdaság szempontjából egyaránt. A tanulási eredmények, kreditek, tanúsítványok rögzítését egyre több helyen digitálisan szeretnék megoldani: a virtuális tároláshoz ma már rendelkezésre állnak megfelelő adatvédelmi és hitelesítési (blokklánc) technikák.

**Kulcsszavak:** mikrotanúsítvány, blokklánc, blockchain, Egyéni Tanulási Számla, digitális átállás, egyéni képzési számla

### Abstract

The paper presents three changes in the education and training system, partly part of the digital transition:

1. the Individual Learning Accounts,
2. Microcredentials and
3. Digital Credentials.

Jobs requiring highly skilled labour have increased by 25% in OECD countries over the last two decades, and workers stay longer in the labour market, making reskilling and upskilling essential. Another major change in the labour market is that individuals' career paths are more fragmented than before, with an increasing share of atypical employment. These changes should be followed by increasing the participation rate in adult learning: the introduction of the Individual Learning Accounts and microcredentials, as well as the digitalisation of certificates, will support this process.



Microcredentials can be completed in a shorter time and more easily, so that the adult population is more motivated to participate in learning. However, the learning outcomes are not properly recorded, and credits are not accumulated, which is important for the trainee, the employer and the economy. Increasingly, the recording of learning outcomes, credits and certificates is being digitalised: appropriate data protection and authentication (blockchain) techniques are now available for virtual storage.

**Keywords:** microcredentials, blockchain, Individual Learning Account, digital transition, individual learning accounts

## Új trendek a tanúsításban

### Bevezetés

A gyorsan változó munkaerő-piaci igények egyrészt a munkahelyi képzés, másrészt a felnőttképzési rendszerek növekvő szerepére, valamint rugalmasságukra és gyors reagálási képességükre mutatnak rá. Ennek ellenére a magyar lakosság felnőttképzésben való részvétele 5% körül mozog, míg az uniós átlag 10%, egyes balti államokban 20%.

A képzésben való részvételt alábbi kihívások befolyásolják:

- 1) A képzés költségei, amelybe beleértendők a közvetett költségek is, így például a kieső jövedelem és lakhatás, amennyiben a lakóhelytől távol valósul meg a képzés.
- 2) Időbeli korlátok: a munkahelyi és családi kötelezettségek mellett nehéz időt találni a képzésnek is.
- 3) A motiváció hiánya: a képzéshez való általános pozitív hozzáállás azért nem kamatozik a gyakorlatban, mert a) az elérhető támogatások és képzések nem elég átláthatók, b) az egyén nincs meggyőződve a képzés színvonaláról és munkaerőpiaci elfogadottságáról, c) nincs megfelelően az egyénre szabva a képzési terv.

A felnőttképzésben való részvételi arány növeléséhez, a végzettségek és kompetenciák egyéni nyomon követéséhez és igazolásához, valamint a hatékonyság követéséhez hasznos eszköz az egyéni tanulási számla (ETSZ) rendszerszintű bevezetése és a mikrotanúsítványok<sup>1</sup> kiadása. Ezen felismerésen alapul két, várhatóan 2022 júniusában elfogadásra kerülő tanácsi ajánlás: egyik az Egyéni Tanulási Számláról, másik a mikrotanúsítványok európai rendszeréről. Jelen tanulmány e két eszközt mutatja be, megemlítve az eszközök digitalizálásában rejlő lehetőségeket is.

### Az Egyéni Tanulási Számla (ETSZ)

Az élethosszig tartó átképzés és készségfejlesztés sokféle – elsősorban egyénhez kötött - finanszírozási lehetőséget igényel, melyet összefoglaló néven Egyéni Tanulási Számlának (Individual Learning Account, ILA) nevezünk. Az ETSZ bevezetésével a képzési jogok és az anyagi háttér (a munkáltató helyett) az egyénhez kötődnek, ezáltal a munkahelyek között átvihetők. Az ETSZ sok előnyt ígér (jogokkal és felelősséggel is felruházva az egyént), de a bevezetés módja, a minőségbiztosítás és a monitoring alapvető a sikerhez: hibás tervezés és bevezetés esetén növelheti a részvételi különbségeket a felül és alul reprezentált csoportok között, ráadásul a család lehetőségét is ki kell küszöbölni.

<sup>1</sup> A mikrotanúsítványokról, valamint a kreditek, tanúsítványok, képesítések, bizonyítványok digitalizálásáról külön összefoglaló készül. Javasolt azonban ezt a három témát (egyéni tanulási számla, tanúsítványok digitalizációja és kisebb egységekben történő kiadása) rendszerben szemlélni.



Az ETSZ-t az alábbiak szerint lehet kategorizálni:

- Individual Learning Account (ILA) – Egyéni Tanulási Számla (ETSZ)  
Virtuális egyéni számla, amelyen a képzési jogok és az anyagiak felhalmozódnak; a forrást csak akkor használják fel, ha ténylegesen megvalósul a képzés. A francia Compte Personnel de Formation (CPF) a legjobb és legközelebbi példa erre.
- Individual Savings Account for Training – Egyéni „spórolási számla”  
Pilot szintjén próbálták csak ki (Kanada, USA), de nem eresztett gyökeret; az egyén tanulási célból nem nagyon spórol.
- Training vouchers – Képzési utalvány  
Az egyén közvetlen támogatást kap a képzéséhez, általában önerő szükséges. Nincs felhalmozás, nem gyűlik a pénz. (Sokszor ezt is ILA-nak hívják.)  
Az állam, a munkaadók, egyéb partnerek is részt vehetnek a rendszer kialakításában és működtetésében.

Magyarországon az ETSZ-hez hasonló tapasztalat a Diákhitel Zrt. által kezelt Képzési Hitel, amelyet aktív tanulói vagy felnőttképzési jogviszonnyal, 18-55 éves kor között lehet igényelni három hónapnál hosszabb idejű képzésekre.

Külföldön az ETSZ-t eddig többek között az Egyesült Királyságban, Franciaországban, Kanadában, az Amerikai Egyesült Államokban, Belgiumban, Dániában és Szingapúrban próbálták ki; alább néhány tapasztalatot mutatunk be:<sup>2</sup>

- Egyesült Királyság: 2000-ben vezették be az egyéni tanulási számlát, de nem volt megfelelő a minőségbiztosítás és a monitorig, ezért sok csalásra adott lehetőséget. A hibákat kijavították, a sémát módosították, így alakult ki az új egyéni képzési számla.<sup>3</sup>
- Skóciában a jelenleg működő rendszert Egyéni Képzési Számlának hívják.<sup>4</sup> Hangsúlyozzák, hogy a foglalkoztatáshoz szükséges készségek fejlesztésére való ez elsősorban, és a kormány által kiemelt területekről való kurzusokat támogatják (jelenlegi felsorolás: mezőgazdaság, üzlet, építőipar, koragyermekkorai nevelés, egészségügy, közlekedés). Van egy online kurzuslista, abból kell választani és siker esetén akár 200 Font támogatást is lehet kapni, amennyiben jogosult valaki.
- Szingapúrban az Edu Save Account (Oktatási Takarékos Alap) program keretében, melyet a szülőkkel együttműködve terveztek és alakítottak ki, az Oktatási Minisztérium letétbe helyez egy bizonyos, a parlament által jóváhagyott összeget a gyermekek számára egy külön számlán, amelyet a szülők a gyerekek kiegészítő, fejlesztő óráira, vagy éppen felsőoktatásuk költségeire használhatnak fel.
- Az Oktatási Minisztérium minden 25-nél idősebb szingapúri számára 500 USD egyszeri SkillsFuture Kreditet tölt fel, amelyet 2025-ig költhetnek el több mint 8000 tanfolyamon. Annak érdekében, hogy ezt az összeget megalapozottan költse el, a MySkillsFuture portál digitális eszközöket is nyújt a döntéshez, itt személyre szabott ajánlásokat is kapnak a felhasználók.  
A SkillsFuture pálya-újrakezdő támogatási csomag célja, hogy több karrierlehetőséget teremtsen a 40-50 év körüli helyiek számára, és segítse őket abban, hogy továbbra is foglalkoztathatók maradjanak, és képesek legyenek jó

---

2 Bővebben lásd OECD (2019), Individual Learning Accounts : Panacea or Pandora's Box? <https://doi.org/10.1787/203b21a8-en>. Hozzáférés: 2022. 05. 26.

3 <https://www.nao.org.uk/report/individual-learning-accounts/> Hozzáférés: 2022. 05. 26.

4 <https://www.myworldofwork.co.uk/learn-and-train/sds-individual-training-accounts-ita>  
Hozzáférés: 2022. 05. 26.



munkahelyekhez jutni. 2025-re az a céljuk, hogy az átképzési programok révén megduplázzák a célcsoport éves elhelyezkedését a következő kezdeményezések révén:

- A kormány hat hónapon keresztül 40%-os fizetéstámogatást nyújt a munkáltatónak, ha 40 évesnél idősebb helyi álláskeresőt alkalmaz, aki átképzési programban vesz részt.
- További 500 USD kreditet tölt fel minden 40-60 év közötti szingapúri álláskeresőnek, aki szakképzési programban kíván részt venni.
- Önkéntes tanácsadókból álló szolgáltatás keretében karriertanácsadást nyújtanak a helyi munkakeresőknek.

Az Egyéni Tanulási Számla egyik lehetséges használata a finanszírozásra irányul: az egyén képzési célú anyagi támogatásának nyilvántartását jelenti. A második, amely a külföldi tapasztalatokban nincs jelen: az egyén tanulási eredményeinek nyomonkövetése, amennyiben a digitálisan tárolt számlára fel lehet tölteni az egyén által megszerzett képesítéseket, mikrotanúsítványokat, krediteket stb.

Az ETSZ sikeres bevezetésének kulcsa az egyszerűség és átláthatóság, a megfelelő és előre kiszámítható támogatás (célcsoport-meghatározás, finanszírozók köre), a rászoruló csoportok nagyobb mértékű támogatása, a tanácsadás és információnyújtás folyamatossága és elérhetősége, a minőségi képzések biztosítása és explicit kapcsolódás a munkaadók által nyújtott képzésekhez, a megfelelő minőségbiztosítás és monitoring.

Fontos eldönteni, melyik aktor (állam, munkaadó, civil szervezetek, munkavállaló) milyen arányban vállal részt a felnőttképzésben való részvétel támogatásában, továbbá, hogy pontosan milyen súlyozással történik a finanszírozás: a munkaerőpiaci kereslet alapján kerül meghatározásra a támogatás mértéke és/vagy az egyén rászorultsága lesz mérvadó a támogatás megítélésékor.

## A mikrotanúsítványok

A mikrotanúsítvány (MT) olyan okmány, amely a rövid távú, átláthatóan értékelt (online, helyszíni vagy vegyes) tanfolyamok vagy modulok révén megszerzett tanulási eredményeket igazolja. Az ilyen típusú képesítések rugalmas jellegének köszönhetően egyre több ember (pl. teljes munkaidőben dolgozók) előtt nyílnak meg tanulási lehetőségek. A mikrotanúsítvánnyal igazolt képzés tehát egy nagyon rugalmas, befogadó tanulási forma, amely képességek és kompetenciák célzott elsajátítását teszi lehetővé. Mikrotanúsítványt felsőoktatási, felnőttképzési és szakképzési intézmények, valamint magánszervezetek nyújtanak.

A Cedefopnak 2021-23 között zajlik egy, a mikrotanúsítványok hasznát és használatát kutató projektje, amelynek első szélesebb körű beszámolójára egy, 2021. nov. 25-26-án lezajlott konferencia keretében került sor.<sup>5</sup> A mikrotanúsítványok főleg a felsőoktatásban jellemzők, de a szak- és felnőttképzésben is lehet helye, valamint a cégek és a munkaerőpiaci szereplők tudnák használni. Luxemburgban például már 8000 tanuló kap digitális mikrotanúsítványokat a szakképzésben; a felnőttképzésben való használatra legjobb példák a Nyitott Online Kurzusok (MOOC), valamint a különböző nagyvállalatok (pl. Google, Amazon, Microsoft, IBM) által munkakör-specifikusan kínált képzések.

5 A konferencia honlapján letölthetők a prezentációk, megnézhetők a videók: <https://www.cedefop.europa.eu/en/events/conference-microcredentials>

A mikrotanúsítványok és a képesítési rendszer lehetséges kapcsolatát hasonlóan lehet leírni, mint a nemzetközi képesítések besorolási lehetőségeit:<sup>6</sup>

1. A mikrotanúsítványok a képesítési rendszer szerves részét képezik. Amikrotanúsítvánnyal rendelkező a képesítési rendszerben elfogadott kreditet kaphatnak a MT-ért. A MT teljes bizalmat élvez, átlátható a rendszer, a minőségbiztosításért felelős állami szerv felügyeli.
2. A mikrotanúsítványok működhetnek közfinanszírozásból (is?), de a képesítési rendszerrel párhuzamosan. A képesítési rendszerbe nem tartoznak bele, de létezik hozzá állami támogatás; a bizalmat a közfinanszírozott szolgáltatók bevonása erősítheti (?).
3. A mikrotanúsítványok a képesítési rendszertől teljesen elkülönülten léteznek. A MT értéke, státusza, a belé vetett bizalom nagyban a kiadón múlik.

Az első esetben erőssége a minőségbiztosítás, ugyanakkor a képesítési rendszerbe történő besorolás elveszi a mikrotanúsítványok legnagyobb előnyét: a rugalmasságot, a munkaerőpiaci igényeknek megfelelő gyors tartalommodosítás lehetőségét. A piaci modell, vagyis a harmadik verzió viszonylag idegen a magyar képzési rendszertől, ráadásul a hazai lakosság képzéssel kapcsolatos tudatossága még nincs azon a szinten, hogy ez a megoldás sikerre vezessen. Mindezek alapján Magyarországon a második verzió lehet célszerű.

A magas színvonalú mikrotanúsítványok legfontosabb szempontjai az alábbiak lehetnek:<sup>7</sup>

- A nemzeti hatóságok elismerik azokat
- Az oktatási és képzési szervezetek elismerik azokat
- A munkáltatók elfogadják azokat
- Elérhetőek a tanuló anyanyelvén
- Elérhetőek egy széles körben beszélt nyelven
- Megszerzésük nem drága
- Tartalmuk elismert oktatási és képzési szolgáltatóktól származik
- Tartalmuk munkáltatókkal, vállalkozói szövetségekkel vagy kereskedelmi kamarákkal együttműködésben került kidolgozásra
- Tartalmuk a munkaerőpiaci igényekhez kapcsolódik
- Elérhetőek online oktatási platformokon
- Olyan készségeket és kompetenciákat kínálnak, amelyek közvetlenül hasznosíthatók jelenlegi vagy jövőbeni munkahelyen
- Olyan készségeket és kompetenciákat kínálnak, amelyek hasznosak a továbbtanulás szempontjából
- Elősegítik az országok közötti mobilitást
- Részleges vagy teljes szakképesítések építőkövei lehetnek
- Megszerezhetőek a készségek validálása révén
- Minőségbiztosításuk átlátható minőségi szabványokon alapul

---

<sup>6</sup> Andrew McCoshan, Cedefop konferencia, 2021. nov. 25-26.

<sup>7</sup> Az Európai Bizottság kérdőíve alapján (Commission Staff Working Document Accompanying the Document Proposal for a Council Recommendation on a European Approach to micro-credentials for Lifelong Learning and Employability. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021SC0367&qid=1639417448217> Hozzáférés: 2022. 05. 26.



A fentiek figyelembevételével célszerű egy, az írhez hasonló Roadmap<sup>8</sup> kialakítása, amelynek keretei között – többek között – készül egy munkaadói és munkavállalói felmérés, hangsúlyt fektetnek a tudatosságnövelésre és együttműködésre az érdekelt felek között, megtörténik a szervezeteken belüli képesség- és kompetenciatérképezés, továbbá a munkaalapú tanulás validálására és értékelésére vonatkozó modellek is kidolgozásra kerülnek.

### A tanúsítványok digitalizálása

A kompetenciák, tanúsítványok, kreditek, egyéni számla tárolására korszerű és lényegében egyetlen lehetőség a blokklánc technológia alkalmazása. A blokklánc technológia oktatási szférában történő felhasználási lehetőségeit áttekintő JRC tanulmány (JRC, 2017) bemutatja a teljes tanulási életút adatainak (formális, nem formális, informális tanulás) rögzítését blokkláncon. Érdemes biztosítani a kompetenciák munkáltatói igazolásának lehetőségét is (pl. Open Badges). Az egyéni tanulási számlák és a tanúsítványok, képesítések hiteles elektronikus rögzítését a Magyarországon is elérhető blokklánc technológiával lenne lehetséges megoldani.

### Összegzés

A mikrotanúsítványok különösen azok számára lehetnek hasznosak, akik saját tudásukra szeretnének építkezni anélkül, hogy egy teljes képzési programot elvégeznének, valamint tovább vagy át szeretnék képezni magukat, hogy megfeleljenek a munkaerőpiaci igényeknek, illetve elhelyezkedésük után szakmai tudásukat kívánják fejleszteni. Ezek nyilvántartására és megszerzésük anyagi támogatására vezették be sok helyen az Egyéni Tanulási Számlát, hogy a munkaképes korú lakosság minél nagyobb arányban vegyen részt képzésekben. Az ETSZ és a mikrotanúsítványok blokklánc alapú nyilvántartása a tanúsítványok kezelésének megkönnyítése érdekében, valamint a zöld és digitális átállás jegyében történik.

Az Egyéni Tanulási Számla, a tanúsítványok „mikrosítása” és digitalizálása egy rendszerben kezelendő: az ETSZ-t blockchain alapon javasolt bevezetni a korábbi tapasztalatok felhasználásával (l. bukott angol verzió, Diákhitel Zrt. tapasztalatai a felnőttképzési hitel kapcsán stb.). Ugyanakkor fontos megkülönböztetni a digitális és a mikrotanúsítványokat. Digitalizálni mindegyik tanúsítványt, képesítést lehet, a nyelvvizsgától az érettségi bizonyítványon át a diplomáig, éppen ezért ezek körét nem érdemes bármelyik szektorra, iskolarendszerre vagy egyéb szempontok szerint leszűkíteni. A mikrotanúsítvány alkalmazhatósága, állami szabályozása meglátásunk szerint már szakpolitikai megfontolás alapján dől el.

8 A Micro-Credential Roadmap: Currency, Cohesion and Consistency. <https://www.skillnetireland.ie/publication/a-micro-credential-roadmap-currency-cohesion-and-consistency/> Hozzáférés: 2022. 05. 26.

## Felhasznált irodalom

- A European approach to micro-credentials. <https://education.ec.europa.eu/education-levels/higher-education/micro-credentials> Hozzáférés: 2022. 05. 26.
- A Micro-credential Roadmap: Currency, Cohesion and Consistency. March 2021, Skillnet  
Cedefop: Individual Learning Accounts.  
[https://www.cedefop.europa.eu/files/5192\\_en.pdf](https://www.cedefop.europa.eu/files/5192_en.pdf) Hozzáférés: 2022. 05. 26.
- Europass Digital Credentials Infrastructure.  
<https://europa.eu/europass/en/what-are-digital-credentials> Hozzáférés: 2022. 05. 26.
- Grech, A. and Camilleri, A. F. (2017) Blockchain in Education. Inamorato dos Santos, A. (ed.)  
EUR 28778 EN; doi:[10.2760/60649](https://doi.org/10.2760/60649)
- National Audit Office: Individual Learning Accounts. REPORT BY THE COMPTROLLER AND  
AUDITOR GENERAL HC 1235 Session 2001-2002: 25 October 2002.  
<https://webarchive.nationalarchives.gov.uk/ukgwa/20170207052351/>  
<https://www.nao.org.uk/wp-content/uploads/2002/10/01021235.pdf> Hozzáférés: 2022.  
05. 26.
- OECD (2019), Individual Learning Accounts: Panacea or Pandora's Box?, OECD Publishing,  
Paris, <https://dx.doi.org/10.1787/203b21a8-en> Hozzáférés: 2022. 05. 26.

## Webshop indítása közkönyvtári környezetben

Tóth Máté

*Hamvas Béla Pest Megyei Könyvtár*

[igazgato@pmk.hu](mailto:igazgato@pmk.hu)

Héjja Balázs

*Hamvas Béla Pest Megyei Könyvtár*

[informatika@pmk.hu](mailto:informatika@pmk.hu)

### Abstract

Launching a public library webshop

In 2021 Autumn Béla Hamvas Pest County Library launched the first public library webshop in Hungary. (Books from the city of arts – online bookshop from Szentendre). The library sells its own publications and other institutions', organisations' books via partnership agreements. The development of the online bookshop is accomplished by the library employees. The paper presents the development process from the emergence of the demand to the realization, and the first experiences of the first 3 months of operation.

During the development we created a WordPress interface in which we elaborated the main function by using the plug-ins of WooCommerce. As far as management issues are concerned it was a significant challenge how a public library can fit the operation of such a webshop into its regular workflows.

After three months of operation we can conclude that the innovative solution supports the accomplishment of our mission. Our further tasks are the promotion of the webshop among potential customers and the broadening of the number of products.

2021 őszén Szentendrén a Hamvas Béla Pest Megyei Könyvtárban elindult Magyarország első közkönyvtári webshopja „Könyvek a művészetek városából – online könyvruház Szentendréről” névvel. Az intézmény a saját kiadványait és bizományosi szerződések útján partnerek kiadványait árusítja az oldalon. A webshop fejlesztése saját erőforrásból történt. A tanulmány bemutatja a webshop fejlesztését az igény megjelenésétől kezdve a megvalósításig és a működés első hónapjainak tapasztalatait.

A fejlesztés során egy WordPress felületet készítettünk, amelyen a *WooCommerce* moduljai segítségével alakítottuk ki az online áruház főbb funkcióit. Menedzselési szempontból kihívást jelentett, hogy egy közkönyvtár hogyan tudja beilleszteni a normál munkafolyamatai közé egy online webáruház működtetéséhez szükséges tevékenységeket.

Az első három hónapi működését követően elmondható, hogy egy olyan innovatív kezdeményezést sikerült megvalósítanunk, amellyel újszerű módon járulunk hozzá a könyvtárunk küldetésének teljesítéséhez. További feladatunk, hogy a webshopot ismertebbé tegyük a vásárlók körében, továbbá, hogy újabb partnerek bevonásával bővítsük az áruház kínálatát.



## 1. Bevezetés

A szentendrei Hamvas Béla Pest Megyei Könyvtár 2021 őszén indította el Magyarország első közkönyvtári webshopját „Könyvek a művészetek városából – online könyvárúház Szentendréről” név alatt. Az intézmény elsősorban a saját és a partnerei kiadványait árusítja ezen a felületen. Az volt a tapasztalatunk, hogy a könyvtár saját – tartalmát tekintve igen értékes – kiadványaiból nagyon sok megmaradt a raktárainkban, amelyeket szerettünk volna eljuttatni az érdeklődőkhöz. Az intézményünk már korábban is árusította a saját helytörténeti témájú kiadványait, de az volt a tapasztalatunk, hogy csak igen kevés vásárlót sikerül elérnünk az alapvetően rétegiglést kiszolgáló termékeinkkel. Az internetes kereskedelemben sokkal nagyobb esély van arra, hogy azok a termékek is megtalálják a vásárlóikat, amelyek iránt a könyvtár fizikai tereiben csak mérsékelt lenne az érdeklődés. Chris Anderson a „long tail” vagyis hosszú fark terminussal írta le azt a jelenséget, amelyet a könyvtár munkatársaival ki szerettünk volna használni.<sup>1</sup> A hosszú fark azokra a termékekre utal, amelyek általában csak kis mennyiségben adhatók el, ezért a hagyományos kereskedelem korszakában esélyük sem volt a piacra jutásra, hiszen nem érte meg belőle annyit készíteni, amely fedezte volna az előállítás és a piacra dobás költségeit. Ezeknek a termékeknek a változatossága sokkal nagyobb, mint a nagy mennyiségben árusítható tömegtermékeké. Az internetes kereskedelem útján távolról is rátalálhatnak azokra a termékekre, amelyeket a kis példányszám miatt korábban akár létrehozni sem nagyon érte meg.<sup>2</sup>

A célunk egy konkrét példán, a Hamvas Béla Pest Megyei Könyvtár online könyvesboltján keresztül ismertetni, hogy egy közkönyvtárnak milyen feladatokkal és közben milyen nehézségekkel kell szembenéznie egy webshop elindítása során. A cikkben azokat a döntési helyzeteket írjuk le, amelyek az oldal létrehozása, a szállítási és fizetési feltételek kialakítása és a webshop szervezeti kereteinek kialakítása kapcsán felmerült.

## 2. Az oldal létrehozása

A webshop a könyvtárhoz kötődő üzleti vállalkozás. Szólhatnak érvek az önálló brand kialakítása mellett és a mellett is, hogy a könyvtár arculatának a részét képezze a webshop. Az online felület kialakításában ez annyiban jelentett döntési helyzetet, hogy az a könyvtár aldoménjeként, annak arculati elemeivel együtt, vagy teljesen önállóan jelenjen meg. Mi az előbbi mellett döntöttünk <http://webshop.hbpmk.hu> címen egy aldomént hoztunk létre, amely arculatában különbözik a könyvtár többi oldalától, de mégis egyértelműen jeleztük, hogy a tartalom a Hamvas Béla Pest Megyei Könyvtárhoz köthető.

A webshop elkészítéséhez a WordPress rendszert választottuk, amelyet installálnunk kellett a saját szerverünkre. Kialakítottuk a StoreFront sablont. Szándékosan egy egyszerű, puritán küllemet választottunk a webshopnak. Létre kellett hoznunk egy „gyermek sablont” annak érdekében, hogy az időről-időre végrehajtott WordPress verziófrissítések alkalmával az egyedi módosítások ne vesszenek el. Így lehetőség nyílt az oldal kinézetét teljes mértékben testre szabni.

---

1 ANDERSON, Chris: *Hosszú fark – A végtelen választék átírja az üzlet szabályait*. Budapest, HVG kiadó, 2007.

2 DIPPOLD Péter: *A long tail és a könyvtárak. Hogyan kapcsolódhatunk a „hosszú fark” jelenséghez?* KISZL Péter, CSÍK Tibor, szerk. *Valóságos könyvtár – könyvtári valóság*. Könyvtár- és információtudományi tanulmányok 2018. (Budapest, ELTE BTK Könyvtár- és Információtudományi Intézet, 2018 167–179.).



Az oldalhoz a következő bővítményeket telepítettük:

- WooCommerce – a webáruház fő motorja, kezeli a rendeléseket és a hozzájuk kapcsolódó összes folyamatot
- Akismet Anti-Spam – általános spamszűrő modul ami meggátolja, hogy a spamrobotok teleszemeteljék az oldalt kéretlen hozzászólásokkal
- Contact Form 7 – kapcsolatfelvételi űrlap modul az esetleges vásárlói reklamációkhoz
- Elementor – a WordPress adminisztrációs felületén megkönnyíti az oldalak szerkesztését
- OTP SimplePay Gateway for WooCommerce – online bankkártyás fizetési modul az OTP Simplepay rendszeréhez
- UpdraftPlus – szükséges automatikus adatbázis mentő modul, minimálisra csökkentjük vele az esetleges adatvesztés kockázatát
- HuCommerce – a WooCommerce magyar fordítása
- WooCommerce PayPal Payments - online bankkártyás fizetési modul PayPal rendszerhez
- WooCommerce PDF Invoices & Packing Slips – a WooCommerce kiegészítő modulja, díjbekérők és csomagcímkék létrehozásához
- WooCommerce Shipping & Tax – a WooCommerce kiegészítő modulja a szállítási díjak és adók kikalkulálásához
- Wordfence Security – szoftveres tűzfal modul a biztonság érdekében
- WP Mail Logging – a rendszer által automatikusan küldött e-maileket tárolja, probléma esetén később visszakereshetők
- Yoast SEO – az oldal keresőoptimalizálását segítő modul

Ezek a bővítmények elegendőnek bizonyultak az alapvető felhasználói élmény biztosításához, amelyeket egy átlagos vásárló egy kisebb internetes áruházról elvár.

### 3. A termékkínálat összeállítása

Az oldal kialakítását követően döntenünk kellett az árusítandó termékkínálatról. Ennek gerincét a könyvtár saját kiadványai adják, amelyeket az olvasóteremben, a rendezvényeinken is árusítottuk már korábban is, de úgy éreztük, hogy nagyon kevés emberhez sikerült eljutnunk ezekkel a könyvekkel. Egy praktikus és felhasználóbarát webshoppal mind térben, mind időben jelentős mértékben kitágítottuk a kiadványainkra való rátalálás lehetőségét.

Önmagukban azonban a könyvtár kiadványai nem lettek volna kellően vonzóak és sokrétűek, amelyeket egy felhasználó elvár egy webáruházról, ezért úgy döntöttünk, hogy felkeressük szentendrei partnereinket, hogy bizományosi szerződések keretében árusíthassuk az általuk kiadott könyveket. Bizományosi szerződéseket kötöttünk a kulturális intézményekkel (Ferenczy Múzeumi Centrum, Szentendrei Teátrum), a helyi idegenforgalom fejlesztéséért felelős Turisztikai Desztinációs Menedzsment Irodával, civil kezdeményezésekkel (pl. Szentendrei Könyvklub, Kóka Ferenc Művészeti Egyesület), amelyek a könyvtárhoz hasonlóan kevés, de értékes, a művészetek városához egyértelműen kötődő könyveket adtak ki.

Felmerült – és részben meg is valósult – a termékkör további bővítésének lehetősége is. Egyrészt újabb érdeklődő szervezetekkel és magánszemélyekkel kötöttünk bizományosi szerződéseket (pl. a Pest megyei nyilvános könyvtárak közül a százhalombattai Hamvas Béla Városi Könyvtárral, Juhász Árpáddal és Zichó Viktorral). Másrészt ezt a fórumot választottuk a könyvtár karácsonyi ajándékkuponjainak az online árusítására.

A termékek jobb kereshetősége érdekében termékkategóriákat határoztunk meg. A projektben résztvevő kollégákkal egy megosztott táblázatban dolgoztuk fel a könyvek bibliográfiai adatait, majd határoztuk meg a bruttó eladási árakat és azt, hogy melyik

termékkategóriá(k)ba soroljuk be. Az induláskor a következő termékkategóriákat határoztuk meg:

- Dunakanyar helytörténet,
- Életrajzok, emlékkönyvek,
- Kiállítás,
- Könyvek olvasásról, könyvtárról,
- Könyvek Szentendréről,
- Művészet,
- Szépirodalom,
- Útleírás.

Szintén a hatékonyabb kereshetőség érdekében készítettünk valamennyi termékünkről leírást, amely az adott tétel mellett jelenik meg az oldalon. Az induláshoz be kellett szkennelnünk valamennyi könyvborítót.

#### 4. A rendelési, szállítási és fizetési feltételek

Az oldal kialakítását követően számos döntést kellett hoznunk a rendelési, szállítási és fizetési feltételek vonatkozásában. A bankkártyás fizetés lehetősége már olyan alapvető követelmény, amelyet nem lehet megkerülni. 2021 februárjában kértünk először ajánlatot az online fizetési rendszerre, de csak 2021 októberére sikerült azt teljes mértékben üzembe állítani. Meg kellett vásárolnunk a TechSource fizetési modult a WooCommerce -hez. Sajnos nagyon sok technikai hiba, lassulás, leállás jellemzi a jelenlegi rendszert. Ezeket a kieséseket, nehézségeket a végfelhasználó velünk, a szolgáltatóval azonosítja és nem az alvállalkozóval. A bankkártyás fizetés mellett lehetőség van online előre utalással is fizetni. Az első félév tapasztalatai azt mutatják, hogy ezt a fizetési módot meglepően sokan (a megrendelők 50%-a) választották a vásárlóink közül, így – bár elsőre kérdés volt, hogy van-e értelme – ezt a lehetőséget is megtartottuk.

#### 5. Szállítási módok és díjak meghatározása

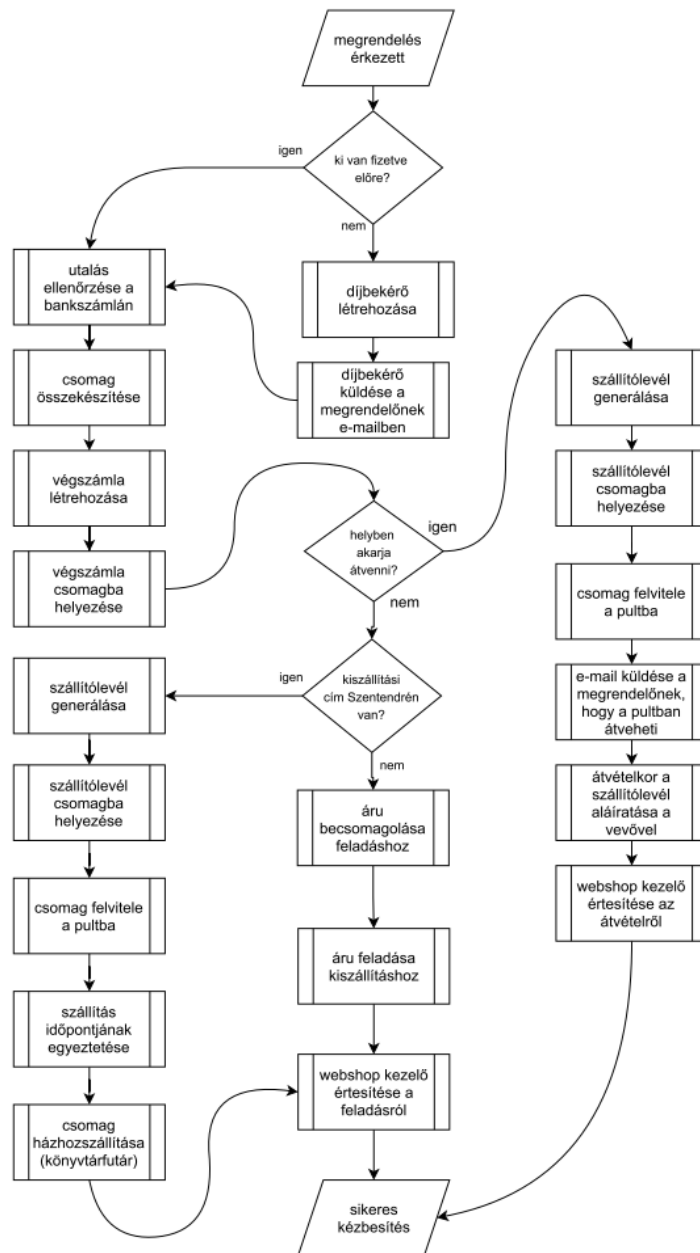
A következő döntéshelyzet a szállítási mód és a hozzá kapcsolódó szolgáltató kiválasztása volt. Tekintettel arra, hogy a könyvtár minden héten összesen 44 órát tart nyitva a szentendrei belvárosban, mindenképpen szerettünk volna biztosítani személyes átvételi lehetőséget. Ezt az olvasószolgálati munkatársak készségesen vállalták. Szentendrén ezen túl mi magunk vállaltuk a házhoz szállítást. Ez utóbbi is könnyen beilleszthető volt a munkafolyamatok közé, ugyanis a könyvtár a COVID-19 járvány ideje alatt az időseknek és a betegség miatt házhoz kötött olvasók számára bevezette a kölcsönzött dokumentumok házhoz szállítását. Egy-egy ilyen házhoz szállítás esetén a sofőrünk egyeztet a vásárlókkal is, akiknek helybe viszi a megrendelt dokumentumokat. Mind a könyvtárban való személyes átvétel, mind pedig a szentendrei házhoz szállítás ingyenes.

A postázás megoldása kapcsán több alternatívát is mérlegeltünk. A Magyar Posta mellett több szolgáltató neve is felmerült. Az ár, a gyorsaság, a könnyű elérhetőség (az ország csomagautomatákkal való lefedettsége) és a megbízhatóság voltak a legfontosabb szempontok. A díjszabás kapcsán úgy döntöttünk függetlenül a vásárolt dokumentum méretétől egységes díjat számolunk fel, így a vásárló számára is tervezhetőbbek a rendelés során felmerülő költségek, illetve ezzel is igyekszünk a vásárlókat abba az irányba terelni, hogy egyszerre több terméket tegyenek a kosarukba.

A webshop elindítása utáni két hónapban kedvezményesen és egységesen 990 forintban határoztuk meg a szállítási díjat, amely alatta volt a jellemzően kifizetendő valós költségnek, de ezzel kívántuk ösztönözni a vásárlásokat. A 10.000 forint fölötti rendelések esetében pedig a könyvtár átvállalta a teljes postaköltséget. Ez az akció szintén csak az első két hónapban élt.

## 6. Menedzselési kérdések

Menedzselési szempontból kihívást jelentett, hogy egy közkönyvtár hogyan tudja beilleszteni a normál munkafolyamatai közé egy online webáruház működtetéséhez szükséges tevékenységeket. A könyvtárunk a minőségirányítás eszköztárával szervezi a tevékenységeit, ezért első lépésként elkészítettük a rendelés teljesítésének folyamatábráját és azonosítottuk a munkafolyamat résztvevényeit.<sup>3</sup>



3 BAJNOK Tamara és BOGNÁRNÉ LOVÁSZ Katalin és FEHÉR Miklós és HORVÁTH Adrienn és MÉSZÁROSNÉ MERBLER Éva és TOPÁR József és TÓTH Máté., *A könyvtárak minőségi működésének értékelési rendszere* (Budapest, Könyvtári Intézet, 2019). [https://ki.oszk.hu/sites/default/files/dokumentumtar/kmer\\_atdolg.pdf](https://ki.oszk.hu/sites/default/files/dokumentumtar/kmer_atdolg.pdf) (2022. június 17.)

Ezt követően határoztuk meg, hogy az intézmény munkatársai közül ki melyik tevékenységet vállalja magára. Mivel a webshop üzemeltetése kapcsán felmerülő feladatok senkinek a munkaköri kötelezettségei között nem szerepeltek, erre dedikáltan új munkatársat pedig nem vehettünk fel, azt az alapelvet fogalmazzuk meg, hogy minden egyes munkafolyamatot az a munkatárs végezze, aki az adott tevékenységet más szolgáltatásunk kapcsán egyébként is végzi. Így a számlázás a gazdasági irodához került, a postázás a postázóba, a személyes átvételek és a kiszállítások szervezése pedig az olvasószolgálatához.

A fentiek mellett kellett egy munkatárs, aki vállalta, hogy a webshop valamennyi rendelését figyelemmel kíséri, nyomon követi azok állását és szükség esetén beavatkozik. Ezt a feladatkört a művelődésszervezőnk vállalta magára, aki eddig is segített a rendezvényeinken a könyvtár kiadványainak az árusításában. Szintén a művelődésszervezőnk vállalta, hogy bizományosi szerződések kötésével folyamatosan bővíti az árusítandó termékkört.

## 7. Az első tapasztalatok

Paradox módon az első piaci tapasztalatokat már a webshop hivatalos elindítása előtt megszereztük, ugyanis a demo állapotban lévő webáruház felületén már annak megjelenésétől kezdve számos rendelést leadtak, annak ellenére, hogy az oldalon nagy betűkkel felhívtuk a figyelmet arra, hogy az oldal jelenleg nem éles, a rendelések pedig nem fognak teljesülni. Ezek a hónapokkal a hivatalos indulást megelőző „rendelések” sokat segítettek abban, hogy képet kapjunk a felhasználói igényekről. Megerősítettek minket abban, hogy lesz igény a webshopra és hogy a megrendelőink számára értéket képviselnek azok a 10-20 évvel ezelőtti kiadású, kereskedelmi forgalomban már nem kapható helytörténeti témájú művek is, amelyek a nyitó árukészletet jelentették.

A hivatalos indulást követő hónapokban különösen nagy érdeklődés mutatkozott a webshop iránt. A helyi média (Szentendre TV, Szentendre és Vidéke újság) segítségével igyekeztünk több csatornán is felhívni a potenciális érdeklődők figyelmét arra, hogy a szentendrei vagy a dunakanyari vonatkozású könyvek online is megvásárolhatók. A marketing kampány eredményeként nagyságrendileg heti 2-5 rendelés futott be az első hónapokban.

## 8. Összegzés

Jelentős értéket jelentett az intézmény számára, hogy a webshop nyomán javult a könyvtárról kialakult kép mind az olvasóink, mind pedig a könyvtáros szakma szemében. A webshop hozzájárult ahhoz, hogy a Hamvas Béla Pest Megyei Könyvtár egy innovatív, folyamatos megújulásra képes intézmény képét sugallja magáról.

A webshop indítása bebizonyította számunkra, hogy kis informatikai tudással, de sok elszántsággal és tenni akarással saját erőből is képes lehet egy közkönyvtári kollektíva létrehozni egy könyvtári webáruházat. A legfőbb nehézséget inkább a piacorientált hozzáállás bevezetése jelentette a közreműködő munkatársak körében. A már megszokott munkafolyamatok átszervezése, új szempontok szerinti újragondolása azonban pozitív hatással volt a munkatársak szemléletének frissítésére is.



## Bibliográfia

1. ANDERSON, Chris: *Hosszú fark - A végtelen választék átírja az üzlet szabályait.* Budapest, HVG kiadó, 2007.
2. DIPPOLD Péter: *A long tail és a könyvtárak. Hogyan kapcsolódhatunk a „hosszú fark” jelenséghez?* KISZL Péter, CSÍK Tibor, szerk. *Valóságos könyvtár - könyvtári valóság.* Könyvtár- és információtudományi tanulmányok 2018. (Budapest, ELTE BTK Könyvtár- és Információtudományi Intézet, 2018 167–179.).  
<https://doi.org/10.21862/vkkv2018.167>
3. BAJNOK Tamara és BOGNÁRNÉ LOVÁSZ Katalin és FEHÉR Miklós és HORVÁTH Adrienn és MÉSZÁROSNÉ MERBLER Éva és TOPÁR József és TÓTH Máté., *A könyvtárak minőségi működésének értékelési rendszere* (Budapest, Könyvtári Intézet, 2019). [https://ki.oszk.hu/sites/default/files/dokumentumtar/kmer\\_atdolg.pdf](https://ki.oszk.hu/sites/default/files/dokumentumtar/kmer_atdolg.pdf) (2022. június 17.)



A kiadás hagyatéka / a hagyaték kiadása:  
A Régi Magyar Költők Tárának hálózati kiadásáról<sup>1</sup>

Etlinger Mihály  
Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet  
[etlinger.mihaly@abtk.hu](mailto:etlinger.mihaly@abtk.hu)

Hernády Judit  
PPKE BTK Irodalomtudományi Doktori Iskola  
[hernady.judit@gmail.com](mailto:hernady.judit@gmail.com)

Vadai István (1960–2018) munkássága szorosan összefonódott a kéziratos és nyomtatott Balassi-szöveghagyomány bonyolult kérdéskörével. A régi szövegek filológiai-textológiai vizsgálata mellett azonban Vadai István a legkorszerűbb digitalizálási módszerek iránt is élenken érdeklődött. Az alábbiakban a régi magyar versek forráskiadásában meghatározó szerepű Régi Magyar Költők Tára digitalizálásában betöltött kezdeményező szerepét és az e folyamattal összefüggő hagyatéki anyag sorsát ismertetjük, betekintést nyújtva a munkálatok jelenlegi állásába is.

A Régi Magyar Költők Tára-sorozat (a továbbiakban RMKT) digitális kiadása több szempontból problematikus. Egyrészt azért, mert a 16. századi művek kiadása 1880-ban kezdődött. A kiadások 1930-ban megszakadtak, az azóta eltelt közel 100 évben nemcsak újabb művek és források kerültek elő, hanem jelentős szakirodalommal és újfajta filológiai eszközökkel is gazdagodtunk, amelyek természetesen a szövegek újrakiadásáért kiáltanak. Ezeket vette figyelembe a sorozat 1990-től indított „új folyama”, amely a korábbi, Szilády Áron-féle kiadások után szövegkritikai szempontból is megújult, a 17. századi sorozat modernebb filológiai módszereit alkalmazta. A sorozat tehát kettéválik a digitális kiadás módszertanában is.

A másik probléma az, hogy a 16. századi RMKT kiadása a mai napig nem fejeződött be. Egy teljes (Bogáti Fazakas Miklós zsolttárai), illetve egy kiegészítő kötettel (kiadatlan gyülekezeti énekek) máig adós a sorozat, így annak digitalizációja sem lehet végleges.

Hogy mégis jobban átlátható és használható online kiadás és szövegbázis jöhessen létre a körülmények ellenére, azt Vadai István (1960–2018) hagyatékának köszönhetjük. A 17. századi sorozat rekonstruált változatát nagy alaposággal ellenőrizte, így annak kereshető PDF-változatát már közre is adta a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézete és a DigiPhil.<sup>2</sup>

Vadai a 16. századi szövegek egységes kiadásán is dolgozott, hagyatékának e része is az Intézethez került. A korpuszon egyszerre két koncepció mentén is dolgozott: a hagyományos RMKT-köteteket felismertette, majd javította OCR-hibáit, emellett a még ki nem adott művek átíratát is elkészítette, majd a már kiadott szövegekkel együtt a 16. századi *Repertorium*<sup>3</sup> számozása alapján létrehozta az általa ismert összes 16. századi vers szövegbázisát. Munkálkodása egészen nagy időtávot lefed: 2002 és 2014 között készítette el Word-átíratát.

---

1 A tanulmány a Digitális Örökség Nemzeti Laboratórium támogatásával készült.

2 *Régi magyar költők tára XVII. század* <http://digiphil.hu/context:rmkt-17>, (2022. 05. 27.) A kötetek első, automatikus optikai karakterfelismerésen átment digitalizált változata (<https://szovegtar.iti.mta.hu/hu/sorozatok/rmkt-xvii-szazad/>) manuálisan nem ellenőrzött, ezért a szövegréteg több helyen hibás.

3 *Répertoire de la poésie hongroise ancienne: A régi magyar vers leltára a kezdetektől 1600-ig*, <https://f-book.com/rpha/v7/> (2022. 05. 27)



A 16. századi anyag kapcsán kijelenthető, hogy egészen vegyes minőségű kiadásokból és forrásokból dolgozott, így Vadai átíratok alaposságuk és pontosságuk ellenére sem felelhetnek meg egy kritikai kiadás elvárásainak. Ez fakad egyrészt abból, hogy az RMKT-sorozatban már megjelent művek esetében sem találkozunk a hagyatékban kritikai jegyzetekkel, saját változtatásait sem indokolta. Másrészt pedig abból, hogy a kritikai kiadásban meg nem jelent műveket elsődlegesen más korábbi kiadásokból írta át, és amelyeknek nem volt modern kiadása, azokat a rendelkezésére álló források alapján közölte. Az ilyen, kritikai kiadással nem rendelkező versek esetében emiatt egészen heterogén forráshalmazzal dolgozott, ezek azonosítása túlnyomórészt megtörtént. Sokszor használta forrásul korábbi kiadások MEK-es változatait, de többször tetten érhető, hogy a Varjas Béla szerkesztette *Balassi Bálint és a 16. század költői* kétkötetes népszerűsítő kiadást is alapul vette az RMKT-ban meg nem jelent históriák és énekek közléséhez.<sup>4</sup> Az online vagy felismertetett nyomtatott kiadások szövegeit természetesen ellenőrizte, ugyanakkor ellenőrzésének szempontjait, esetleges másodlagos forrásait nem rögzítette. Bármennyire is volt Vadai István a modern filológia és textológia egyik legnagyobb képviselője és alakítója, szempontrendszerének és átírási/javítási elveinek ismerete elengedhetetlen feltétel lenne ahhoz, hogy az RMKT-sorozathoz mérhető kritikai szövegkiadásként foglalkozzunk vele.

A hagyaték ezen része ráadásul több szövegkritikai problémát is magában hordoz. Bogáti Fazakas Miklós nagyszámú, kritikai kiadásban meg nem jelent műve esetében legtöbbször Gilicze Gábor szövegkritikailag egészen önkényes és pontatlan kiadását veszi alapul.<sup>5</sup> Alapos vizsgálat után kiderült, hogy Vadai elődjének hibáit több helyen ugyan javítja, de olyan változatok segítségével, amelyek a Bogáti-szöveg-hagyományban nem feltétlenül rendelkeznek szövegkritikai értékkel. És mivel nem a legjobb változatnak vélt *Péchi Simon-énekeskönyvből* dolgozott, átíratok, ha pontosabbak is a korábbiaknál, nem pótolhatják az RMKT-vállalkozás tervezett Bogáti-kötetét.

Vadai szövegbázisát modernizált, ejtéstükröztető átíratok alkotják zömmel, ám akad néhány betűhű változat is. Ezek modernizálása is szerkesztői feladat, amely Vadai átírási elveinek ismerete és érvényesítése nélkül nem lehetséges. Éppen ezért a rögzítési elveket deduktív módon a hagyaték szövegeiből kell kinyernünk.

Vadai forrásainak vizsgálata alapján leginkább arra lehet következtetni, hogy munkájának szövegkritikai igénye eltörpül elsődleges célja mellett, mely az összes 16. századi verses mű digitalizált változatának elkészítése volt. Ezért a hagyatékból készülő adatbázis alapja is erre épül: a DigiPhil-munkacsoport Vadai átíratok ellenőrzését és a Text Encoding Initiative ajánlásának megfelelő jelölőnyelvi átíratban (XML-ben) rögzíti,<sup>6</sup> változatainak forrásait minden esetben jegyzetben hivatkozza. A korpusz és a készülő adatbázis lehetőséget ad arra, hogy különböző szempontok alapján szűrjük és vizsgáljuk a 16. századi költészetet. Ahhoz, hogy az elsődlegesen a szövegből kinyerhető információkat közöljük, a következő metaadatkategóriákat vezettük be: RPHA-szám, cím, alcím, dallamjelzés, mottó, vers, versszak, verssor, üres sor, kolofon, akrosztichon. Emellett szükségesnek láttuk a saját szerkesztői jegyzeteink mellett Vadai saját jegyzeteit illetve a kódoló által felismert szövegkritikai megjegyzéseket

4 Varjas Béla szerk., jegyz., *Balassi Bálint és a 16. század költői* I-II (Budapest: Szépirodalmi Könyvkiadó, 1979).

5 Bogáti Fazakas Miklós, *Magyar Zsoltár*, s. a. r. Gilicze Gábor (Budapest: OSZK, 2009), <https://mek.oszk.hu/07900/07965/07965.htm> (2022. 05. 27.)

6 TEI Consortium, „TEI P5: Guidelines for Electronic Text Encoding and Interchange”: <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (2022. 05. 11.) A TEI kapcsán lásd még: Bobák Barbara, Fellegi Zsófia, Palkó Gábor, *Szövegtárolás és a Text Encoding Initiative*, <https://teach.dariah.eu/course/view.php?id=40> (2022.05.11); Palkó Gábor, „Digitális filológia: számítógép anyaszerepben”, *Filológiai Közöny* 61, 2. sz. (2015), 187–199, 195–198.

feltüntetni, hogy a kiadás transzparens lehessen a Vadai-variánsok és a létrejövő saját változatok esetében. Ez a gesztus, ha nem is pótolja a szövegek kritikai kiadását, fogódzkodat nyújthat a közölt szövegek helyes felhasználásában és egyfajta forráskritikai szemlélet kialakításában.

Az RMKT 17. századi sorozatát a Magyar Tudományos Akadémia Irodalomtudományi Intézete indította el 1959-ben, Klaniczay Tibor és Stoll Béla szerkesztésében. A kritikai kiadás gyűjtőkörébe tartozik minden magyar nyelvű verses szöveg, az egyetlen sornyi lírai daraboktól egészen a több részes, olykor igen terjedelmes epikus költeményekig. A versanyagot az egyes kötetek különféle csoportosításban (korszak, tematika vagy felekezeti hovatartozás szerint) adják közre, illetve vannak egy vagy több szerzői életművet közlő sorozatdarabok is. A szövegek betűhű átíratban, szövegkritikai és tárgyi jegyzetekkel ellátva kerülnek kiadásra, a szerzői életműveknél életrajzi áttekintés is olvasható.<sup>7</sup> A sorozat darabjai néhány éve az Irodalomtudományi Intézet textológiai portálján kétrétegű PDF fájlok formájában is elérhetővé váltak.<sup>8</sup> Az elektronikus hozzáférés lehetővé tételében azonban nem fedti le a digitális kiadás fogalmát, a nyomtatott kiadások informatizálásával együtt járó lehetőségek ugyanis jóval túlmutatnak a csupán digitális formátumra való áttétel kínálta, igen szűkös kereteken.<sup>9</sup>

A sorozat jelölőnyelvi átírásának megkezdése szintén Vadai István hagyatékának képezi részét. A projekt azonban Vadai váratlan halála miatt megakadt, s csupán több évnyi szünet után, 2021-ben indult újra a DigiPhil-ben, a fájlok azonban már nem az Arany kritikai kiadások informatizálása során kidolgozott egyszerű XML formátumban,<sup>10</sup> hanem szabványos TEI XML-ben készülnek. A munkálatok alapanyagát azok a szövegfájlok képezik, amelyek OCR-hibáit még Vadai István javította ki. Mivel ezek sem mentesek a kisebb hibáktól, minden kötet kritikai szöveganyaga egy további manuális ellenőrzési fázison is keresztülmegegy, hogy az OCR-hibákból eredő betűtévesztéseket a lehető legminimálisabbra tudjuk korlátozni.

Az RMKT kódolása során számos nehézséggel talákoztunk. Ezek főként a sorozat versanyagának nagyfokú változatosságával, illetve a szövegkritikai jegyzetekkel állnak kapcsolatban. E problémák közül néhányat már Vadai Istvánnak is szembesülnie kellett, az ezzel kapcsolatos kommunikáció a Trello kollaborációs eszközön, levélváltás formájában zajlott, e felületen Vadai kommentjei a ma napig visszaolvashatók. Hamar kiderült, hogy a modern kiadásoknál alkalmazott automatikus szöveggenerálás az RMKT esetében nem valósítható meg, mivel a kritikai kiadás negatív apparátust alkalmaz, vagyis a szövegtanúk teljes anyaga vonatkozásában az eltérő nyomdai betűkészletből eredő különbségek nem ismerhetők meg, így nem is reprodukálhatók. A gépi szöveggenerálás ezért a valóságban soha nem létező, a főváltozat betűkészletét a szövegtanúkéval összekeverő szövegváltozatokat tudna csak létrehozni. E probléma megoldásához minden egyes szövegtanút egyenként kellene ellenőrizni, illetve a negatív apparátusból eredő gépi tévesztéseket is manuálisan kellene kijavítani, ezt azonban a szöveganyag terjedelme lehetetlenné teszi.

---

7 Bisztray Gyula, Klaniczay Tibor, Nagy Lajos, Stoll Béla, szerk., *A tizenöt éves háború, Bocskay és Báthori Gábor korának költészete*, Régi magyar költők tára: XVII. század 1 (Budapest: Akadémiai Kiadó, 1959), 5–14.

8 <https://szovegtar.iti.mta.hu/hu/sorozatok/rmkt-xvii-szazad/> (2022.05.11.)

9 Az informatizálás fogalma kapcsán lásd: Fellegi Zsófia, „Átmenet az analógból a digitális filológiába: médiumváltás? Madách Imre: Az ember tragédiája”, *Helikon* 67, 1. (2021.), 84–102, 91–92. Vö. Maróthy Szilvia, „Tudományos szövegkiadások a hálózaton: Áttekintés”, *Irodalomtörténeti Közlemények* 122, 5. sz. (2018), 617–633.

10 Fellegi Zsófia, Palkó Gábor, „Arany-kéziratok és kritikai kiadások közzététele az Arany János Emlékévben”, *Helikon* 66, 1. sz. (2020), 82–98, 85–90.

Ezt az alapvető problémát ezért csak úgy tudjuk kiküszöbölni, hogy teljes szöveggel kizárólag a főváltozatot jelenítjük meg, a szövegváltozatokat pedig egyenként, a megfelelő sorhoz rendelve tüntetjük fel:

```
<lg>
<1>Nem lehete, mert az Isten</1>
<1><app><lem wit="#KK">Megmarad diczsőségében,</lem><rdg wit="#W1">düczőségében</rdg></app></1>
<1><app><lem wit="#KK">Ő törvény székit készíteti</lem><rdg wit="#W1 #W2">késcitete</rdg></app></1>
<1><app><lem wit="#KK">Az itéletre letéteti.</lem><rdg wit="#W1">itiletre</rdg></app></1>
</lg>
<lg>
<1>Dolgában tart igasságot,</1>
<1>Elöhivja ez világot,</1>
<1>Es igassággal mindeneket</1>
<1>Megitél minden nemzeteket.</1>
```

1. ábra: Egy versrészlet jelölőnyelvi átírata

A DigiPhil korábbi projektjeihez képest az RMKT abban is újdonságot jelent, hogy az egyes kötetek anyaga gyakran többretegűen strukturált. Azaz tartalmaz szerzői köteteket, versciklusokat, illetve az ismeretlen szerzőjű vagy a csak néhány művel rendelkező szerzőkhöz köthető darabok gyakran különféle szerkesztői versblokkokba vannak sorolva. A kódolás során fontos, hogy ezek a rétegek is láthatóvá és követhetővé váljanak.

Az egyik legnagyobb nehézséget a szövegkritikai jegyzetek jelölőnyelvi átírata képezi. A *digital born* kiadásokkal szemben az eredetileg nyomtatvány formájában megjelent kritikai kiadások utólagos informatizálása esetében gyakran adódik probléma abból, hogy a jelölőnyelvi átírást többnyire nem a szövegkiadás készítője végzi. A sokszor több évtizeddel – de olykor akár egy évszázaddal – korábbi textológiai gyakorlatot tükröző nyomtatványok filológiai jelölései ugyanis a kódolás sajátos igényei szerint nem mindig értelmezhetők megfelelő pontossággal. De az esetleges apparátusbeli hibák vagy hiányosságok is komoly gondot jelenthetnek a helyes jelölőnyelvi átírat elkészítése során.

Az RMKT vonatkozásában különösen élesen jelentkezik ez a probléma, hiszen a sorozat több szerkesztő munkájának eredményeként jött létre. Ráadásul a feldolgozott versanyag sajátosságai is gyakran igen eltérőek. A sorozat ezért nem használ teljesen egységesített jelölési rendszert, a szövegkritikai jegyzetek jelöléseit az egyes szerkesztők gyakorlata és a versanyag egyedi jellemzői alakítják ki. Ebből kifolyólag lehetséges, hogy amit az egyik szerkesztő jelöl, a másik esetleg nem, vagy ha igen, akkor máshogyan: más szóhasználattal vagy eltérő részletességgel. A jelentékenyebb szerzői életművek például a közköltészeti darabokkal szemben jellemzően nagyobb részletességű leírást kapnak. További problémát jelent, hogy a kötetekben a jelölések pontos meghatározása sem található meg, vagyis nem tudjuk, hogy például a *javítva ebből*: vagy az *előbb*: kitétel egészen pontosan milyen kéziratbeli jelenségeket fed le. Ez a kritikai kiadás szokásos használata során természetesen nem okoz problémát. A jelölőnyelvi átírat elkészítése szempontjából azonban igencsak megnehezíti a helyzetet. Ha ugyanis nem tudjuk rekonstruálni, hogy a módosítás pontosan hogyan történt (áthúzás, átsatírozás, fölé írás stb.), akkor a jelölőnyelvi átírat pontatlanná, vagy esetleg félrevezetővé is válhat, például törlést jelölhet ott, ahol a kéziratban nincs áthúzás, vagy ha van, annak kiterjedése esetleg a valóságostól eltérő. A kódolandó anyag mennyisége miatt az utólagos ellenőrzés természetesen ebben az esetben sem lehetséges. Az alapvető cél tehát az, hogy a lehetőségekhez képest a kötet jelölését leghűbben tükröző átíratot készítsünk.

Ez az említett példa esetében azt jelenti, hogy a *javítva ebből*: kitétel által jelölt szerzői vagy másolói változat nem <del>, hanem <corr> jelölőbe kerül, az alábbi módon:

```
<choice ><orig>szívemek</orig><corr corresp="#corrector">szívet</corr></choice>
```

Mint látható, a javítás végrehajtójának a jelölése ekkor a *corresp* attribútummal történik, mivel a <corr> jelölőnél a *hand* attribútum nem használható, a TEI szintaxisa ugyanis ezt a megoldást nem teszi lehetővé. E kereten belül a kötetekben helyenként előforduló *azonnali javítás* jelölőnyelvi átírása az *instant* attribútum segítségével valósítható meg: <choice ><orig>szívemek</orig><corr corresp="#corrector" instant="true">szívet</corr></choice>

A szövegkritikai apparátus jellegéből kifolyólag vált szükségessé például az emendáció eltérő kódolása egyetlen kézirat vagy nyomtatvány, illetve több szövegtanú megléte esetében. Ha csupán egyetlen szövegforrás áll rendelkezésre, vagyis a javítás ugyanazon a kézíraton (vagy esetleg nyomtatványon) szerepel, a kód magában foglalja a korrekció előzményét is: <choice><orig>hák</orig><corr>ház</corr></choice>.

Több forrás esetében viszont a főváltozat szövegében csupán egy <corr> tag áll, a javítás előzménye pedig a *readingek* között marad olvasható, ebből a szempontból sértetlenül megtartva a nehezebben félreérthető eredeti jelölési rendszert. Ha a közlés alapjául szolgáló kéziratban szereplő adatot a <corr>-hoz kapcsolódóan tüntetnénk fel, akkor vagy feleslegesen megkétszereződne a szövegrész (a *readingek* mellett az emendáció kódolásánál is megismétlődve), vagy – a kritikai apparátusból történő elhagyása esetében – eltűnne a változatok közül, ami valószínűleg csak még átláthatatlanabbá tenné például a Rozsnyai-kéziratváltozatok egyébként is nehezen áttekinthető szövevényét.

Előfordul, hogy a szerkesztő az alapszövegbeli változattal szemben a szövegtanúk egyező variánsát tekinti helyesebb alaknak, s ezért ezeknek megfelelően javít a főváltozaton. Ezekben az esetekben a szövegforrás jelölése a *source* attribútummal vagy az <orig>, vagy a <corr> tagba kerülhetne. Ez a megoldás azonban az előzőekben már érintett kettős közlés vagy elhagyás problémáját vetné fel. Az emendáció alapjául vett változat kiválasztása, majd elhagyása az apparátusból azonban itt nagyfokú bizonytalanság és komoly adatvesztés nélkül nem is volna kivitelezhető.

Természetesen az RMKT sem mentes a hibáktól, és ezek olykor szintén megnehezítik a kódolás folyamatát. Előfordul például, hogy hiányzik a szövegtanút jelző szigla, vagy éppen hibásan szerepel. E hibák jelölésére fog lehetőséget nyújtani az új DigiPhil-ben alkalmazott annotációs réteg, amelynek segítségével a kutató anélkül fűzhet megjegyzéseket az egyes szöveghelyekhez, hogy megváltoztatná annak tartalmát, vagyis a jelölőnyelvi átíratot.

Összegzésként elmondható, hogy a fentikben vázolt, Vadai István munkáját alapul vevő informatizálási munka túlmutat az elektronikus hozzáférés biztosításán, a szövegek könnyebb, korszerűbb és számos új felhasználói lehetőséggel bővített kezelését teszi lehetővé. A két sorozat elkészült kötetei a DigiPhil oldalán lesznek elérhetők, úgynevezett „Free Access – no reuse”<sup>11</sup> felhasználási joggal. A TEI XML fájlok a lektorálást követően a kutatóközönség számára a DigiPhil nyilvános GitHub felületén, azonos feltételekkel válnak hozzáférhetővé.

---

11 <http://www.europeana.eu/rights/rr-f/> (2022.09.10.)



„Ki a fenének kell collstok?”<sup>1</sup>  
A digitális szöveg rejtett mértékegységei

„Who the hell needs measuring stick?”  
The hidden units of digital text

Varga Emese  
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ  
[varga.emese@oszk.hu](mailto:varga.emese@oszk.hu)

Makkai T. Csilla  
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ  
[makkai.csilla@oszk.hu](mailto:makkai.csilla@oszk.hu)

**Abstract:** How to reveal the results of the philological research behind a digital edition? These data are often hidden in the text of letters and critical notes, but when grouped and quantified they provide significant information.

In our presentation, we will present the results of machine-assisted research based on the digital critical edition of Móricz Zsigmond's correspondence and novel corpus, outlining how new contexts in the correspondence and novel corpus become visible with machine assistance. It allows us to look at our corpus and the author's oeuvre from a different perspective.

**Keywords:** data visualization, digital edition, graph visualization, stylometry, hungarian literature, 20th century literature, corpus analysis

**Absztrakt:** Hogyan lehet megmutatni egy digitális kiadás mögött álló filológiai kutatás eredményeit? Ezek olyan adatok, amelyek a levélszövegben, kritikai jegyzetekben, sokszor rejtve vannak jelen, csoportosítva és számszerűsítve viszont szignifikáns információval bírnak. Előadásunkban a Móricz Zsigmond-levelezés digitális kritikai kiadása, illetve a regénykorpusz digitális kiadása alapján gépi eszközökkel készült, további kutatásokra támaszkodó eredményeit prezentáljuk, felvázolva, hogy gépi támogatással milyen új összefüggések válnak láthatóvá a levelezésben. Ezek segítségével más perspektívából tudunk rátekinteni a korpuszunkra, innen tovább lépve pedig az írói életműre.

**Kulcsszavak:** adatvizualizáció, digitális kiadás, gráf alapú vizualizáció, stilometria, magyar irodalom, 20. századi irodalom, korpuszelemzés.

Egy digitális szövegkiadás készítése során szükségszerűen keletkeznek számszerűsíthető adatok, legyenek azok tulajdonnevek, évszámok, kritikai jegyzetek, műfaji kategóriák. Viszont az adatok közötti összefüggések felvázolása is hozzátartozik a kutatómunkához annak ellenére, hogy azok sokszor hosszú, nehezen átlátható Excel táblákba tömörülnek.

Tanulmányunk legfőbb kérdése, hogy milyen adatvizualizációk segítségével lehet megmutatni egy digitális szövegkiadás mögött álló filológiai kutatás eredményeit, hogyan lehet

---

1 Móricz Zsigmond, *Tükör IV.*, lásd: <https://dhupla.hu/text/o:mzs-tukor.tei.4>.



átláthatóbbá tenni egy strukturált adathalmazt, milyen eszközök léteznek az összefüggések megmutatására.

A továbbiakban három különböző adatvizualizációt szeretnénk prezentálni: két gráf alapút és egy stilometriai elemzés eredményeit. Ez esetben nem volt szükség *collstokra*, gépi támogatású méréseket végeztünk nagyméretű, digitális szövegtörzseken.

A Móricz-műhely keretein belül elkezdett közös munka fókuszában állt a PIM (Petőfi Irodalmi Múzeum) Kézirattárában található Móricz-különgyűjtemény kutathatóvá tétele. Ebből a sokéves munkából született meg a [Tükör](#) című világháborús jegyzetkötetek digitális kiadása, illetve a [Móricz-levelezés](#) digitális kritikai kiadása, melynek 2.0-s verziója olvasható a 2021-ben létrehozott [dHUpIa](#) nevű digitális bölcsészeti platformon. Többfajta technika többféle korpuszt igényel: a gráf alapú adatvizualizációk kiindulópontjaként a Móricz-levelezés digitális kritikai kiadása szolgál, a stilometriai elemzést pedig a Móricz-regénykorpuszon végeztük, melyet az Országos Széchényi Könyvtár [Magyar Elektronikus Könyvtárából](#) értünk el.

Munkánkat tehát a megszámlálható adatok elemzése jellemzi. A nyelv statisztikai alapú vizsgálatának különböző módzatai már a korai irodalomtudományi elemzésekben, például szógyakoriság mérésekben is felfedezhetőek. Hasonlóképpen a stilometriai mérések első kísérletei közé tartoznak a szerzőségi kérdések statisztikai alapú vizsgálatai, mellyel már 1851-ben foglalkoztak. Augustus de Morgan angol tudós és Thomas Mendenhall amerikai fizikus kutatásaikban a szavak hosszúságát és gyakoriságukat vizsgálták.<sup>2</sup> Ebből is látszik, hogy a számolás és az olvasás műveletei nem függetlenek egymástól, mivel a nyelv természetének statisztikai alapú rendeződése is szignifikáns információkkal bír.<sup>3</sup>

A digitális irodalomtudomány pedig erre a jellegre alapozva, a szöveg részeinek számszerűsíthetőségéből indul ki. A gépi támogatással végzett elemzések nemcsak felgyorsítják ezeket a statisztikai alapú vizsgálatokat, hanem a korábbiakhoz képest több és más jellegű adatra építve adnak lehetőséget irodalmi jelenségek értelmezésére. Természetesen a gépi, statisztikai alapú elemzések nem függetlenek a hagyományos filológiai elemzésektől, hanem új perspektívaként kiegészítik azokat.

A digitális kiadásoknak fontos szerepük van az adatok hatékony számszerűsítésében, hiszen azok önmagukban is adatbázisok – a metaadatolt, annotált szövegkiadások kereshető, lekérdezhető, számszerűsíthető adatokat tartalmaznak. A Móricz levelezés kiadás szűrési funkciója (ún. facettált kereső) listázott, számszerűsített adatokat jelenít meg, melyek segítségével az olvasó hamarabb eljut az információhoz, amit keres.

A tudományos kutatások egyre inkább adatközpontúvá válnak, a webes kultúra pedig egyre vizuálisabbá.<sup>4</sup> A nagyméretű adathalmazok értelmezhetősége, befogadhatóvá tétele önmagában még nem valósul meg, ehhez különböző vizualizációs eszközök használata szükséges, melyek szemléletessége folytán a kutatási eredmények jobban megmutathatóak.

## Gráf alapú adatvizualizációk

A genealógiai kutatás elkerülhetetlen egy teljességre törekvő levelezés kiadás készítésekor, ugyanígy az író kapcsolatrendszerének átláthatóbbá tétele is a kutatás lényeges részét képezi. Ennek érdekében indítottuk el Móricz Zsigmond leveleinek publikálása mellett

2 David HOLMES – Judit KARDOS, *Who Was the Author? An Introduction to Stylometry*, *Chance* 16. (2003/2.), 6. lásd: <https://id.art1lib.org/ireader/42465305>.

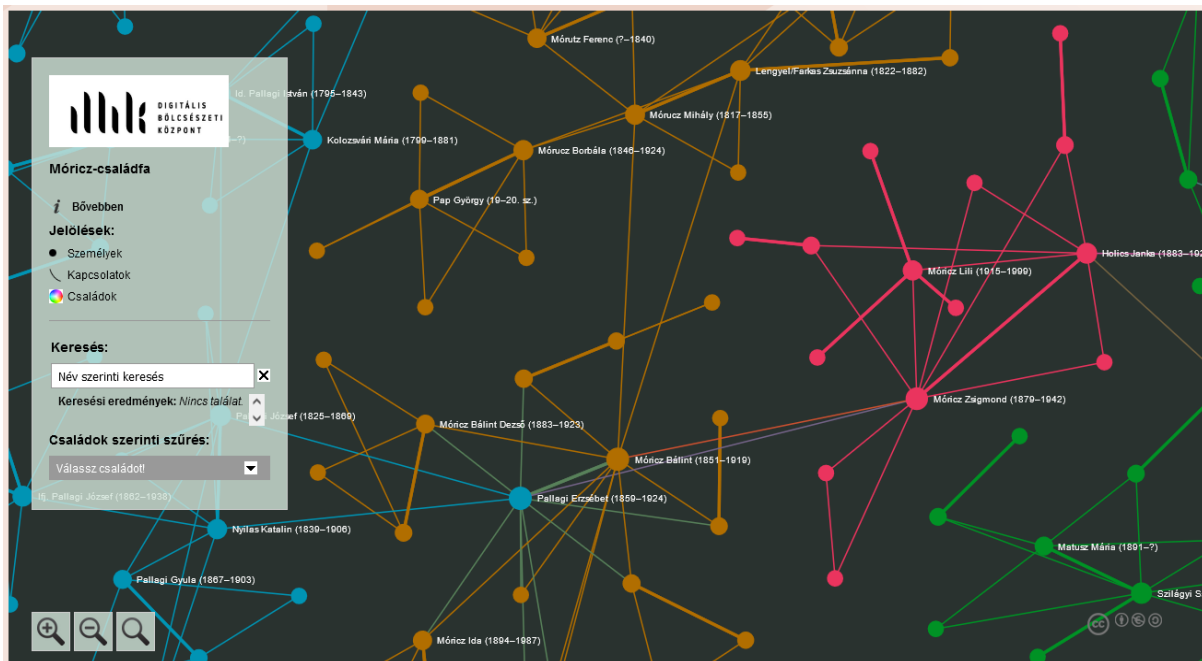
3 SZEMES Botond, *A digitális bölcsészet mint kódfejtés. A nyelv statisztikai vizsgálatának hagyományáról*, Műút 2021/079. Lásd: <https://www.muut.hu/archivum/37522>.

4 MARÓTHY Szilvia. 2020. *Digitális bölcsészet. A szövegtől az adatig*, Akadémiai, Budapest, 2020. Lásd: <https://doi.org/10.1556/9789634545781>.

Móricz családfájának és szakmai kapcsolati hálójának kutatását és interaktív vizualizációs megjelenítését.

Hogy minél pontosabb legyen a családfához készülő adatbázisunk, továbbléptünk a levelezésen és meglévő adatainkat összevetettük a PIM Kézirattárában lévő Móricz-hagyatékkal és a Móricz szakirodalom adataival, valamint különböző internetes, digitalizált adatbázisokban kutattunk. (Huntéka, [PIM Névtér](#), [Arcanum](#), [Hungaricana](#), [FamilySearch](#), [OSZK-Gyászjelentések](#))

A kutatómunka eredményeként a családtagok száma a Móricz ágon 35-ről 62 főre nőtt, az anyai, Pallagi-ágon 25-ről 90 főre gyarapodott. Holics Janka oldalán a Holics ág 45-ről 142-re, a Szklenár család pedig 55 főről 101-re növekedett. Így a családfa összesen 409 nevet tartalmaz. Ennek megjelenítésére a [Gephi](#) nyílt forráskódú, hálózatelemző és -ábrázoló szoftvert választottuk, amely gráf alapú adatvizualizációk készítésére alkalmas. Ez az adatvizualizációs eszköz egy kapcsolatrendszer olyan egyszerűsített ábrázolását teszi lehetővé, mely az aktuális szempontoknak megfelelően kiemeli az összefüggéseket, és ez által az eredmények könnyebben értelmezhetővé válnak<sup>5</sup>. Előnye, hogy interaktívá, dinamikussá tehető, kutatható, nagy adatmennyiség esetén is jól átlátható (lásd: 1. ábra), a kapcsolódási pontok mentén többféle irányból elindulva lehet végigjárni a hálózatot és összerakni a puzzle darabkait.

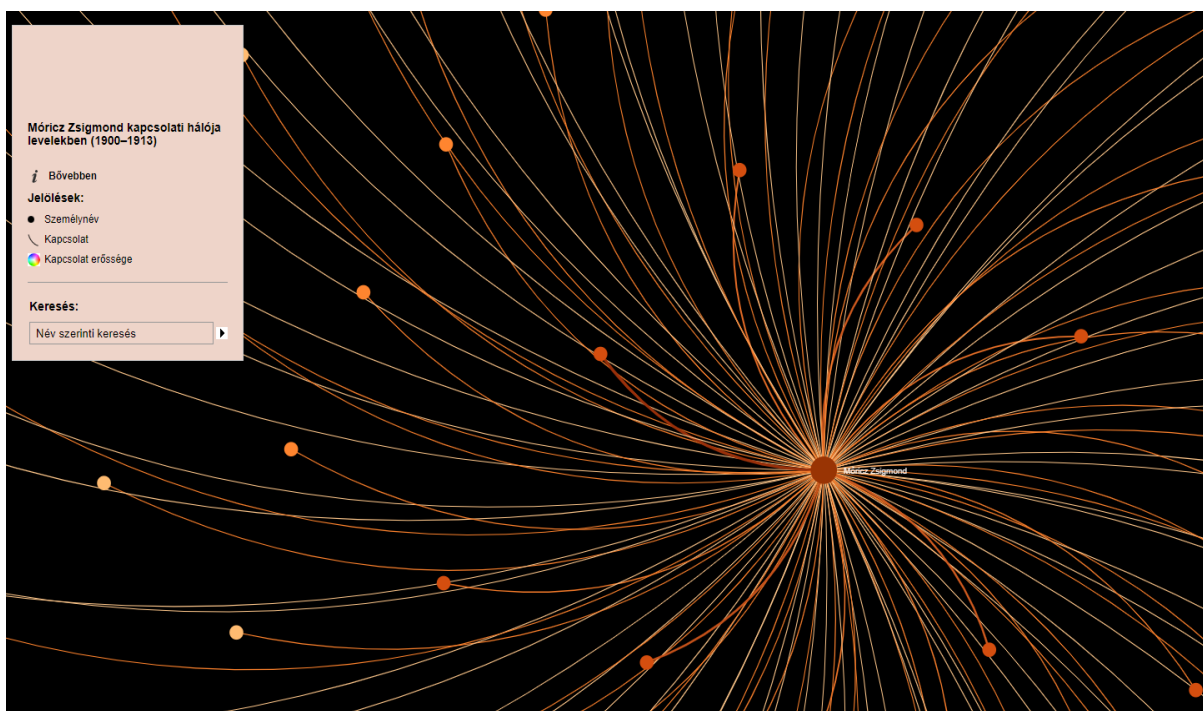


1. ábra: Móricz családfa vizualizáció  
<https://dhupla.hu/s/dataviz/moricz-csalad/>

A családtagok között kétfajta kapcsolatot jelenítettünk meg: a gyerek-szülő és házastársi viszonyt. Számokkal súlyoztuk a kapcsolat erősségét, melyet a vizualizáción az élek vastagsága mutat - a gyerek-szülő kapcsolatot a vékonyabb, a házastársi kapcsolatot a vastagabb él jelöli. A bal oldali panelben külön rászűrhetünk a családi ágakra - a Móricz, a Pallagi, a Szklenár és a Holics család mellett Móricz szűk családi ágát is külön klaszterbe rendeztük. Emellett konkrét

5 KRÉSZ Miklós – TÓTH Attila, *Gráf alapú adatbányászat és vizualizáció: egy esettanulmány = Tudományos és művészeti műhelymunkák: új utakon a Pedagógusképző Kar című konferencia előadásai*, szerk. Tóth Szergej, Szegedi Tudományegyetem, Szeged, 2009, 358. Lásd: <http://acta.bibl.u-szeged.hu/id/eprint/64233>.

személynevekre is rákereshetünk. Minden családtag rendelkezik egy személyi adatlappal, ami a jobb oldali panelben ugrik elő, ahol az illető személyi adatai, a születési hely- és ideje, a halálozási hely- és ideje, valamint a közvetlen hozzátartozóinak a nevei láthatóak.



2. ábra: Mórícz kapcsolati háló adatvizualizáció  
<https://dhupla.hu/page/kreativ/moricz-kapcsolatihalo>

Hasonló séma alapján, ugyancsak a Gephi szoftverben készült a kapcsolati háló is. (Lásd 2. ábra) 520 kézirat alapján 136 levelezőpartner állítja fókuszba az író, amely a pályakezdés tizenhárom évét sűríti egybe. A középpontban Mórícz Zsigmond látható, körülötte a levelező pályatársak állnak.

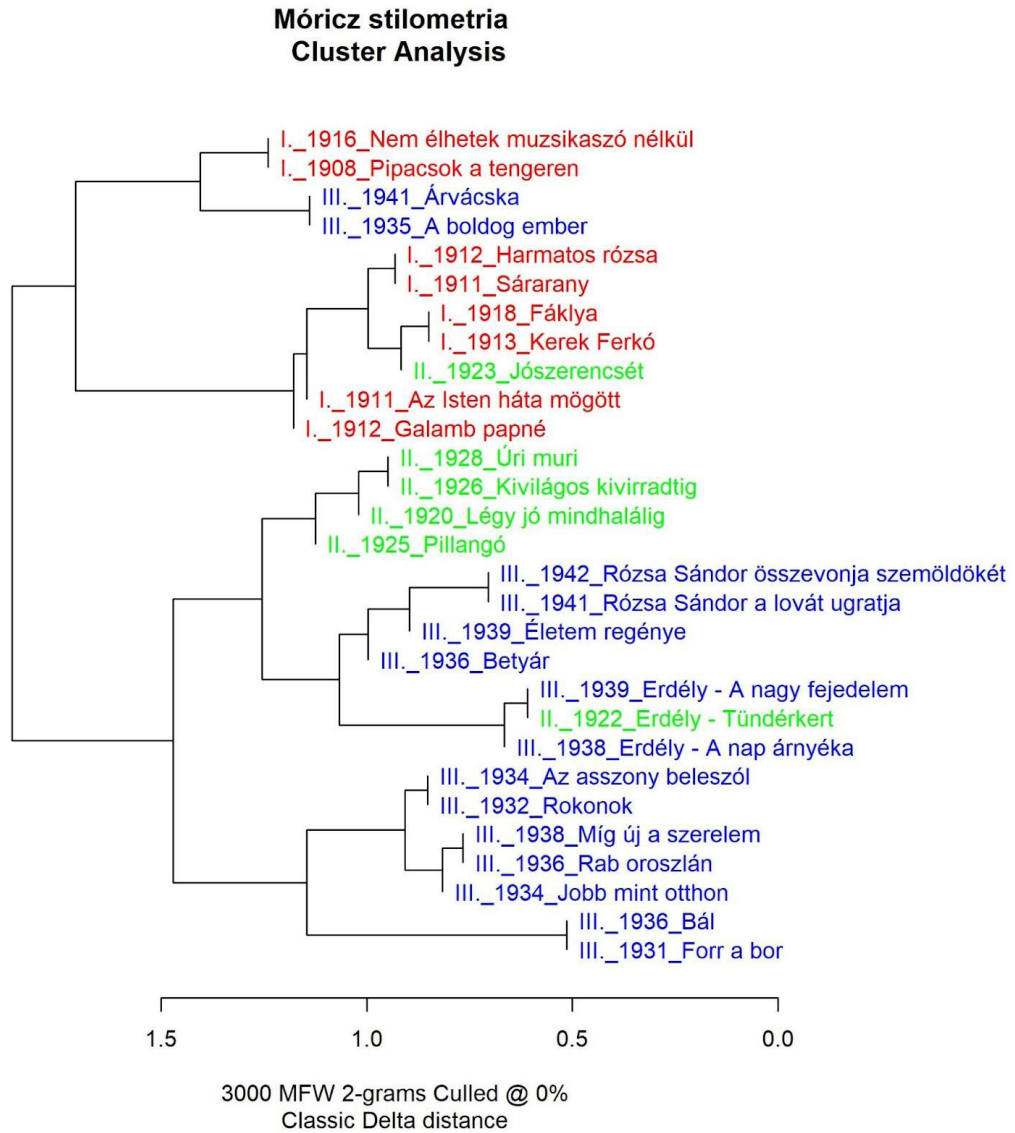
## Stilometriai elemzés

A stilometria nagyméretű szövegtömbök stílusának statisztikai alapú elemzésére alkalmas, melyben a szövegek összehasonlítása számszerűsített nyelvi jellemzők és különféle távolságmértékek alapján történik. Ez által tetten érhető az ún. „szerzői ujjlenyomat”, mely a szövegek tematikus szintje alatt létezik.<sup>6</sup>

Az elemzés szorosan kapcsolódik a Digitális Bölcsészeti Központban korábban készült, Mórícz-levelezésen alapuló [stilometriai vizsgálat](#)hoz, melyet továbbgondolva készítettük el a regénykorpuszon alapuló elemzést. Magyarországon találtunk példát korpuszok gépi elemzésére, stilometriai vizsgálatokra is, viszont olyan vizsgálat, amely egyetlen szerző életművének feldolgozásával foglalkozik, csak elvétve olvasható. Ebből kiindulva, arra tettünk kísérletet, hogy Mórícz Zsigmond regényeit statisztikai alapon elemezzük, és megnézzük, hogy felállítható-e egy hozzávetőleges időrend, láthatóvá válnak-e eltérések az életműben, megcélozható-e valamiféle korszakolás? Emellett, méréseink eredményét összevetettük a Mórícz-szakirodalom vonatkozó részeivel, abból a tapasztalatból kiindulva, hogy a legtöbb stilometriai elemzés megáll a vizualizációnál és az irodalomtörténeti következtetések csak ritkán kerülnek előtérbe.

6 SZEMES Botond, *l. m.*

Az R programozási nyelv erre alkalmas *stylo* csomagjával végeztük az elemzést, mely távolságmétrikák alapján állít fel szógyakorisági listákat, majd ezeket egymáshoz viszonyítva hoz létre egy dendrogramot. Ahhoz, hogy a vizsgálatunk minél megbízhatóbb eredményeket adjon, többféle beállítási paraméterrel kísérleteztünk. A legjobb eredményeket végül a klasszikus delta, az Eder delta és a Canberra távolságmértékek adták.

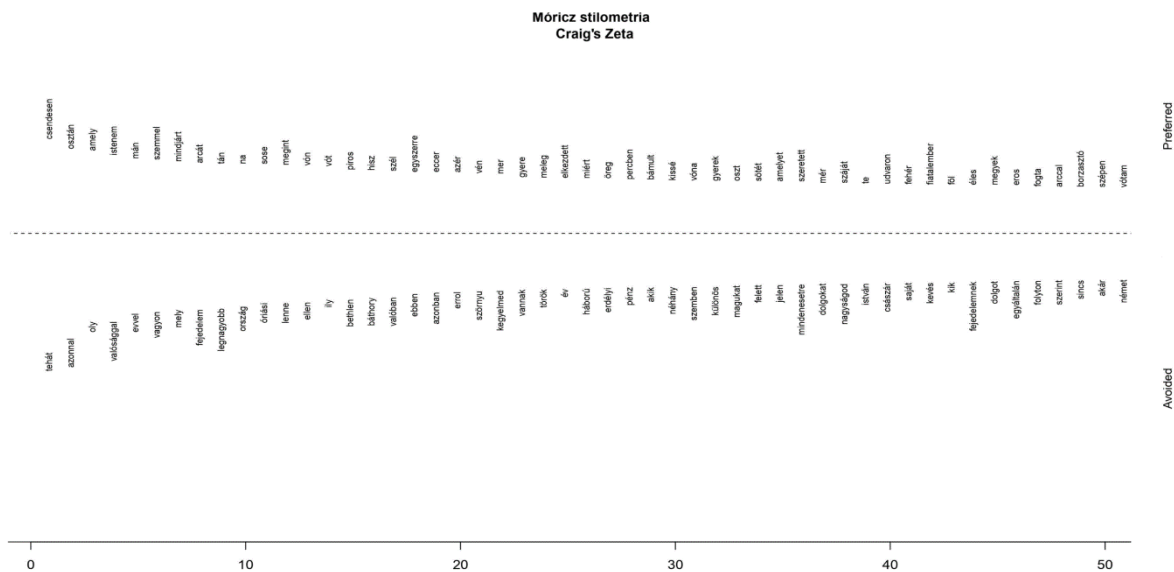


3. ábra: Klaszteranalízis klasszikus delta távolsággal

Az itt látható, 3-as ábrán a szerzőazonosítási célra létrehozott klasszikus delta távolsággal mértünk. A beállított paraméterek szerint a 3000 leggyakoribb szó-bigramot (tehát egymást követő lexémákat) vette figyelembe a szoftver, illetve nem használtuk a cullingot, ami azt jelenti, hogy nem szűrtünk ki egyetlen szó-bigramot sem a korpuszból.

Előzetesen a korpuszt három alkorpuszra bontottuk, ez tükröződik a színek szerinti besorolásban. Az 1908 és 1918 között írt szövegek pirossal, az 1920-as években írt szövegek zölddel, az 1930 után írt regények késsel jelöltek. A mérés után nagyjából kialakul egy időrend, a 10-es, 20-as, 30-as évek szerint. Az 1908 és 1923 között kiadott regények szorosan egymás mellett helyezkednek el, ugyanígy az 1930-as évek szövegei is egy nagyobb tömbbe rendeződnek. Érdekes viszont, hogy nem mindegyik szöveg került közel a "kortársához", illetve egészen eltérő tematikájú szövegek is egymás mellé rendeződtek.

A 3-as számú ábra alapján az látszik, hogy *A boldog ember* (1935) és az *Árvácska* (1941) című szövegek több rokonságot mutatnak a korai művekkel, mint a későbbiekkel. A móríci írásmódra mindvégig jellemző az elhangzott és lejegyzett dialógusokból való építkezés, mely módszer végletekig vitt megtestesítője lett végül *A boldog ember* és az *Árvácska*.<sup>7</sup> Az a tény, hogy a tárgyalt szövegek közelebb kerültek az életmű első szakaszában írt regényekhez, arra enged rákérdezni, hogy a korai művek erőteljesebben közvetítik-e a hangzó nyelvet, az élőbeszédet. Emellett megfigyelhető még, hogy a történelmi regények (Erdély-trilógia és Rózsa Sándor regények), szorosan illeszkedtek egymáshoz, melynek oka a létrehozott archaizáló nyelvben keresendő.



4. ábra: Stilometriai elemzés Oppose függvénnel

Ehhez a kísérlethez a korpuszunkat az eddigi eredményeink alapján osztottuk két részre, vagyis a korai és a kései művekre, ahol a korszakhatár az 1920-as évek közepén húzódik meg. Az R-ben az *oppose* függvényt használtuk, amely két szövegtörzset összehasonlítására alkalmas. A függvény két szólistát állít elő egy diagramon, az egyik az első korpusz kifejezetten preferált szavait tartalmazza a másik korpuszhoz képest. A másik szólista pedig az első korpusz kifejezetten került szavait tartalmazza a második szövegtörzshöz képest.

A mérésből láthatóvá válik, hogy az első csoportba tartozó művek preferált lexémái között gyakoriak a tájnyelvi elemek: 'mán', 'vót', 'eccer', 'azér', 'oszt', 'má' stb. Ezek nem kizárólag tájnyelvi jellegük miatt érdekesek, hanem a hangzó nyelv móríci alkalmazása szempontjából is, mely szövegépítésének egyik jellemzője.

A második csoportban túlsúlyba kerülnek az anyagi világhoz, történelemhez, társadalomhoz kapcsolható kifejezések. Előfordulnak a 'pénz', 'fejedelem', 'ország', 'kegyelmed', 'háború', 'erdélyi', 'nagyapod', 'császár', 'dolog' kifejezések, melyek különböző szemantikai mezőkön helyezhetők el, mégis összekapcsolhatóak. Mintha az írói tekintet a premier plámban készült portré műfajától eltávolodva, átkapcsolt volna a nagytörténelmi történő ábrázolásmódra.

<sup>7</sup> Cséve Anna, *A tükör mint hallucinációs fal (Móríci Zsigmond kéziratos könyveiről)*, *Alföld* 2005/9., 74–82., 79.; Szilágyi Zsófia, *Móríci Zsigmond*, Kalligram, Pozsony, 2013, 221.





## Kitekintés

A digitális szövegkiadások metodikája és a szövegelemző szoftverek párhuzamosan fejlődnek, kölcsönhatásban állnak egymással, éppen ezért a kutatás során felmerült kérdésekhez újabb megoldási lehetőségek is társultak. Tervezzük a munka folytatását, pontosabban azt, hogy a digitális kiadások készítésével párhuzamosan a stilometriai vizsgálatokat is folytassuk a Móricz szövegeken. Célunk egy konszenzusháló létrehozása a Móricz-korpusz elemeiből, melyhez új eszközöket, illetve a *stylo* újabb funkcióit fogjuk tesztelni. Ugyanígy a levelezéskiadás alapján készült kapcsolati háló bővítése és az entitástár gazdagítása is a további munkafolyamatok részét képezi.

## Bibliográfia

CSÉVE Anna, *A tükör mint hallucinációs fal (Móricz Zsigmond kéziratos könyveiről)*, Alföld 2005/9., 74–82., 79.

CSÉVE Anna, MAKKAI T. Csilla, TÖRÖK Sándor Mátyás, VARGA Emese, VÉTEK Bence: *Rokonok a gráfon*. <https://pim.hu/hu/digitalis-bolcseszeti-kozpont/moricz/rokonok-a-grafon>

David HOLMES – Judit KARDOS, *Who Was the Author? An Introduction to Stylometry*, Chance 16. (2003/2.), 6. lásd: <https://id.art1lib.org/ireader/42465305>.

KRÉSZ Miklós – TÓTH Attila, *Gráf alapú adatbányászat és vizualizáció: egy esettanulmány = Tudományos és művészeti műhelymunkák: új utakon a Pedagógusképző Kar című konferencia előadásai*, szerk. Tóth Szergej, Szegedi Tudományegyetem, Szeged, 2009, 358. Lásd: <http://acta.bibl.u-szeged.hu/id/eprint/64233>.

MAKKAI T. Csilla, VARGA Emese, Vétek Bence: *Egy életmű metszete. Stilometriai vizsgálat Móricz Zsigmond regényein* Szépirodalmi Figyelő, 2022/1.

MARÓTHY Szilvia. 2020. *Digitális bölcsészet. A szövegtől az adatig*, Akadémiai, Budapest, 2020. Lásd: <https://doi.org/10.1556/9789634545781>.

SZEMES Botond, *A digitális bölcsészet mint kódfejtés. A nyelv statisztikai vizsgálatának hagyományáról*, Műút 2021/079. Lásd: <https://www.muut.hu/archivum/37522>.

SZILÁGYI Zsófia, *Móricz Zsigmond*, Kalligram, Pozsony, 2013, 221.



## ITldata – Egy irodalmi adatbázis fejlesztése Wikibase alapon és ennek hasznosítása Kosztolányi Dezső forrásjegyzékénél<sup>1</sup>

Dobás Kata

MTA Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet  
[Dobas.Kata@abtk.hu](mailto:Dobas.Kata@abtk.hu)

Fazekas Júlia

MTA Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet  
[Fazekas.Julia@abtk.hu](mailto:Fazekas.Julia@abtk.hu)

### Absztrakt

Az ITldata-projekt célja egy olyan irodalomtudományos adatbázis kialakítása, mely hálózatos formában képes különböző kutatások adatait tárolni, megjeleníteni és rendszerezni. Az adatbázis Wikibase szoftverrel működik, a Wikidata struktúráját veszi alapul, ugyanakkor független tőle. Az adatbázisban tárolt információk SPARQL lekérdezésekkel rendszerezhetők, mely a tájékozódáson túl elősegíti a mostani és a jövőbeli kutatásokat. Az ITldata specifikációja úgy lett kidolgozva, hogy minél többfajta irodalomtudományos projekt adatainak tárolására szolgáljon, használható legyen könyvtári katalógusok, kéziratgyűjtemények, sajtóbibliográfiák számára egyaránt. Az ITldata ezentúl lehetőséget biztosít a különböző elemek annotálására, valamint összekapcsolható más digitális rendszerekkel, adatbázisokkal.

Az ITldata-ba elsőként fölvitt projekt Kosztolányi Dezső folyóiratokban megjelent munkáihoz kapcsolódik. Kosztolányi művei kritikai kiadásának munkálatai 2008-ban kezdődtek el. A forrásgyűjtő munka előrehaladtával kirajzolódott, hogy az életmű jelentős hányada periodikákban lelhető fel. Jelenleg 250 olyan időszak kiadványról tudunk, amelyben megjelent Kosztolányitól írás, a végső szám valószínűleg ennél jóval nagyobb lesz. Az eddigi eredményekből állítottuk össze a Forrásjegyzék első hat kötetét, amely 11 ezer tételt tartalmaz. Kosztolányi publikálási gyakorlatának jellemzője volt, hogy ugyanazt a művet újra megjelentette másik periodikában, kis változtatásokkal, esetleg új címmel, de nem volt ritka az sem, hogy ugyanaz a szöveg évekkel később ismét napvilágot látott. Ebből következett, hogy egy-egy mű publikálási hálózatának láthatóvá tételéhez adatbázisra volt szükségünk. 2018-tól a Koha könyvtári rendszerben, 2022-től pedig az ITldata-ban dolgozunk.

### Abstract

The ITldata-project aims to create a database for literary studies that collects, displays and organizes data for various research projects. The database uses the Wikibase software, and although it follows the structure of Wikidata, operates separately from it. ITldata primarily stores data about literary works or anything connected to literary studies. The stored information can be processed through SPARQL queries which not only help with navigating the database but can advance ongoing and future research projects as well. The specification of ITldata aims to collect and display data from literary collections of various nature, therefore, it will be useful for library catalogues, manuscripts

---

1 Jelen tanulmány a Digitális Örökség Nemzeti Laboratórium támogatásával készült.



collections, press bibliographies alike. Furthermore, ITldata presents an option to annotate all items, and it can be linked to other databases and networks as well. The first project that began to store its data in ITldata is concerned with the works of Dezső Kosztolányi. Work on a critical edition of Kosztolányi's texts began in 2008. As the fundraising work progressed, it became clear that a significant part of Kosztolányi's oeuvre could be found in periodicals. We currently know of 250 periodicals in which writings from Kosztolányi were published, the final issue is probably much larger. From the results so far, we have published the first six volumes of the Kosztolányi's bibliography, which contains 11,000 items. It was characteristic of Kosztolányi's publishing practice that the same work was republished in another periodical, with changes and possibly a new title, but it was not uncommon for the same text to be published again years later. We needed a database to make the publishing network of a work visible. From 2018 we worked in the Koha library system and from 2022 in ITldata.

## I. Az ITldatáról általában

Tanulmányunkban a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézete által fejlesztett ITldata adatbázist mutatjuk be, amely irodalomtudományi kutatások segítésére, azok adatainak tárolására és szemantikus kapcsolataik rendszerezésére jött létre. Az ITldata a DigiPhil projekt<sup>2</sup> keretében valósul meg. Az alapok kidolgozása után különböző kutatások adatainak a bevitelét kezdtük meg, melyek közül az első a Kosztolányi-forrásjegyzék volt. Az anyaggal való munka és a kutatókkal közös együttműködés során alakítottuk ki az adatbázis végleges struktúráját, azt a keretet, amelyben a jövőbeli kutatások adataival is dolgozni tudunk majd. Mostani cikkünkben előbb az ITldatáról szóló általános tudnivalókat és fejlesztési lépéseket, majd egy konkrét projekt, a Kosztolányi-forrásjegyzék esetén azt is bemutatjuk, hogyan vált maga az ITldata az adatrendszerezés eszközévé.<sup>3</sup>

### 1. Wikibase használata

Az ITldata egy olyan szemantikus adatbázis, mely (hasonlóan a Wikidatához) a Wikibase szoftvert használja. Az Irodalomtudományi Intézet DigiPhil csoportja közel tíz éve tesz közzé digitális kritikai kiadásokat, a kiadások adatgazdagításának támogatására háttéradatbázisként korábban egy nyílt forráskódú integrált könyvtári rendszert, a Koha-t használta. A rendszer személy- és földrajzi neveket, bibliográfiai adatokat, valamint kéziratok metaadatait tartalmazta. Az elmúlt évek tapasztalatai és a készülő új born digital kritikai kiadások igényei azonban rávilágítottak a könyvtári rendszerek kötöttségeire, ezért a cél egy rugalmas, digitális filológiai és egyéb kutatási igényeket támogató rendszer kiépítése lett. A Wikibase könnyen áttekinthető és használható struktúrával rendelkezik, mely jól alakítható különböző projektek elvárásainak megfelelően, folyamatosan bővíthető és fejleszthető. Az ITldata létrehozásakor már alkalmaztuk és a használatban folyamatosan alkalmazzuk a Wikibase különböző bővítményeit, kiterjesztéseit. Erre példa a *related items* nevű kiterjesztés, melynek segítségével származtatott állításokat tudunk megjeleníteni különböző elemeknél. Ezzel megkönnyítjük a kétirányú azonosítást, valamint a felhasználók számára

<sup>2</sup> <https://digiphil.hu>

<sup>3</sup> Az ITldata specifikációjának kidolgozását Fazekas Júlia, a Kosztolányi-forrásjegyzékkel való munkát Dobás Kata vezette, a tanulmány két részét is ennek megfelelően dolgoztuk ki.

is egyszerűen átláthatóvá válnak a kétirányú (például *testvér*) vagy ellentétes (például *apa – gyerek*) kapcsolatok.

Kiemelt fontossággal bírt számunkra, hogy a Wikibase segítségével egy többnyelvű adatbázist hozunk létre, mely nemzetközi projektekhez is használható. Különösen lényeges ez olyan kutatásoknál, melyek többnyelvű anyagot tartalmaznak. Alapvető elvárásunk jelenleg, hogy az adatbázis minden eleme rendelkezzen legalább magyar címkével és leírással, valamint angol címkével (nem-egyedi elemek esetén pedig leírással is) – így egy külföldi kutató számára szintén használhatóvá válik az adatbázis.

A Wikibase egy emberek és számítógépek által is jól olvasható és rendszerezhető adatbázis alapját adja. Ez bővíthető természetesen az elemek szintjén túl oldalak (vagy akár vitalapok) létrehozásával, mely szintén szerepel a terveink között egyes projekteknel. Így az adatbázis nemcsak adatok tárolását és rendszerezését segíti majd, de adott esetekben azok szöveges tartalmakkal (szócikkkel) való összekapcsolását is lehetővé teszi. Az ITIdata ezen túl külső azonosítókön keresztül kiválóan kapcsolható más adatbázisokhoz, így Wikibase alapú rendszerekhez (mint például a Wikidata), vagy könyvtári és egyéb rendszerekhez is (például a számunkra különösen fontos MTMT-hez, vagy az OSZK és az EPA katalógusához).

## 2. Elvárások és működés

Az ITIdata adatbázist úgy dolgoztuk ki, hogy képes legyen különféle, eltérő adatstruktúrával rendelkező projektek elemeit kezelni. Így törekedtünk arra, hogy bibliográfiai adatok, kéziratgyűjtemények és hagyatéki elemek, valamint sajtóbibliográfiák adatai egyaránt betölthetőek és rendszerezhetőek legyenek. Fontos volt, hogy az ITIdata kapcsolódjon könyvtári katalógusokhoz, de olyan viszonyok megjelenítésére is képes legyen, melyre azok nem feltétlenül biztosítanak lehetőséget. Gondolhatunk itt különböző szövegváltozatok közötti kapcsolatra, sajtóbeli- majd kötetes megjelenésekre, vagy akár arra, hogy egy szerző hogyan építette be egy kézirat részletét egy későbbi publikációba. Hasonló dokumentumok mellett kutatási projektek adatainak betöltését is szem előtt tartottuk, melyek segítségével összekapcsolhatók az egy korpuszhoz tartozó elemek, vagy jelölhető és lekérdezhető egyetlen projekt keretében végzett és az adatbázisba betöltött anyag.

Az ITIdatában minden elem látható a felhasználók számára, ezek pedig különböző módszerekkel rendszerezhetőek. A munka első szakaszában a megfelelő tulajdonságok összeállítására és specifikálására volt szükség, ezt a folyamatos, kutatókkal való együttműködés során fejlesztettük. Jelenleg több, mint 140 tulajdonság áll rendelkezésre az adatbázisban, de a bekapcsolódó projektek változatos természete miatt ez adott esetben bővíthet. A Wikidatából kiindulva létrehoztunk számos olyan elemet, mely több projekt számára alapvető fontosságú. Ide tartoztak többek között a leggyakoribb foglalkozások, irodalmi- és publicisztikai műfajok, dokumentumtípusok és nyelvek. Ezen túl az adatbázisban egy irodalmi személynévtér adatainak betöltésére is lehetőségünk nyílik, amely a szerzők életrajzi adatait tartalmazza. Ez a munka során bővílni fog, létrehozva egy egyedi és gazdag személynév-adatbázist. Az ITIdata földrajzi helyek adatainak tárolására szintén alkalmas, ez lényeges mind a személynevekhez kötődő helyszíneknél, mind a különböző publikációk megjelenési helyeinél. A földrajzi helyekhez koordinátákat is megadunk (főként a Wikidata alapján), ennek segítségével a térképes vizualizáció is lehetségessé válik.

### 3. ITIdata használati lehetőségei az irodalomtudományban

Az ITIdata adatok tárolására, kapcsolatok megjelenítésére kiválóan alkalmas, valamint segíti majd a későbbiekben a készülő digitális kritikai kiadásokban a különböző elemek beazonosítását. Konkrét kutatási kérdéseknél azonban még lényegesebb, hogy hogyan tudjuk a tárolt adatokat rendszerezni és felhasználni. Egy SPARQL végponton<sup>4</sup> keresztül lehetőségünk van lekérdezéseket írni, megadott paraméterek alapján elemezni az adatbázisba betöltött adatokat. Egyszerű lekérdezésekkel így például kinyerhetjük az adatbázisban jelenleg szereplő összes személynevet, műfajt (kategóriákkal, alá-fölé-rendeltséggel együtt) vagy egyetlen szerző publikációit. Az ITIdata segítségével azonban ennél bonyolultabb kérdésekre is választ kaphatunk, táblázatokat, grafikonokat hozhatunk létre. Ha például azt szeretnénk megtudni, hogy egy szerző arányosan milyen műfajokban alkotott, ezt ábrázolhatjuk akár egy buborékábrán is. Lekérdezhetjük akár azt is, hogy milyen volt a szerzőségi köre egy folyóiratnak, valamint az adott személyek milyen gyakorisággal és mértékben publikáltak. De akár térképen is ábrázolhatjuk egy szerző leveleinek feladási helyét, mely adott esetben segítheti az utazásairól kapott képünk kiszélesítését.

Az ITIdata jelenlegi formájában számos projekt számára használható és hasznosítható. Cél, hogy folyamatosan fejlesszük és bővítsük, és mind a hazai, mind a nemzetközi irodalmi (vagy ahhoz részben kötődő) kutatások számára jelentős eszközzel tudjunk szolgálni. Az adatbázisban jelenleg több, mint húszezer elem szerepel, azonban *A magyar irodalomtörténet bibliográfiája* teljes adatállományának betöltését követően<sup>5</sup> ez a szám a tízszeresére fog nőni.

## II. Az ITIdata és a Kosztolányi-forrásjegyzék

Jelen tanulmány második részében azt foglalom össze, hogy a felvázolt, Wikibase adta lehetőségekhez hogyan illesztettük a Kosztolányi-forrásjegyzék adatait, milyen előzetes tapasztalataink voltak, milyen kérdésekbe ütköztünk, illetve milyen egyéni igényeknek kellett megfelelnünk a specifikáció során.

### 1. A Kosztolányi-forrásjegyzék

Amikor elkezdődött Kosztolányi Dezső műveinek kritikai kiadása,<sup>6</sup> a filológiai munka első fázisaként Kosztolányi életében megjelent műveit, azok különböző megjelenéseit szeretnénk volna összeszedni, katalogizálni. Ez a nyomtatott sajtóban megjelent szövegeket jelentette, a kéziratok mellett. 1903 és 1936 közé tehető az az időszak, amikor Kosztolányi publikált, tehát a kutatómunka erre az időintervallumra koncentrált. Budapesti, vidéki, határon túli lapokat szemléztünk, az összegyűlt adatokat a Forrásjegyzékekben folyamatosan publikáltuk.<sup>7</sup> Mennyiségileg egy nem várt adatállománnyal találkoztunk. *A Forrásjegyzék* hat kötetében *jelenleg* 11 ezer tétel körüli nagyságrendben található adatok. Számtalan eddig nem ismert Kosztolányi-művet is találtunk a rengeteg ismeretlen szövegközlés mellett. Kosztolányi publikálási gyakorlatának megfelelően több mű látott napvilágot akár évekkel később ugyanazon címmel egy másik folyóiratban, de némileg eltérő változatban, vagy évtizedekkel

4 Elérhető: <https://query.ITIdata.abtk.hu/> (hozzáférés: 2022. 06. 18.)

5 Lásd Fellegi Zsófia előadása, «A digitális filológia infrastruktúrái. A DigiPhil megújulásáról», Networkshop 2022. április 22.

6 Hivatalosan 2008-tól indultak meg a Kosztolányi életművének kritikai kiadásával foglalkozó kutatócsoportok munkálatai. Azóta számtalan pályázat keretében és kutatócsoportban folytatódott a munka.

7 *Kosztolányi Dezső napilapokban és folyóiratokban megjelent írásainak jegyzéke 1–6.*, szerk. Arany Zsuzsanna, majd Dobás Kata, Ráció Kiadó, Budapest, 2008–2018.

később jelent meg teljesen más címmel ugyanazon mű.<sup>8</sup> A Kosztolányi Forrásjegyzékek ennek megfelelően egymásra épülnek.

A publikálási gyakorlatra példa: a Komáromi Újságban 1916-ban megjelent *Ilike az asztalnál* címmel egy novella. A szöveg *Ozsonna* címmel 1911-ben már a Hét című lapban is napvilágot látott, némileg eltérő változatban. Ugyanez a mű 1924-ben megjelent a Bácsmegeyi Naplóban is, *Aurélia napja* címmel, szintén változtatásokkal. A szöveg *Ilike az asztalnál* címmel került be két Kosztolányi-kötetbe, az 1911-es *Bolondokba* és az 1908-as *Boszorkányos estékbe*:

FJ6/491. Kosztolányi Dezső: „Ilike az asztalnál”, *Komáromi Újság*, XVI. évf. 31. sz., 1916. aug. 3., 1–2. [Műfaj: novella; első sor/mondat: „Ilike, négy éves. Ilike az asztalnál ül a keresztmama ebédlőjében.”; lásd FJ1/209: „Ozsonna” + FJ2/272: „Aurélia napja” + FJ3/88: „Ozsonna” és 650: „Aurélia napja” is; kötetben: *Bolondok*, 38–43, *Boszorkányos Esték*, 126–130.]

FJ1/209: Kosztolányi Dezső: „Ozsonna”, *A Hét*, XXII. évf. (1. köt.) 19/1104. sz., 1911. máj. 7., 298–299. [Műfaj: novella; első sor/mondat: „Piroska négy éves. Piroska az asztalnál ül a nagymama ebédlőjében.”; kötetben: B, 38–43., BE, 126–130: „Ilike az asztalnál”.]

FJ2/272. Kosztolányi Dezső: „Aurélia napja”, *BMN*, XXV. évf. 129. sz., 1924. máj. 11., 11. [Műfaj: novella; első sor/mondat: „Piroska négy éves. Az asztalnál ül nagymama ebédlőjében.”; lásd FJ1/209: „Ozsonna”; kötetben: B, 38–43., BE, 126–130: „Ilike az asztalnál”.]

stb.

Egy másik példa, ami szintén az anyagaink természetéhez tartozó jelenség: egy 1924-es, Pesti Hírlapban megjelentetett tárca a pár évvel később kiadott *Édes Anna* című regényben kapott helyet, de némileg eltérő változatban. Vagyis nem egy regényrészlet publikálásáról beszélhetünk jelen esetben, hanem egy annál lazább kapcsolatról.

FJ1/1131. Kosztolányi Dezső: „Együtt, mindig együtt”, *PH*, XLVI. évf. 89. sz., 1924. máj. 11., 4. [Rovat: Vasárnap; műfaj: tárcanovella; első sor/mondat: „Vasárnap, sötét alkonyatkor ballagok a hegy alján, budai kis utcákban, egyikből ki, másikba be.”]

→ *Édes Anna* című regény

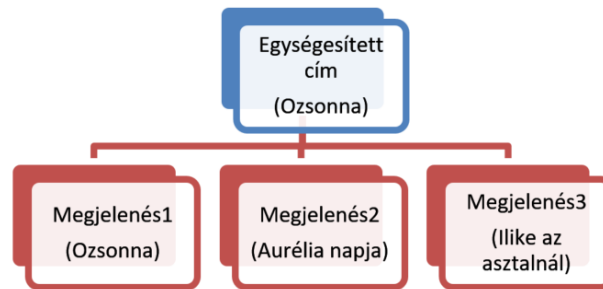
Ahogy az talán márből a pár példából is érzékelhető, az adatstruktúra, amivel szembesültünk, hálózatos megjelenítésre volt ítélve. Már csak azért is, mert ez a struktúra szétfeszítette a nyomtatott bibliográfiák kereteit. A kutatók számára egyre több időt vett igénybe, hogy a *Forrásjegyzékek*ben található adatokat összefésüljék, a visszautalások mentén egy-egy mű szövegvariánsait feltérképezzék. Két elvárásunk volt tehát a leendő adatbázisunkkal szemben azon túlmenően, hogy természetesen szerettük volna, ha az alapvető bibliográfiai adatokat is rögzíteni tudjuk. Az egyik, hogy a kötetes megjelenéseket a megfelelő módon tudjuk megjeleníteni, a kötetben belüli szövegek oldalszámra lebontva megjelenhessenek, és az eltérő címváltozatok is láthatóak legyenek. A másik pedig annak a publikálási gyakorlatnak a megjelenítése volt, ami az egyes szövegváltozatok alapján megmutatkozott.

---

8 A jelenség koránt sem egyedülálló voltára példa a korszakban: Wirágh András, *Nyelven innen, határokon túl: Hálózatos magyar irodalom 1920 és 1929 között*, Irodalomtörténeti Közlemények 2020/2., 259–268. ([http://real-j.mtak.hu/14954/2/itk\\_2020\\_2.pdf#page=115](http://real-j.mtak.hu/14954/2/itk_2020_2.pdf#page=115))

## 2. Első megoldás: Koha

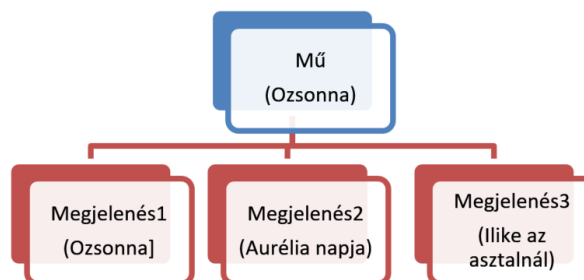
2017-től a DigiPhil csoporttal együttműködésben az általuk használt Koha könyvtári rendszerben kezdtünk el dolgozni.<sup>9</sup> A Koha-ban létrehoztunk egy úgynevezett egységesített címet, ami lehetővé tette, hogy az egyes tételeket összekapcsoljuk. Ez egy olyan elem volt, amelyhez felvezettük az adott forrásjegyzékes tétel címét, majd ugyanezt az egységesített címet adtuk az összekapcsolandó tételünknek is. Vagyis ezen a mezőn keresztül tudtuk összekötni az egyes szövegváriánsokat.



A rendszer működött, némi hiányossága azonban volt. A kötetes megjelenésnél nem lehetett feltüntetni azt, hogy az adott mű milyen oldalszámokon található, a megjelenítésnél pedig csupán egy listát láttunk a kapcsolódó tételekről. Meg kell azonban jegyezni, hogy a Koha minden merevsége ellenére is létrehozta azt a struktúrát, vagyis az egységesített címmel azt a típusú hálózatot, amit a későbbiekben is meg kívántunk őrizni. Ez tehát a projekt szempontjából nagyon fontos állomás volt.

## 3. Második megoldás: ITIdata

Az egységesített cím struktúráját követtük az ITIdata specifikációja során is. Az egységesített cím a **mű** nevet kapta. A konkrét példánál maradva: az *Ozsonna* című novella mű osztályú elemként is az *Ozsonna* címet kapja, a leírást, címkét követően pedig hozzá tudjuk kapcsolni a két konkrét megjelenést is, vagyis a kötetbelit és a folyóiratost.



A tesztelési fázisban felmerült az a kérdés, hogy szükséges-e minden egyes tételnél létrehozni mű típusú elemet, hiszen nem minden tételnek van kötetes megjelenése, és számos olyan tétel is van, ami nem jelent meg többször egyik periodikában sem. Végül arra a pragmatikus megoldásra jutottunk, hogy ha egy tételnél nincs több megjelenés, akkor csak a konkrét folyóiratost megjelenés lépéseit követjük, és egyetlen elemet hozunk létre. Ha újabb

<sup>9</sup> <http://biblio.digiphil.hu/>



szövegváltozatot találunk, akkor hozzuk létre a mű osztályú elemet, és ehhez kapcsoljuk a változatokat.

A kötetes megjelenéseket a **szerkesztett gyűjteményben**, a konkrét, tehát nyomtatott sajtóbeli megjelenéseket pedig a **kiadás** tulajdonsággal kapcsoljuk a mű elemhez, ami összeköti a különböző változatokat. Vagyis itt, a mű típusú elemnél láthatóak a különböző megjelenések, felhasználóként innen lehet eljutni a konkrét folyóiratos megjelenések leírásáig vagy a szöveg kötetben elfoglalt helyéig.

**OZSONNA** (Q1862)  
Kosztolányi Dezső, novella  
Illike az asztalnál | Aurélia napja

**ÁLLÍTÁSOK**

Nyelv	Címke	Leírás	Más néven
magyar	Ozsonna	Kosztolányi Dezső, novella	Illike az asztalnál Aurélia napja
angol	Nincs címke	Nincs leírás	

**osztály, amelynek példája** mű

**szerző entitás** Kosztolányi Dezső

Az adatbázisban a sajtóorgánum, illetve a Kosztolányi-kötet leírása is megtalálható. A periodikák esetében a saját sajtótörténeti kutatásaink eredményeit publikáljuk. Az adatok felvitelét követően több mint 250 lap leírása lesz olvasható az adatbázisban, köztük számos olyané is, amely nem kutatható vagy hozzáférhető ma Magyarországon.

**kiadás**

**szerkesztett gyűjteményben**

- Bolondok**  
oldal: 38-43  
cím: Illike az asztalnál (magyar)
- Boszorkányos esték**  
oldal: 126-130  
cím: Illike az asztalnál (magyar)

Az *Édes Anna* című regényhez kapcsolható tárcatípusú esetre a **kapcsolatban áll** tulajdonságot dolgoztuk ki. Ez bizonyult kellően tágnak ahhoz, hogy az irodalmi téma vagy motívum beledolgozását egy adott műbe érzékeltetni lehessen.



#### 4. Összegzés, további tervek

Az általunk kezelt adatsorok mögött egy kutatócsoport évtizedes kutatómunkája áll. A kutatómunkából következően sok esetben saját, akár más országos könyvtári intézményektől eltérő adatokat viszünk fel az ITIDatába, legyen szó egyes Kosztolányi-kötetek datálásáról vagy ismeretlennek vélt periodikák beazonosításáról.

Jelenleg három munkatársam van, akik az adatfeltöltést végzik: Páji Gréta, Bodovics Ticiána és Gönczi László. Közülük ketten részt vettek a forrásgyűjtés munkálataiban is, így az ő sajtótörténeti tudásuk, valamint Kosztolányi-szövegismeretük miatt a specifikációban is részt vállalnak, illetve az anomáliák feloldásában is segítenek.

Munkánk jelenlegi fázisában egy nagyon izgalmas helyzet állt elő. Az a bibliográfiai anyag, amit mi bevallottan segédanyagként szántunk a kritikai köteteken dolgozóknak, elkezdett önálló életet élni, és újabb irodalomtudományos határterületek bontakoztak ki előttünk. A jelenlegi adatbázis struktúra ugyanis nem csupán filológiai kutatásokra alkalmas, hanem például olyan intézményi összefüggésekre is ráirányíthatja a figyelmet, amelyek eddig rejtve voltak előttünk. Az egyébként még mindig kissé hanyagolt sajtótörténeti kutatásokhoz is hozzájárulhat az ITIDdata, de az is láthatóvá válhat Kosztolányi publikálási hálózatának vizsgálata során, hogy az egyes lapok között milyen kapcsolatrendszerek működtek a 20. század elején.

Kézai Simon Program – digitális családi fotóarchívum  
„Kézai Simon Programme” - digital family photo archive

Sörény Edina  
Digitális Jólét Nonprofit Kft. (Budapest)  
[soreny.edina@djnkt.hu](mailto:soreny.edina@djnkt.hu)

## Összefoglalás

A Kézai Simon Program célja, hogy megőrizze a magántulajdonban lévő családi fotókhoz kapcsolódó pótolhatatlan emlékeket és történeteket, amelyeket elsősorban az idősebb családtagok őriznek. A Program keretében létrehozott digitális családi fotóalbumok hozzájárulnak a 19-20. századi Magyarország történetének széleskörű megismeréséhez. Mindezt beépítve nemzeti kultúránk örökségébe és kollektív emlékezetébe, segítve ezzel a nemzeti történelem elemzését és kutatását.

## Abstract

The “Kézai Simon Programme” aims to preserve the irreplaceable memories and stories associated with privately owned family photos that are preserved primarily by older family members. The digital family photo albums created within the framework of the Program contribute to the extensive knowledge of the history of Hungary in the 19th and 20th centuries. We incorporate all of this into the heritage and collective memory of our national culture, assisting in the analysis and research of national history.

**kulcsszavak:** digitalizáció, digitális archívum, családkutatás, Kézai Simon Program, Kézai Hálózat, online szolgáltatások, értékmérés

**keywords:** digitization, digital archive, familysearch, Kézai Simon Programme, Kézai Network, online services, salvage

## 1. Bevezető

Az Innovációs és Technológiai Minisztérium támogatásával a Digitális Jólét Nonprofit Kft. 2020. december 9-én indította el a Kézai Simon Programot, új országos közösségi mozgalomként. A Program alapvető célja, megőrizni a magántulajdonban lévő, családi fotókon rögzített, pótolhatatlan emlékeket és a hozzájuk tartozó történeteket, amelyeket főként az idősebb családtagok őriznek. Mindezt beépítve a nemzeti kultúránk örökségébe és kollektív emlékezetünkbe, segítve a nemzet történelmi elemzéseket és kutatásokat.

## 2. Kihívások: régen és most

A Program névadója, a 13. századi Magyarország leghíresebb krónikása Kézai Simon, aki legfőbb művével, a Gesta Hungarorummal, megalkotta a magyarok krónikáját. A Program az ő történelmi munkássága jegyében indult el azzal a feladattal, hogy létrehozzunk egy platformot a családok digitális krónikáinak.

Elmondhatjuk, hogy egészen az őskortól kezdődően az emberek fontosnak tartották, hogy a velük megtörtént eseményeket valamilyen módon feljegyezzék. Kezdetben a barlangok falain, majd kőtáblákon, agyagtáblákon, papirusztekercseken, vagy éppen pergamenen, és később papíron, illetve ma már digitális tartalomhordozókon is. A kódexek illusztrációi és

képek, valamint fali festmények sokasága őriz korabeli eseményeket, sőt családok életével kapcsolatos történeteket. Később a fotózás megjelenésével a családok élete egyre inkább dokumentálttá vált, minden társadalmi csoport körében elterjedt annak igénye, hogy megőrizték és megmutassák utódaik számára a fontos családi eseményeket, és mára elmondhatjuk, hogy régi fotókkal teli dobozokban magányos családi fényképek milliárdjai gyűjtik a port a szekrényekben és a padlásokon szerte a világon.

Míg szüleink és nagyszüleink főként papír alapú fotókat gyűjtöttek össze az évtizedek alatt, addig mi számtalan digitális eszközön tárolunk az életünk valamely eseményéhez kapcsolódó digitális felvételt. Ezek számossága akár a több ezret is meghaladhatja és változatos formátumokban pihen egy-egy floppy-n, CD-n, DVD-n, és számítógépek winchesterén.



1. ábra: Gesta Hungarorum<sup>1</sup>



2. ábra: Digitális képtároló<sup>2</sup>

A 20. század nagy történelmi eseményei a családok életében is jelentős változásokat hozott. A háborúk, az újratekzés, majd a városiasodás és a generációk mobilizálódása a családi emlékek megőrzésére és hatással volt. A régi papíralapú fotók és akár családi iratok az évek során folyamatosan sérülhetnek, megsárgulhatnak, sőt el is kallódhatnak, alkalomadtán meg is semmisülhetnek, így velük együtt elveszthetnek a hozzájuk kapcsolódó emlékek és értékek is. Ma már jellemzően ezen fotók nagy száma főként az idős generáció birtokában van, így azok értékmentése kiemelt prioritást jelent.

1 Az ábra forrása [https://hu.wikipedia.org/wiki/K%C3%A1lti\\_M%C3%A1rk](https://hu.wikipedia.org/wiki/K%C3%A1lti_M%C3%A1rk) (2022. 06.17.)

2 Az ábra forrása: <https://24.hu/tech/2018/07/28/csaladi-fotoalbum-a-digitalis-korban-mutatjuk-az-ertelmes-megoldasokat/#> (2022. 06.17.)



3. ábra: Családi fotó<sup>3</sup>



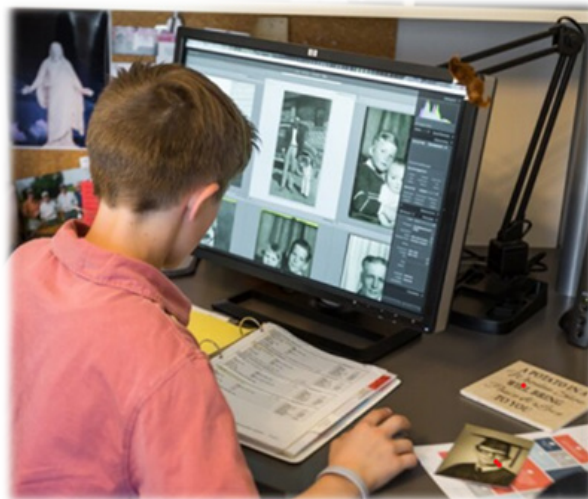
4. ábra: Családi fotó<sup>4</sup>

### 3. Digitális technológiák a családi emlékek megőrzéséért – a Kézai Simon Program



**kézai simon  
program**

Mit kezdünk ezekkel a régi családi fotókkal? Talán nem klónozzhatjuk a fizikai tárgyakat, például a családi örökséget és emlékeket, de átalakíthatjuk őket digitális formába, ami legalább megkönnyíti a másokkal való megosztásukat, és megőrzi a digitális emlékeket fizikai elvesztés esetén is.



5. ábra: Családi képek digitális feldolgozása<sup>5</sup>

3 Az ábra forrása: <https://kezai.hu> (2022. 06.17.)

4 Az ábra forrása: <https://kezai.hu> (2022. 06.17.)

5 Az ábra forrása: <https://www.familysearch.org/en/blog/six-steps-to-digitizing-your-family-photos> (2022.06.15.)





A Digitális Jólét Nkft. ezen nemzeti kulturális örökségünk megőrzése érdekében egy értékmentő programot indított útjára és működtet, amit nevezhetünk többféleképpen is:

- Családi krónika digitalizált albumba rendezve – családok digitális krónikája
- Online digitális családi fotóarchívum
- Magyarország digitális családi albuma
- Közösségépítő országos mozgalom
- Digitálisan életre kelt családtörténet
- Digitális Családtörténeti Album
- Nyilvános és kutatható családi fotótár

A Program elsődleges céljaul tűzte ki, hogy a családok és az intézmények birtokában lévő felbecsülhetetlen értékű analóg technikával megörökített fényképeket, diákat, negatívokat és a hozzájuk kapcsolódó információkat megőrizze a jövő nemzedékei számára. Ez nemcsak a szűkebb családok számára fontos, hanem közös, nemzeti kulturális érdekünk, hiszen segítségükkel átfogóbb képet kaphatunk hazánk 20. századi történelméről, életéről, eseményeiről.

A Program minden állampolgár számára lehetőséget nyújt a papíralapú, illetve analóg formában megtalálható családi fényképei és egyéb dokumentumai digitalizálására, valamint azok tárolására 100 kép erejéig.

Az egyszerű regisztráció után a fotók akár otthonról történő önálló digitalizálása a Kézai Simon Program elnevezésű – Android és IOS operációs rendszerre egyaránt letölthető – ingyenes applikáció segítségével végezhető el, illetve a szkennelt képek számítógépről feltölthetők a [kezai.hu](http://kezai.hu) weboldalon is. Ehhez kapcsolódóan a Program egyik nagy erőssége, hogy a generációk közötti kapcsolat erősítését szolgálja, hiszen a digitális világban már otthonosan mozgó fiatal korosztályt, például az unokákat a nagyszülőkkel egy újabb közös platformra helyezi. Egyszerre van mód az idősek digitális kompetenciájának fejlesztésére és a családi történetek átadására a következő generációk számára.



6. ábra: részlet a Kezai.hu oldaláról





**KÉZAI SIMON  
ARCHÍVUM**



7. ábra: Kézai.hu applikációs megjelenítése<sup>6</sup>

Azonban, számos olyan, - főként idős - felhasználó is szeretné megőrizni emlékeit, akik a regisztrációt nem tudják önállóan elvégezni, mert nincs hozzá eszközük, vagy nem rendelkeznek kellő szintű digitális kompetenciával. Számukra a Digitális Jólét Program Hálózat Pontjain dolgozó Mentorok segítenek a Programhoz történő csatlakozásban, a fotók digitalizálásában és feltöltésében. Továbbá fejlesztés alatt van Mobil Kézai Pontok kialakítása, amelyek újabb lehetőséget és szolgáltatást kínálnak a történelmi értékű fotódokumentumok feltárására és megőrzésére.



8. ábra<sup>7</sup>: Kézai Simon Pont átadása Győrben, az egyetemi könyvtárban



9. ábra: Kézai Simon Program digitális plakátja<sup>8</sup>

6 Az ábra forrása: <https://play.google.com/store/apps/details?id=com.codecluster.ksp&hl=hu&gl=US> (2022.06.15.)

7 Az ábra forrása: <https://www.youtube.com/watch?v=XJzAwF2KV28> (2022. 06.17.)

8 Az ábra forrása: <https://en-gb.facebook.com/pages/NMI-Erd%C5%91smecske/206995419847013> (2022. 06.17.)



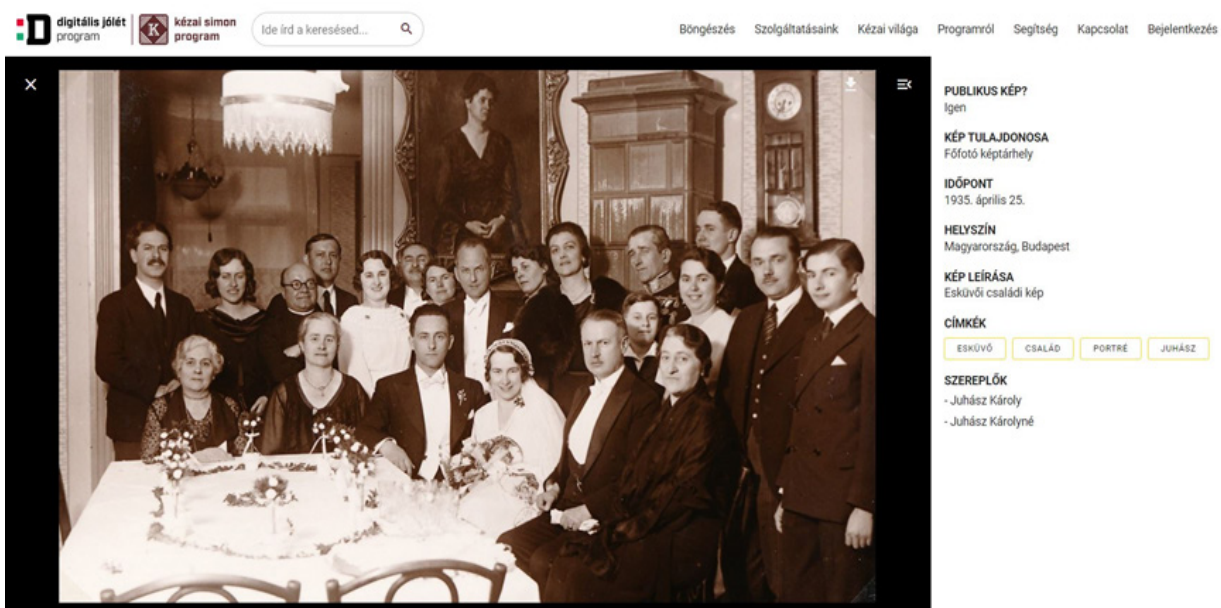
10. ábra: Kézai Simon Program digitális plakátja<sup>9</sup>

A családi fotókból a weboldalra történő feltöltését követően kialakíthatók saját galériák. A képek tetszőleges témák szerint albumokba rendezhetők. A családtagok a fényképekhez megjegyzéseket fűzhetnek, megjelölhetik a fényképeken szereplő személyeket, elmesélhetik a fényképhez kapcsolódó történeteket, ezáltal digitálisan életre keltve a fotókat és mások számára is megismerhetővé téve azokat. Továbbá különböző címkékkel láthatjuk el a fotókat, ami abban segít, hogy publikus képek esetén azok könnyebben előkereshetőek legyenek.

Minden kép esetében a tulajdonos dönthet azok felhasználásáról. A felület lehetőséget biztosít arra, hogy a fotó birtokosa milyen körben osztja meg a családi emlékeit. Ezek alapján az alábbi típusokat különböztetjük meg:

- **Privát:** Alapbeállítás: Ha a fotó tulajdonosa a feltöltés során a privát módot jelöli be, akkor kizárólag ő láthatja fényképeit.
- **Családtagok:** Opcionális: Ha a fotó tulajdonosa a feltöltött fényképeket és albumokat szeretné megosztani az általa megjelölt regisztrált családtagokkal, ismerősökkel. Ebben az esetben csak a kiválasztott személyek férhetnek hozzá a képekhez.
- **Ismerősök:** Opcionális: A feltöltött fotók megoszthatók más regisztrált felhasználókkal.
- **Publikus:** Opcionális: A legfelső szint, amikor a fotók tulajdonosai a publikus beállítást választják. Ebben az esetben a képek nyilvánossá válnak, vagyis bárki számára megtekinthetőek.

<sup>9</sup> Az ábra forrása: <https://www.facebook.com/kezaisp/> (2022. 06.15.)



11. ábra: Publikus fotó a Kezai.hu-ról

Az online felületre feltöltött családi képeket és adatokat egy magyar tulajdonban lévő szoftver rögzíti, és egy biztonságos, magyar központi adatbázisban kerülnek tárolásra. Ez állami garanciavállalást jelent a fotók hosszú távú megőrzésére úgy, hogy azok bármikor elérhetőek legyenek a képek tulajdonosai számára, a szigorú adatvédelmi előírásokat betartva. A publikussá tett képek, a közzététel előtt minősítési eljáráson esnek át, hiszen a feltöltött képeken nem szerepelhet sértő tartalom vagy tiltott önkényuralmi jelkép.

A Kézai Simon Program igyekszik minél szélesebb körben információt adni tevékenységéről és online közösségét is bővíti: a Kézai Simon Program Facebook oldalon, a Kézai Simon Program Instagram oldalon, a Kézai Simon Program YouTube csatornán, és a Digitális Témahét YouTube csatornáján.

#### 4. Elért eredmények, tervek és kihívások

*Mit nyújt tehát a Kézai Simon Program?*

- 1990 előtti papír/analóg formában megtalálható családi fényképek és egyéb dokumentumok digitalizálását ingyenesen 100 kép erejéig;
- Digitális Családtörténeti Albumok kialakítását és gondozását;
- Digitalizálást a „Kézai Simon Program” elnevezésű ingyenes applikáció segítségével a kezai.hu online felületre;
- Tájékoztatás és segítségnyújtás a regisztrációban Digitális Jólét Program Hálózat Pontjain;
- A digitáliskompetencia-fejlesztő tevékenységek elérése a Program segítségével
- Új szolgáltatásokat alakít ki a családtörténeti kutatások támogatására.



### Eddig elért eredmények

- A kezei.hu oldalon több mint 42 ezer db fénykép található.
- Az összes kép 70 %-a, publikus kép.
- Az összes kép több, mint 750 albumban található
- Felhasználók száma megközelíti 1.200 főt.
- A Program közösségi felületeit több mint 1.000 fő felhasználó követi.

Emellett a Program folyamatosan építi partnerhálózatát, ezek közül néhány: múzeumok, könyvtárak, Magyar Nemzeti Levéltár, Népfőiskola-Lakitelek, Belváros-Lipótváros Budapest V. Kerület Önkormányzat, Edutus Egyetem, Országos Széchényi Könyvtár, AGORA Szombathelyi Kulturális Központ, Széchényi István Egyetem.

### A jelen digitális emlékei...a kihívások

Digitálisan integrált életünk során képeink és otthoni videóink nagy részét pompás digitális fényképezőgépekkel, hihetetlen okostelefonokkal vagy más könnyen hordozható eszközökkel készítjük és osztjuk meg, de mit kezdünk ezeknek az eszközöknek a gyümölcseivel – a digitális fájlok és források sokaságával, amelyet naponta, hetente, havonta, évente felhalmozunk? Mit kezdünk a digitális életünket alkotó személyes digitális tartalmainkkal? Hogyan tudunk ezután közösen digitális családtörténetet építeni?

## 5. Összefoglaló

Kézai Simon Program célja és törekvése tehát egy hiánypótló kezdeményezés a magyar lakosság körében. Felbecsülhetetlen a közösségteremtő ereje, valamint a generációk összekapcsolása. A Program egyedülállósága, annak a lehetőségnek a biztosítása, hogy a határon belüli és kívüli magyar lakosság az applikáció segítségével saját maga dolgozhassa fel, őrizhesse meg emlékeit az utókor számára.

### Bibliográfia

Collectionaire for families, 2022: <https://collectionaire.com/> (2022.06.06.)

Creating Family Archives, 2019: <https://www2.archivists.org/publications/creating-family-archives> (2022.06.06.)

FamilyArchive Business, 2018: <http://familyarchivebusiness.com/> (2022.06.07.)

FamilySearch, 2022: <https://www.familysearch.org/en/> (2022.06.09.)

Six Steps to Digitizing Your Family Photos, 2017: <https://www.familysearch.org/en/blog/six-steps-to-digitizing-your-family-photos> (2022.06.06.)

## Open Monograph Press e-könyvplatform a Szegedi Tudományegyetemen

Fülöp Tiffany\*

[tiffany.fulop@ek.szte.hu](mailto:tiffany.fulop@ek.szte.hu)ORCID: [0000-0002-2219-8455](https://orcid.org/0000-0002-2219-8455)

Molnár Tamás\*

[tamas.molnar@ek.szte.hu](mailto:tamas.molnar@ek.szte.hu)ORCID: [0000-0003-2571-3595](https://orcid.org/0000-0003-2571-3595)

Hoczopán Szabolcs\*

[szabolcs.hoczopan@ek.szte.hu](mailto:szabolcs.hoczopan@ek.szte.hu)ORCID: [0000-0002-7892-9974](https://orcid.org/0000-0002-7892-9974)

\* SZTE Klebelsberg Könyvtár

**Absztrakt**

The Open Journal System (OJS), developed for journals, has been a huge success at the University of Szeged. Most of the editorial offices have moved to the platform hosted by the Klebelsberg Library.

Encouraged by the success of the OJS and assessing how many e-books are produced annually at the University of Szeged, we started to develop our own professional e-book platform in 2021, using the Open Monograph Press system of the Public Knowledge Project. Our aim is to move the e-books of the University of Szeged away from the ever-changing departmental websites and to give them a worthy place on a platform that allows for long-term storage, retrieval and citation.

**Keywords:** Open Monography Press, e-book platform, open access publishing

**Bevezető**

A Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára mindig is nagy hangsúlyt fektetett az egyetem láthatóságának növelésére, folyamatosan javítva és bővítve szolgáltatásai körét. Ennek egyik legújabb elemét a Tartalomszolgáltatási Osztályon elinduló e-könyv platform<sup>1</sup> jelenti, aminek célja az egyetemi egységekhez kötődő kurrens könyvsorozatok elérhetőbbé és átláthatóbbá tétele. Ehhez a szolgáltatáshoz az Open Monograph Press<sup>2</sup> (továbbiakban OMP) szoftvert használjuk, és ugyan még a tapasztalatszerzés kezdetén járunk, ez a tanulmány szeretné bemutatni az elképzeléstől a megvalósítás kezdetéig tartó szakasz tapasztalatait.

Az OMP annak a több nemzetközi egyetemet érintő Public Knowledge Project nevű kezdeményezésnek az ingyenes szoftvere, amihez az Open Journal System is tartozik, és aminek célja olyan nyílt hozzáférésű szoftverek fejlesztése, amik a tudományos publikációk minőségét és elérhetőségét támogatni, segíteni tudják, főleg egyetemi kiadványok keretében. A szervezet így az OJS mellett többek között preprint kiadás menedzselését és megjelentetését kezelő (Open Preprint System), metaadat indexelő (Open Harvester System) és – az

---

1 URL: [www.ebook.ek.szte.hu](http://www.ebook.ek.szte.hu)

2 URL: <https://pkp.sfu.ca/omp/>





általunk jelenleg célkeresztbe kerülő - tudományos könyvek létrehozását, menedzselését és megjelentetését segítő szoftvert is fejlesztett a tudományos információáramlás segítése céljából.

### Előzmények, avagy hogyan merült fel az igény az OMP használatára?

A Klebelsberg Könyvtár évek óta „kihelyezett” ISSN/ISBN irodaként is működik<sup>3</sup> a Szegedi Tudományegyetemen. Ennek következtében teljes rálátásunk van az egyetemünkön megjelenő jelentős e-könyv mennyiségre. Ezek az e-könyvek korábban elsősorban tanszéki honlapokon voltak megtalálhatóak (már ameddig megtalálhatóak voltak) nélkülözve egy professzionális tartalomszolgáltató rendszer előnyeit és biztonsági szolgáltatásait.

Könyvtárunk számtalan, különböző tematikus repozitóriumot szolgáltat az EPrints 3 szoftver segítségével. Ezekből az egyik az SZTE Egyetemi Kiadványok repozitóriuma<sup>4</sup>, ami a Szegedi Tudományegyetemhez és jogelődjéhez köthető lezárt és kurrens, tudományos és szépirodalmi kiadványok lelőhelye. Tartalmának kategorizálását tekintve láthatóvá válik, hogy a folyóiratok mellett mennyi könyvsorozat és önálló könyv teszik ki az állomány egy részét.

Természetesen ezek egy része már megszűnt sorozatok, nem működő szerkesztőségek kiadványai, azonban készítettünk egy felmérést, hogy mennyi ebből a kurrensnek tekinthető elem. Kurrensnek ítéltük azokat a címeket, amelyeknek az elmúlt kb. három évben jelent meg kiadványa és hellyel-közzel folyamatos volt a megjelenésük. Mivel a Szegedi Tudományegyetemhez kötődő kiadványokról van szó, ezért az egyetem szervezeti egységei mentén (tanszékek, intézetek, kutatócsoportok) mértük fel, hogy mely egységeknek lehetne szolgáltatni az OMP-t. A felmérés azt mutatta, hogy közel 26 tanszéknek 35 olyan sorozata van az egyetemen, ami kurrensnek tekinthető, ebből 13 esetben határozottan célszerű lenne a kiadványok megjelentetése - a repozitórium mellett - a tanszéki weboldalakon megjelenő listászerű felsoroláson túl is.

Mindezek mellett, ha megnézzük a tanszéki kiadványokat, azt láthatjuk, hogy a társadalom és humán tudományok a reál szakterületekhez képest a folyóiratok mellett vagy helyett, de nagy számban jelentetik meg tudományos eredményeiket könyvekben (monográfiákban és tanulmánykötetekben), amiket jelentős részben könyvsorozatokba tömörítenek.

Az általunk szolgáltatott Open Journal System nagy igényét az a 27 egyetemi-tanszéki folyóirat mutatja meg, ami jelenleg rajtunk keresztül megjelenik. Az OJS-ben lassan, de évről-évre biztosan gyarapodó és azt különféle mértékben kihasználó szerkesztőségek száma azt mutatja, hogy igényt lehet kelteni szolgáltatással. Akár a rendszer tanszékek között terjedő hírneve, akár az általa kínált szolgáltatások kecsegtető lehetősége miatt, egyre többen döntenek a csatlakozás mellett.

Előfordult már néhányszor, hogy könyvsorozat behozatalának kérésével keresték fel OJS szolgáltatásunkat. Ez pedig felvetett olyan problémákat és kérdéseket, amik a két rendszer különbözőségén alapulnak, és rávilágítanak az OMP rendszer szükségességére. Az egyik legszembetűnőbb probléma természetesen az, hogy az OJS által létrehozott weboldalakon megjelenő feliratok a folyóiratok terminusait tükrözi, míg egy könyv/könyvsorozat esetében sorozatokról, könyvekről, kötetekről, kötet szerkesztőkről, sorozatonkénti ISSN-ről és könyvenkénti ISBN-ekről kell beszélni, amiknek megjelenítésére az OJS rendszer értelemszerűen nem a legalkalmasabb.

3 URL: <http://www.ek.szte.hu/isbnissn-szam-igenyles/>

4 URL: <http://acta.bibl.u-szeged.hu/>



## Az OMP felépítése

A keretrendszer alapegysége a kiadó, amiből egy telepítésen belül többet is létre lehet hozni. Ennek a kiadónak van egy általános leírása, hozzá tartozó szerkesztősége és munkafolyamata, valamint akármennyi sorozatot létre lehet hozni egy kiadón belül, aminek van saját ISSN száma és leírása, illetve hozzá tartozó művei.

Alapvetően, ha a kereskedelmi kiadókat és a könyvesboltokat nézzük, akkor az OMP a versenyszférában megjelenő struktúra mentén épül fel. Egy kiadó egy szerkesztőséggel, sok sorozattal és rengeteg könyvvel a katalógusában, amiket kategóriák használatával tovább tudunk tagolni - új megjelenésekkel és kiemelésekkel a kezdőoldalon. Ha ezt adaptálni akarjuk egy egyetem kiadványaira, akkor el kell döntenünk, hogy mi feleljen meg a kiadó alapegységének.

The screenshot shows the website of the SZTE BTK Medieval and Early Modern Hungarian History Department. The header includes the department name, navigation links for 'Katalógus' and 'Információk', and a search bar. The main content area is divided into sections: 'Válogatott' (Selected) and 'Új megjelenések' (New arrivals). The 'Válogatott' section features a book titled 'URBS, CIVITAS, UNIVERSITAS: Ünnepi tanulmányok Petrovics István 65. születésnapja tiszteletére' (1. kötet), edited by Papp Sándor, Tóth Sándor László, Zoltán Kordé, and Kordé Zoltán, published in January 2018. The 'Új megjelenések' section shows two new books: 'A szultán hadainak tápláléka: Az oszmán sereg élelemellátási rendszere az 1658. évi hadjárat tükrében' (2021 January 1) and 'New approaches to the Habsburg-Ottoman diplomatic relations' (1. kötet), edited by Papp Sándor and Marton Gellért Ernő, published in January 2021. On the right side, there are filters for 'Nyelv' (Language: Magyar, English), 'Böngészés' (Browse), 'Sorozatok' (Series: Fontes et Libri, Fontes et Libri. Tanulmányok / Studies, Fontes et Libri. Források / Sources, Fontes et Libri. Monográfiák / Monographs), and 'Open Monograph Press'.

1. kép Kiadó kezdőoldala

Azokban az esetekben, ahol az egyetemnek van egy hivatalos kiadója, viszonylag egyszerű a megoldás, hiszen az OJS telepítésen belül csak egy darab kiadót kell létrehozni, ami alá felkerülhet minden tartalom. Viszont gyakori, hogy a kiadott könyvek kiadói különböző egyetemi egységek, ami már felveti a kategorizálás jogosságát. Ebben az esetben a kiadónak több mindent is meg lehet feleltetni, annak fényében, mit találunk a legmegfelelőbbnek. Ilyenkor egy egység neve fogja össze az alá tartozó könyvsorozatokat és önálló műveket, így az egység alá több tényleges kiadó is kerülhet. Az OMP rendszer rugalmassága biztosítja ennek a kivitelezésnek a lehetőségét.

Ez az egység lehet tehát kar, tanszék, intézet, kutatócsoport vagy akár a tényleges kiadó is. Az előnyök és hátrányok megvitatása közben az alábbi pontokat kell átgondolni:

- A kiválasztott egység alatt egyértelműen reprezentálódik az oda tartozó tudományos munkásság egésze. Ha ez egy kar, akkor látványos a mennyiség, de jobban elvész a tanszékek tevékenysége. Ha ez egy kiadó, akkor nagyon elaprózódhat az OMP telepítés a sok, esetleg akár kisebb számú könyvet is megjelentető kiadó miatt. Arról nem beszélve, hogy abban az esetben, ha egy tanszéknek például több kiadója is van (például a tanszék és egy kutatócsoportja külön kiadóként van jelen), nem lesz átlátható a tanszék munkássága a külön telepítések miatt.



- A belső könyvkiadói munkafolyamat használata esetén, ha egy karról van szó, akkor az összes egyetemi egység, kiadó ugyanabba a rendszerben kell, hogy dolgozzon, ami jelentős mennyiségű kéziratot és így szervezést, elkülönítési nehézséget jelent.

Az általunk kínált szolgáltatásnál azt a megoldást választottuk, ahol a tanszékek jelentik az alapegységet.

Mostanra már láthatóvá vált az OMP felépítése: Kiadónként (ami tartalmaz saját leírást, szerkesztőséget, kézirat beküldési információkat, kapcsolati adatokat) több sorozatról, vagy önálló kiadványokról beszélhetünk, amelyek egy összesített katalógusban és sorozatonként vagy kategóriánként is meg tudnak jelenni. A rendszer kezeli a kötetszerkesztőket és sorozatszerkesztőket, a monográfia szerzőit és a fejezet/tanulmány szerzőket egyaránt.

### A szolgáltatás kialakításának folyamata

A már korábban olvasható szükséglet-felmérés és a fejlesztők által elérhető alapos útmutatók elolvasása után informatikus kollégáink készítettek egy teszt telepítést, ahol mindent alaposan át tudtunk tanulmányozni az OMP működését tekintve. Ezek megismerése után meghatároztuk a korábban említett alapegységet, ami mentén végül dolgoztunk.

A könyvkiadással kapcsolatos szerkesztői munkafolyamatról annyit érdemes megemlíteni, hogy a szoftver egy teljes mértékben online végigkísérhető, automatizált és nagyon rugalmas menetet kínál a számunkra. Az öt szakasz során (kézirat beküldése, belső majd külső szakmai lektorálás, technikai szerkesztés és előállítás) a kettős-vak lektorálástól az automatikus e-maileken át az egyértelmű nyomon követhetőségig kapnak segítséget a szerkesztők, sőt a technikai szerkesztéshez kapcsolódó olvasószerkesztői, illetve korrektori szerepkörök még alaposabban kidolgozottak, hiszen itt a nyomdai rendszert képezi le.

Bár a szoftver felépítése, ahogy a PKP maga is az open access (OA) típusú elérhetőség irányába történő támogatást célozza meg, mégis a rendszer nem csak nyilvánosan elérhető könyveket tud szolgáltatni, hanem a könyvkiadáshoz tartozó szinte összes elem beállítható, kitölthető a rendszerben a marketing információktól kezdve a könyv metaadatolásán át az esetleges fizetési lehetőségekig. A szerepkörök, a rendszer automatikus értesítései, az OMP egésze mind az OA-t, mind pedig a fizetős modellt teljeskörűen támogatja. Éppen ezért következő lépésként meghatároztuk azokat a metaadatokat és egyéb kiadással kapcsolatos adatokat, amelyeknek a megadását szükségesnek gondoltuk. Természetesen ezek a későbbiek során igény vagy kérdés esetén bármikor bővílni tudnak.

2. kép Metaadatok

A továbbiakban még két nagy lépés állt előttünk a szolgáltatás üzembe helyezése felé vezető úton: az OMP nem rendelkezett magyar fordítással, és szükségünk volt egy archívumépítési technikára.

## A fordítás elkészítése

A PKP lehetőséget biztosít arra, hogy (nyílt forráskódú szoftver lévén) a közösség, ha szüksége van rá maga fordítsa le a szoftvereiket. Mindezt egy viszonylag kényelmesen működő online felületen, a Weblate-en<sup>5</sup> tehetjük meg, amely többek között olyan opciókat biztosít számunkra, mint a mechanikus gépi fordítás, a már korábban fordításra került hasonló szövegek nyomon követése vagy saját szótár létrehozása és bővítése. Egyszerre többen is tudnak dolgozni a rendszerben, ahol megjegyzésekkel és javításokkal is el tudják látni a szövegeket.

3. kép Weblate felület - OMP szövegelem

5 URL: <https://weblate.org/hu/>



Az OMP fordítása során több mint ezer szövegrészletet kellett lefordítanunk az egyszavas címkéktől a teljes e-mailekig, amibe a szoftverhez tartozó bővítmények még nem lettek bevonva. A rendszer rugalmas működését jelzi az is, hogy abban az esetben, ha valami nem került még lefordításra, vagy egy későbbi fejlesztés újabb szövegrészleteket ad a szoftverhez, akkor nem jelentkezik hiba, egyszerűen átmenetileg az angol eredeti szövegrészletek jelennek meg a felületeken.

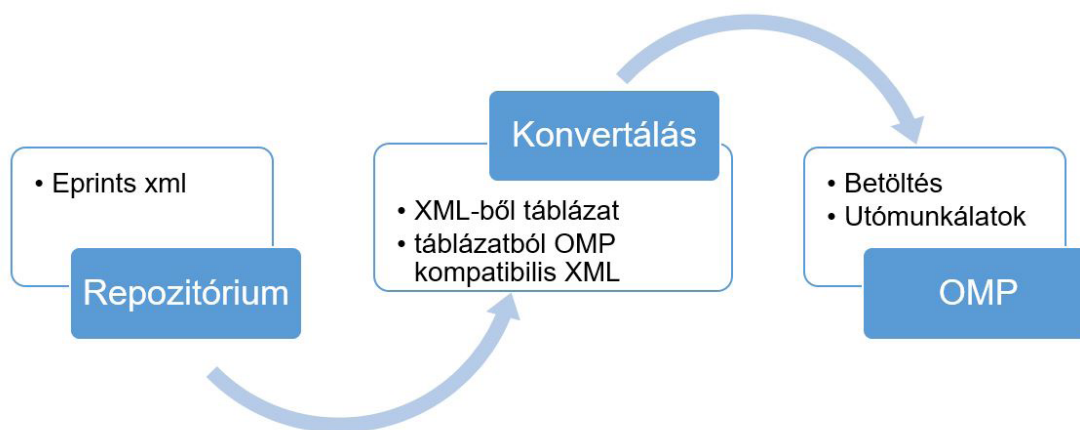
A nyersfordítást több kolléga segítségével készítettük el, amin ezek után a szoftveren dolgozó kollégákkal a rendszer ismeretében egy utólagos tisztázást végeztünk. Erre azért volt szükség, mert nehézséget okozott a fordítás elkészítése során az, hogy a szövegrészleteket kontextustól, funkciótól, a rendszertől függetlenül kellett lefordítani kollégáimnak. Így az egy-két szavas szövegekről egy külsős szem számára nem volt mindig egyértelműen eldönthető, hogy azt például cselekvő vagy passzív módba kellene fordítani. Másrészről pedig szükség volt a nyomdai és kiadói szakszavak egyértelmű meghatározására és azok következetes használatára. Mindezek végén egy utolsó nagy átfésüléssel lezártuk a többhetes munkát, ami során annak ellenőrzésére fektettük a hangsúlyt, hogy a rendszerben előforduló kifejezések, utasítások és szakszavak minden esetben azonosak legyenek, hiszen a szoftver használhatóságát nagyban befolyásolja az, hogy ha például formanyomtatványként hivatkozunk egy elemre, akkor annak menüpontja és az össze hozzá tartozó utasítás is így legyen címkézve.

Amikor elkészül egy fordítás az online felületen, azt jelezni kell a szoftver fejlesztőinek, akik majd egy későbbi frissítésnél már a fordítást is beemelik a szoftver kódjába, így az a többi nyelv között automatikusan elérhetővé válik. De amíg erre nem kerül sor, addig lehetőség van arra, hogy manuálisan letöltsük ezt a nyelvi fájlt csomagot és kézzel betegyük a megfelelő helyre, hogy elérhetővé váljon a nyelv az OMP rendszerében. Az általunk szolgáltatott OMP az utóbbi módon kapta meg a magyar nyelvet.

### Az archívumépítés problémái


Habár az általunk kiválasztott könyvsorozatok kurrensek, de azok korábbi könyveit is szükséges betölteni ahhoz, hogy megfelelően jelenítsen meg egy tanszéket/intézetet/egységet vagy egy-egy kutatási területet. Sajnálatos módon jelenleg még nincs lehetőség annak a viszonylag egyszerű mechanizmusnak a használatára, ahol az egy könyvhöz tartozó összes adatot ki lehet tölteni egy űrlapon a beküldéshez. Így vagy beküldjük a kéziratokat a szokványos úton, kihagyjuk a szükségtelen munkafolyamat állomásokat, mint a szakmai lektorálások és azután tölthetjük ki az adatokat, jelentethetjük meg a könyvet. Vagy pedig használhatjuk az OMP saját metaadat struktúráján alapuló xml-re épülő betöltést.

A mi célunk az volt, hogy ezt a munkafolyamatot, amennyire tudjuk gyorsítsuk a PKP egy másik szoftverének (OJS) szolgáltatása során megszerzett tapasztalok alapján. Ez pedig azt jelentette, hogy mivel ezek a tanszéki, egyetemi kiadványok már betöltésre és feldolgozásra kerültek az SZTE Egyetemi Kiadványok repozitóriumában, ezért a metaadatokat onnan szerettük volna átvenni. Ez azt jelentette, hogy a repozitóriumból kiexportált xml fájlt Excel táblázattá alakítottuk, amit megfeleltettünk egy olyan Excel struktúrának, amiből már tudtunk OMP kompatibilis xml-t gyártani.



4. kép A konvertálás folyamata

Az xml-el dolgozva láttuk át azt, hogy mennyire másképp működik az OMP alapfelépítése az OJS-hez képest. Mivel a rendszer támogatja azt is, hogy az egyes fejezeteket vagy tanulmányokat is kezelni tudjuk, tehát lehessen nekik saját pdf fájljuk, címük, szerzőik és leírásuk, ezért az xml sablonjuk felépítése ezeknek a kiadványoknak jóval komplexebb. Amíg az OJS-ben a tanulmányok a folyóiratszámok metaadataitól függetlenül léteztek addig az OMP-ben a kiadványhoz tartozó adatokat a teljes mű összes adata és a fejezetek összes adata együttesen jelenti. Ezzel együtt az xml betöltés megvalósítása nagyrészt sikerült, a legfontosabb elemek megfelelően működnek benne, egyedül a kiadványokhoz tartozó adatok változó számának kezelhetőségén kell még dolgozni.

<h3>Új nemzedék: A szegedi Régészeti Tanszék tehetséggondozásának elmúlt évtizedei</h3>		Nyelv <a href="#">English</a> <a href="#">Magyar</a>
Gyöngyvér Bíró (ed) Katalin Pintér-Nagy (ed) Tamás Szebenyi (ed)		Böngészés  Sorozatok <a href="#">Monográfiák a Szegedi Tudományegyetem Régészeti Tanszékéről</a>
<b>Kulcsszavak:</b> Régészet , Magyarország , tanulmányok, Tehetséggondozás	<input type="button" value="PDF"/>	
<b>Tartalom</b> Ünnepi kötet B. Tóth Ágnes, Kulcsár Valéria, Vörös Gabriella és Wolf Mária tiszteletére	Közzétett január 1, 2020	
<b>A lajismizei La Tène üveggarperec-töredékekről</b> Zsófia Sz. Osváth, István Fórizs, Máté Szabó, Bernadett Bajnóczy <input type="button" value="013-027. o."/>	Sorozatok <a href="#">Monográfiák a Szegedi Tudományegyetem Régészeti Tanszékéről</a>  Nyomtatott ISSN 2062-9877	
<b>Elfelejtett ezüstök görög érmék a pécsi Janus Pannonius Múzeum gyűjteményéből</b> Mira Majdán <input type="button" value="029-060. o."/>	Copyright (c) 2020 a szerzők; a szerkesztők	

5. kép Könyv és fejezeteinek adatai



Ezek a korlátozottabb betöltési lehetőségek azonban megfontolandóvá teszik azt a kérdést is, hogy a katalógusba bekerülő művek esetén csak a teljes mű adatait, vagy a benne található részek (fejezetek, tanulmányok) adatait is fel tervezi-e tüntetni a szolgáltató vagy a kiadó.

### További tervek

Jelenleg három egységet költöztettünk be a felületre, akik úgy érkeztek hozzánk, hogy igazán mi magunk még nem kezdtük el direkt megkeresni a tanszékeket. Ez már pozitívnak tekinthető. Első körben szeretnénk főleg azokat a kurrens megjelenésű sorozatok kiadóit megkeresni ezzel a lehetőséggel, akikre már tanulmányunk elején is utaltunk, hogy a lehető legjobb megjelenést és láthatóságot tudjuk biztosítani a tudományos műveik nyílt hozzáférésű megjelentetése céljából. Ráadásul az OMP standard metaadatoknak köszönhetően lehetőséget biztosít a nemzetközi adatbázisok, aggregátorok által történő aratásra is.

Célunk még a belső szerkesztőségi munkafolyamatok használatának bátorítása, főleg azokban az esetekben, ahol elsődlegesen tanulmánykötetekben gondolkodnak a kiadók. Lehet, hogy a könyvkiadói munkafolyamat nem tekinthető annyira komplexnek, mint egy folyóirat szerkesztése, ettől függetlenül a rendszer által történő automatizálás (emailek, sablonok, útmutatók, szerepkörök) sokat könnyíthet a könyvkiadás munkafolyamatán.



## Mesterséges intelligencia, digitális bölcsészet, kulturális örökség: trendek és eredmények

Palkó Gábor

*ELTE Digitális Bölcsészet Tanszék, Digitális Örökség Nemzeti Laboratórium*

[palko.gabor@btk.elte.hu](mailto:palko.gabor@btk.elte.hu)

ORCID: [0000-0002-4394-8577](https://orcid.org/0000-0002-4394-8577)

A XXI. század első évtizedeiben a kulturális, tudományos és oktatási területen két párhuzamosan zajló, egymáshoz szorosan kapcsolódó trendnek lehetünk tanúi. Egyik oldalról a mesterséges intelligencia (MI) beláthatatlan ütemben ír át és vált le különféle bevett kutatási, vizsgálati módszertanokat, de túlzás nélkül állíthatjuk, az MI a mindennapi életünkre is egyre nagyobb hatást gyakorol. A másik oldalról pedig – részben a kulturális örökség digitalizálása révén, részben pedig a digitálisan létrejövő (born digital) anyagok termelődésének óriási volumene következtében – szó szerint – beláthatatlan nagyságrendű adatterek és -hálózatok jönnek létre.

Ha a digitálisan hozzáférhető dokumentumok mennyisége többszörösen meghaladja a források feldolgozása során következtetéseket megfogalmazni képes szakember befogadóképességét, és ha a digitális dokumentumok özöne számítógép segítségével már nem tekinthető át, akkor a források közreadása, kutathatóvá tétele, illetve a jövő generáció digitális kompetenciáinak formálása terén a kutatásnak, az oktatásnak és az archiválásnak eleve számolnia kell a gépi intelligencia teremtette új helyzettel.

A mesterséges intelligencia alapú nyelvelemző eszközök fejlesztéséhez szükség van nagy tömegű – történeti forrásokat, sajtóanyagot, médiatermékeket, szépirodalmat és web 2.0-es forrásokat egyaránt tartalmazó – magyar nyelvű anyagra. A heterogén szövegtörzs egyrészt a többcélú nyelvmodellek készítéséhez szolgál alapul, másrészt olyan bölcsészeti, társadalomtudományi vagy éppen piaci horizontú kutatásokat, fejlesztéseket tesz lehetővé, amelyek az oktatásban, a tudományos diskurzusban, valamint a nagyközönség számára készülő szolgáltatásokban is hasznosíthatók.

Ahhoz, hogy nagy méretű és kitűnő minőségű magyar korpuszok, az ezekből előállított tanítóanyag, továbbá magas színvonalú programozási architektúra álljon rendelkezésre, digitális bölcsészek, számítógépes nyelvészek, informatikai szakértők összehangolt munkájára van szükség. De még ez sem elég: elengedhetetlen a kutatóhelyek, az egyetemek, a kulturális örökséget megőrző intézmények (a „memória-intézmények”), a piaci és kormányzati szereplők szoros együttműködése.

Ebben a rövid előadásban arra vállalkozom, hogy ennek a nemzetstratégiai szempontból elodázhatatlan együttműködésnek a kulturális örökség és a bölcsészettudomány szempontjából legfontosabb gyakorlatait és követendő példáit bemutassam. Előadásom első felében a mesterséges intelligencia, illetve a digitális kulturális örökség kultúrtechnikái közül emelem ki azokat, amelyek – meglátásom szerint – a legnagyobb jelentőséggel bírnak a jövőre nézve. Az előadás második felében pedig az együttműködés intézményi formáira térek ki.

### Mélytanulásra épülő magyar nyelvmodellek fejlesztése

Az okoseszközök elterjedése révén a tény, hogy a hangzó és írott nyelvi megnyilvánulásainkat a mesterséges intelligencia alapú alkalmazások nemcsak értelmezik, de reagálnak is rájuk, mára általános tapasztalattá vált. Azt azonban, hogy ez a változás milyen hatással lesz a kultúra létrehozására és közvetítésére, senki sem tudja.



Az a tény, hogy a magyar nyelvű MI fejlesztésre szorul, és ez hatalmas költségekkel jár, sokak számára csak az ITM és az OTP Bank közös szuperszámítógép-projektjének bemutatása kapcsán vált világossá.<sup>1</sup> Az elmúlt években a mélytanulós technológiára épülő nyelvmodellek fejlesztése terén több kutatóhely is jelentős eredményeket ért el, és ezek a teljesítmények lassan beépülnek a kutatási és kulturális célokra fejlesztett szolgáltatásokba. Csak két példa: a Nyelvtudományi Intézet és a Pécsi Tudományegyetem (PTE) közös terméke, a HILBERT, melynek létrehozását a Microsoft Hungary segítette, jelentős korpuszépítési és számítástechnikai kapacitást mozgósított; a DH-LAB munkatársa, Nemeskey Dávid által készített HuBERT a magyar nyelvre jelenleg a legpontosabb eredményeket produkáló modell. A hírek szerint viszont Feldmann Ádám (PTE) május elején egy új nyelvmodellt mutat be a Budapest ML Fórumon, a gépi tanulással és mesterséges intelligenciával foglalkozó konferencián.

Az ITM-OTP projekthez visszatérve: ez a fent említett együttműködés eminens példa lehet arra, amikor a magyarországi kutatóhelyek és piaci szereplők egy jól körülhatárolt, közös cél érdekében egyesítik szellemi és fizikai erőforrásaikat, belátva, hogy jó minőségű eredmény csakis így érhető el.

Miért fejlesztünk gigantikus nyelvmodelleket? A bankoktól a bölcsész kutatócsoportokig miért van szükségük nyelvmodellekre azoknak, akik emberi megnyilvánulásokat elemeznek és az emberek számára érthető szövegeket generálnak? A válasz elég egyszerű: a mesterséges intelligenciára épülő új technológia információkinyerési és szövegalkotási képességei messze meghaladják a korábbi technológiák lehetőségeit. Az MI néhány év alatt nemcsak a nyelvtechnológiát forradalmasította, de olyan versenyhelyzetet teremtett, amelyben Magyarország nem engedheti meg magának a lemaradást, mert az kulturális és piaci értelemben egyaránt hatalmas veszteséget jelentene. Nem kell magyaráznunk: a magyar nyelvre optimalizált modellek kiemelkedő minősége csak magyar szakemberekkel garantálható.

A létrejövő modellek felhasználása széles körű, egyaránt szolgálja a nemzeti örökség feltárását, a határon túli örökség integrálását, valamint a közvetlen, innovatív piaci megoldások fejlesztését, a beszéd- és kézírás-felismerés, szövegkivonatolás, szemantikus keresés, ügyfélkapcsolat-automatizálás stb. területén.

Egy valamiről azonban kevesebb szó esik a nyelvmodellek fejlesztésében, ahol többnyire a hatalmas számítástechnikai kapacitás kerül csak reflektorfénybe. Az, hogy egy olyan, csekélyebb méretű digitális adatforrásokkal rendelkező nyelv esetén, mint a magyar, a modell tanításához szükséges nyelvi nyersanyag, a korpusz előállítása rendkívüli erőforrásokat igénylő munka, amely nélkül azonban az egész projektum végkimenetele kétes. (Ezért vagyunk különösen büszkék arra, hogy az ITM-OTP projektben a Digitális Örökség Nemzeti Laboratóriumot bízták meg annak a tanítókorpusznak a felépítésével, amely minden eddiginél nagyobb mennyiségű magyar nyelvű szöveg összegyűjtését feltételezi.)

Milyen forrásokból állítható elő ez a bizonyos tanítókorpusz? A szövegek döntő többsége webes eredetű, born digital, vagyis digitálisan született: nyilvános weboldalakról származik. A repozitált gyűjteményi tartalmak, digitálisan született vagy digitalizált könyvek és dokumentumok is részei azonban ennek a minden képzeletet felülmúló méretű és komplexitású korpusznak, ami természetesen nem csak arra való, hogy hatalmas nyelvmodelleket tanítsanak rajta. A korpusznak magának is a kutatás tárgyának kell lennie, hiszen olyan mintázatok felismerését teszi lehetővé, amelyek korábban kivehetetlenek lettek

1 Ld. pl.: Átadták az ITM és az OTP Bank együttműködésével elkészült szuperszámítógép első egységét. <https://kormany.hu/hirek/atadtak-az-itm-es-az-otp-bank-egyuttmukodesevel-elkeszult-szuperszamitogep-első-egyseget>

volna. Mindez persze csakis a számítógéppel támogatott olvasás, a „distant reading” (Franco Moretti) horizontján értelmezhető.

Nincs tér itt a korpusz- és szolgáltatásépítés jogi kérdéseit részletesen tárgyalni. Egyetlen megjegyzés: a hazai digitális bölcsészeti kutatások, és általában valamennyi MI alapú fejlesztés számára kedvező fejlemény, hogy a magyar jogalkotó implementálta a szerzői jogi törvénybe azt az EU által kezdeményezett módosítást, amely a szabad felhasználás körébe vonta a szöveg- és adatbányászatot.<sup>2</sup>

De visszatérve a gigantikus magyar nyelvű korpuszok kutatásához: ezek értelmes feldolgozása, kutathatóvá tétele elképzelhetetlen a szemantikus technológiák fejlesztése nélkül. Ezek felhasználják a fent említett, már részben rendelkezésre álló mélytanulósos nyelvmodelleket, de messze nem azonosak velük.

### Szemantikus technológiák a távoli olvasás szolgálatában

A digitális kulturális örökség szövegtére a folyamatos digitalizáció és a digitálisan született anyagok mennyiségének gyorsuló növekedésével elérte azt a kritikus tömeget, ami a hagyományos lassú vagy közeli olvasáson alapuló adatgyűjtés számára áttekinthetetlen. Igen korlátozott megoldást jelent a szabad szöveges keresés a dokumentumok szövegeiben, különösen az agglutináló nyelvek esetében, hiszen a releváns találatok gyakran elvesznek a kutató számára irreleváns szöveghelyek tengerében. A probléma kézenfekvő megoldása a névelemfelismerő és névelem linkelő algoritmusok használatában rejlik, ezek azonban a magyar nyelvre jelenleg kevésbé hatékonyak, másrészt pedig ezen technológiák komoly informatikai ismereteket feltételeznek, egészen addig, amíg be nem épülnek a szélesebb közösséget megcélzó szolgáltatásokba.

Az egyik legkomplexebb ilyen szemantikus technológia a wikification (wikifikálás). Arepozitált anyagok (pl. OCR-ezett dokumentumok vagy a webaratásból származó cikkek) egy úgynevezett wikifikációs eljárás mennek keresztül, melynek során a szövegobjektumokból kinyert szöveg egyes szavaihoz és szókapcsolataikhoz a Wikidata tudástár elemeit rendeljük hozzá. Léteznek többnyelvű hasonló eszközök, de ezek a magyar nyelvre nem hatékonyak.

Egy pilot keretében a SZTAKI és a Társadalomtudományi Kutatóközpont a már rendelkezésre álló wikification algoritmusok felhasználásával metaadatokat rendelt repozitóriumi tartalmakhoz. (Micsik András, Gárdos Judit).<sup>3</sup>

A Digitális Örökség Nemzeti Laboratórium konzorciumi együttműködésben saját, magyar nyelvre optimalizált wikifikáló eszközt fejleszt, amelyet a holnapi napon mutatunk be a Networkshopon.<sup>4</sup>

A névelem-felismerés, vagyis amikor egy szöveg egyes szavairól, kifejezéseiről megállapítjuk, hogy földrajzi, személy- vagy egyéb tulajdonnevek-e, vagy az érzelemfelismerés, amikor a szöveg mondatainak érzelmi töltetét a számítógép ismeri fel, a wikifikáció, amikor a szöveg szavait egy ontológia elemeihez kötjük: mindez nehezen ér el a nagyközönséghez, ameddig nem integrálják népszerű szolgáltatásokba. Mindez elmondható a konkrét gyűjteményi anyagokon finomhangolt szövegfelismerő eszközökről is. Ez a következő kultúrtechnika, amelyre előadásomban röviden kitérek.

---

2 Vö. pl. <https://ekk.org.hu/2021/10/05/az-uj-szerzoi-jogi-torveny-a-konyvtaroknak-jobb-1-resz/>

3 A kutatók eredményeiket itt ismertették: <https://milab.tk.hu/hirek/2021/06/2021-junius-17-gardos-judit-es-micsik-andras-eloadasa-a-tk-milab-speaker-series-sorozat-kovetkezo-allomasakent>

4 Nemeskey Dávid Márk, Palkó Gábor – Szemantikus névelem-azonosítás (NEL) magyar nyelvű szövegeken (a HuWikifier bemutatása)

## Írott szövegek MI alapú feldolgozása

### Finomhangolt karakterfelismerés (Optical Character Recognition – OCR)

A gyűjteményi és a kutatási szférában az esetek többségében olyan eszközökkel és munkamenetek során állítanak elő digitális szövegobjektumokat, amelyek végeredményeként MI feldolgozásra kevésbé alkalmas produktumok állnak elő (pl. képi és vagy szövegréteg szintjén gyenge minőségű PDF fájlok). Olyan munkamenet kidolgozására van szükség, amely már a gyűjteményi feldolgozás során tekintettel van a gépi szövegfeldolgozás igényeire. Az egyes nyelvekre optimalizált alapmodellek használata mellett szükség van ugyanakkor a gépi tanulás alapú karakterfelismerésre is. Le kellene váltani azt a gyakorlatot, hogy a könyvtárak és más gyűjtemények anélkül választanak eszközt, illetve alakítanak ki munkamenetet a nyomtatott anyagok OCR-ezésére, hogy mérlegelnék a technológia előnyeit és hátrányait, összevetnék a különféle szoftverek, különféle szoftverbeállítások eredményeit és a saját gyűjteményükre optimalizált eszközöket és munkameneteket dolgoznának ki. Egy dokumentumtípus vagy gyűjteményrész feldolgozásánál kis mennyiségű szöveg kézi javítása is az egész korpusz radikális szövegminőség-javulásával jár.

### Kézírás-felismertetés (Handwritten Text Recognition – HTR)

A digitális örökség diskurzusában a könnyen feldolgozható és közzé tehető nyomtatott, illetve digitálisan született anyagok mellett a „valódi” – vagyis kézzel írt – kéziratok háttérbe szorulnak, hiszen általános, nem az adott kézre vagy gyűjteményrészre hangolt modellekkel nem tehetők kereshetővé. Szerencsés fejlemény, hogy létezik egy könnyen kezelhető kézírás-felismerő keretrendszer, amelybe az OSZK munkatársainak jóvoltából már magyar kézírásra tanított kész modell is elérhető.<sup>5</sup> Az eszköz hátránya, hogy a szoftver nagyobb gyűjtemények feldolgozásánál igen költséges, azért csak néhány intézmény használja. Szükség van tehát ingyenesen elérhető, magyar nyelvre optimalizált, nagy gyűjteményi egységek feldolgozására is használható kézírás-felismerő eszközre.<sup>6</sup>

Aszemantikus szövegfeltárás és a gépi tanulással optimalizált írásfelismerés munkamenetének széles körű elérhetővé tétele a Digitális Örökség Nemzeti Laboratórium egyik fő feladata, de erre később, az intézményi fejlemények kapcsán térek vissza. Előbb még egy negyedik kultúrtechnika, a „born digital curation” témáját emelném ki.

### A digitálisan született anyagok kezelése („born digital curation”)

A magyar kulturális örökség nagy volumenű adatvesztést szenved el amiatt, hogy a magyar közgyűjtemények és kutatóhelyek többsége nem rendelkezik sem kompetenciával, sem infrastruktúrával, de még legtöbbször tervekkel sem a digitálisan született anyagok archiválására és kezelésére. A digitálisan született anyagok kezelése nem csak gyűjteményi vagy tudományos feladat, de piaci szereplők is végezhetik azt. A különféle, adott esetben elavult hordozókon tárolt, és/vagy elavult formátumú digitálisan rendelkezésre álló anyagok kezelése speciális szaktudást és eszközkészletet igényel. Nemzetközi összefogásban

5 A Transkribus szolgáltatás első magyarországi modellépítéséhez l.: Bobák, Barbara és Gábori Kovács, József (2019) Kézírásfelismerés Arany János levelein. In: Networkshop 2019. HUNGARNET Egyesület, Budapest, pp. 38-44. A PIM (majd OSZK) DBK projektjéről: Szűcs, Kata Ágnes (2021) Automatikus kézírás-felismertetés Kiss József levelezésén. In: Online térben az online térért : Networkshop 30: országos online konferencia. 2021. április 6-9. Eötvös Loránd Tudományegyetem. HUNGARNET Egyesület, Budapest, pp. 73-80.

6 L. ehhez a jelen konferencián: Lévai Dániel, Szekrényes István, Palkó Gábor: *Kézírásfelismerés saját modellek létrehozásával a Digitális Örökség Nemzeti Laboratórium szuperszámítógépén*

eszközök, szabványok és jó gyakorlatok sora jött létre, és ezek jelentős része ingyenes, nyílt eszköz.

Maga a probléma nem ismeretlen a hazai közgyűjteményi diskurzusban. Bár a levéltárak a tömegesen keletkező digitális iratok megőrzésére fejlett technológiákkal rendelkeznek, az egyedi dokumentumok és az adatkészletek feldolgozásában rosszabb a helyzet.

2017-ben az ELTE és a PIM közreműködésével szerveztem workshopot és konferenciát neves külföldi vendégelőadók részvételével, majd a PIM Digitális Bölcsészet Központ tavaly novemberben műhelykonferenciát tartott a témában. Igen öröndetes, hogy az OSZK-ba átkerült Központ – folytatva a fél évtizede megkezdett munkát – az e-mailek archiválásáról és kutathatóvá tételéről szervezett tutoriált.<sup>7</sup>

A Digitális Örökség Nemzeti Laboratórium keretei között a Magyar Nemzeti Levéltár, a Miskolci Egyetem, a BTK Irodalomtudományi Intézet és az ELTE közös projektet indított, amelynek célja a digitálisan született kutatási anyagok archiválása és közzététele a nyílt tudományosság jegyében. A holnapi napon elhangzó előadás az egyik pilot projektet ismerteti, melynek keretei között olyan, a digitális bölcsészet legveszélyeztetettebb anyagai között számontartható adatbázisok szakszerű kezelésének problémáit és nemzetközi jó gyakorlatait mutatja be, mint a kutatási célú adatkészleteket tároló relációs adatbázis (SQL).<sup>8</sup> Újabb pilot keretei között pedig egy megszűnt határon túli magyar nyelvű hírportál, a manna.ro webportál szakszerű archiválását és kutathatóvá tételét teszteljük.

És ezzel át is térnék előadásom rövidebb második részére, amely a téma szempontjából releváns intézményi fejlemények közül emel ki néhányat.

## Open Science, Open Data

Magyarországon a kutatási adatok kezelésével kapcsolatos diskurzus, a nyílt tudományosság jegyében az elmúlt években rendkívüli módon megélénkült.

Megalakult a kutatási adatok kezelését támogató Research Data Alliance (RDA) globális szervezet magyar csoportja, a Research Data Alliance Hungarian National Node (HRDA). Alapító tagjai a Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI), az MTA Könyvtár és Információs Központ (MTA KIK), a HUNgarian Open Repositories (HUNOR) és a Kormányzati Informatikai Fejlesztési Ügynökség (KIFÜ) - és nemrégiben a DH-LAB is csatlakozott a szervezethez. A KIFÜ pedig az EOSC – Európai Nyílt Tudományos Együttműködésben képviseli Magyarországot. Meetupok és workshopok során hallhattunk a kutatási adatok összefüggésében a FAIR alapelvekről, az Open Science fontosságáról és lehetőségeiről, illetve az adatrepozitóriumok jellemzőiről. (A napokban érkezett a meghívó a Nyílt tudományos fórum újabb, már negyedik workshopjáról!)

Egyre több intézmény alkalmaz vagy képez ki data steward-ot (adatgazdászt), és mind több kutatóhelyen merült fel egy nem pusztán publikációk, de különmemű kutatási adatelemek kezelésére is alkalmas – akár a szemantikus hálózatok kezelését is lehetővé tevő – intézményi vagy intézményközi adatrepozitórium létrehozásának terve. Ebbe a folyamatba a Digitális Örökség Nemzeti Laboratórium több szinten is bekapcsolódik.

Még a kutatási infrastruktúrák kiépítésénél is fontosabb talán a kutatók és az archívumi szakemberek képzése, ha a nyílt tudományosság jegyében a kutatási adatok felelős és

---

7 L. a PIM workshopjának programját itt:

<https://pim.hu/hu/esemenyek/digitalisan-letrejott-born-digital-keziratok-kezelese>

8 Dióssy Anna Laura, Alföldi István: Digitálisan született kutatási anyagok megőrzése: a relációs adatbázis mint born-digital objektum





intelligens kezelésének széles körű előmozdítása a célunk. A DH-LAB ezért az ELTE-n 2022-ben adatgazdász képzést indít.

A szakképzés célja olyan szakemberek képzése, akik a könyvtártudomány, a digitális bölcsészet és a könyvtári informatika határmezsgyéjén a digitalizált vagy eleve digitálisan keletkezett adatok kezelési technikáinak magas szintű elsajátítását követően képesek ezek egyéni, illetve csoportmunkában történő, a jogi keretek adta mozgástéren belüli, akár projektszerű kezelésére. Mindebbe beleértendő az adatok biztonságának és hitelességének megőrzése, a „fair use” körülményeinek megteremtése, valamint a kezelt adatok másodlagos felhasználási célokra történő kijelölése/felhasználása, az erre alkalmas rendszerek adatokkal történő feltöltése, üzemeltetése.

### Szemantikus technológiák szolgáltatássá fejlesztése

A fent vázolt képzés elengedhetetlen, hogy a kutatók és a memória-intézmények szakemberei tudják, a XXI. században hogyan kell és hogyan lehet az adatokkal bánni, de mindehhez természetesen szükség van dedikált infrastruktúrára. Számos adatrepozitórium-fejlesztési projekt indult el az elmúlt években. A SZTAKI és az ELKH például a Dataverse szoftverre építi a Concorda repozitóriumot, míg számos helyen a szemantikus web elveinek megfelelő Invenio RDM-et tesztelik. A KIFÜ és a DH-LAB mellett az ELTE-n is indul egy adatrepozitórium pilot. Az adatrepozitórium funkciókat messze meghaladja az a fejlesztés, amely elnyerte az NKFIH-ITI megtisztelő TOP50 kutatási infrastruktúra elismerést.

A Laboratórium a Monguz Információtechnológiai Kft.-vel együttműködve egy olyan hardver- és szoftver-infrastruktúrát fejleszt, amely lehetővé teszi a nemzeti kulturális örökség mesterséges intelligencia alapú feldolgozását, kutatását, oktatását és közzétételét saját fejlesztésű, magyar nyelvre optimalizált nyelvfeldolgozó alkalmazások segítségével.<sup>9</sup> A Nemzeti Kutatási és Fejlesztési és Innovációs Hivatal ezt az úttörő fejlesztést díjazta.

Egyetlen mondatban összefoglalva a felsoroltakat: a mesterséges intelligencia mint kultúrtechnika nemcsak a kulturális örökség, de a kutatási adatok kezelésének minden szintjén forradalmi változások kezdeményezője, ám az ezekben rejlő potenciált csak a fent vázolt intézményi együttműködések formájában leszünk képesek kihasználni.

<sup>9</sup> L. a konferencián: Kiss Tamás-Palkó Gábor: *Magyarország Ígéretes Kutatási Infrastruktúrája: A Digitális Örökség Nemzeti Laboratórium (DH-LAB) és a Qulto Kutatási Infrastruktúra (Qulto KI)*



A tudományos publikálás támogatása  
a Debreceni Egyetemi és Nemzeti Könyvtárban

Pergéné Szabó Enikő  
Debreceni Egyetem Egyetemi és Nemzeti Könyvtára (Debrecen)  
[perge@lib.unideb.hu](mailto:perge@lib.unideb.hu)

Bátfai Mária Erika  
Debreceni Egyetem Egyetemi és Nemzeti Könyvtára (Debrecen)  
[ebatfai@lib.unideb.hu](mailto:ebatfai@lib.unideb.hu)

A Debreceni Egyetem Egyetemi és Nemzeti Könyvtárban (DEENK) évtizedek óta küldetésünknek tekintjük az egyetemen folyó oktató- és kutatómunka támogatását. Hallgatói, oktatói és kutatói munkát segítő tréningek, kurzusok tartása mellett, a tudományos kutatás és publikálás nemzetközi trendjeinek megismertetéséért honlapot<sup>1</sup> és blogot<sup>2</sup> is üzemeltünk. Kiadónk, a Debreceni Egyetemi Kiadó az egyetemen készülő jegyzetek, szakkönyvek és folyóiratok hagyományos és elektronikus formátumú kiadása mellett azok széleskörű terjesztését is biztosítja.

Stratégiai célkitűzéseink közt elsődleges feladatunknak tekintjük az egyetemi tudásvagyon disszeminációját. Az iDEa Tudóstér kutatói platform<sup>3</sup> és az intézményi repozitórium működtetésével az egyetem kutatói teljesítményének hatékonyabb láthatóságát biztosítjuk.

Jelen tanulmányunkban az egyetemen megjelenő folyóirataink nemzetközi – elsősorban a DOAJ, Scopus és Web of Science - adatbázisokban való megjelenését támogató, a Nemzetközi Oktatást Koordináló Központtal (NOKK) duális szervezésben, 2019 őszén indított *Támogatási program az egyetemi folyóiratok Scopus-ba kerüléséhez* c. projekt keretében végzett munkánkat és azok eredményeit ismertetjük.

**Kulcsszavak:** láthatóság, nyílt hozzáférésű folyóiratok, OJS, indexelés, ranking, egyetemi könyvtár, egyetemi kiadó

The University of Debrecen University and National Library (DEENK) has been committed to supporting teaching and research at the University for decades. In addition to providing training and courses to support students, teachers and researchers, it operates a website and a blog to inform about international trends in scientific research and publishing. The Debrecen University Press - and its predecessor - has been publishing notes, textbooks and journals in print and, in the last decade, in electronic form since 1995. Among the DEENK's primary objectives, it considers the dissemination of the university's knowledge assets as one of its strategic priorities, and also ensures the visibility of the university's research achievements by operating the iDEa Knowledge Platform and its institutional repository.

---

1 <https://openscience.hu/>

2 <https://instantscience.hu/>

3 <https://tudoster.idea.unideb.hu/>



In this study, we present our work and results in increasing the international visibility of the journals published at the university, with a special focus on the multi-year programme “Support Programme for the inclusion of university journals in Scopus”, launched in autumn 2019 in a dual organisation with the Coordinating Center for International Education (NOKK).

**Keywords:** visibility, open access periodicals, OJS, indexing, ranking, university library, university press

A Debreceni Egyetem küldetésnyilatkozata szerint „a legmagasabb szintű, sokoldalú, multidiszciplináris képzéssel, kutatással és fejlesztéssel járul hozzá az egyetemes tudomány és a magyar társadalom fejlődéséhez.”<sup>4</sup>

Az egyetemek finanszírozásban *végbemenő változásoknak, a modellváltással vállalt publikációs és kutatási teljesítményeknek köszönhetően*, a korábbinál is fontosabb feladat lett a tudományos kutatómunka és a minőségi oktatás disszeminációja. A fenntartható finanszírozás, és a *pályázati források eléréséért* az egyetem hazai és külföldi rangsorokban elért helyzetének előremozdítása - amely a hazai és külföldi hallgatók elsőhelyes jelentkezésének arányait növelheti – az egyetem minden szervezeti egységében prioritással bír.

## 1. Tudományos folyóiratok láthatósága

A technológiai, politikai és tudományos környezetben végbemenő változásoknak köszönhetően a 2000-s évek végén megjelent a folyóiratok online, open access elérésének az igénye. A szerkesztőségek felismerték, hogy a folyóiratok online láthatósága, a tudományos eredmények gyors elérhetősége mellett az nyílt hozzáférés biztosítja a szerzők tudományos láthatóságát, emellett a kapcsolati háló építésének és a kutatástámogató pályázatok elnyerésének eszköze lehet.

A folyóiratok ismertségének és elismertségének egyik fontos mutatója a cikkekre történő hivatkozások száma, azonban a folyóirat online elérhetőségének biztosítása nem eredményezi automatikusan a nagyobb hazai és nemzetközi láthatóságot, a folyóirat online elérésének biztosításával nem törvényszerű a kutatók által is használt nemzetközi adatbázisokban való megjelenés.

Egy 2015-ben végzett felmérésünk szerint, változó rendszerességgel ugyan, de közel 60 folyóirat, évkönyvmegjelenések köthető az egyetem karaihoz, intézeteihez<sup>5</sup>, közelegyharmaduk a Debreceni Egyetemi Kiadónál jelenik meg. A statisztikai adatok azonban azt is mutatták, hogy a debreceni lapok nemzetközi láthatósága (a rendszeresen megjelenő lapok számát tekintve) rendkívül alacsony. Az open access folyóiratokat jegyző nemzetközi adatbázisban ([Directory of Open Access](https://doaj.org/) (DOAJ)<sup>6</sup>-ban) 2015-ben csupán egy debreceni folyóirat szerepelt, az egyetemi rangsorok adatait szolgáltató nemzetközi indexelő adatbázisok (Web of Science és a Scopus) is csupán egy egyetemi kiadványt indexeltek. Egy 2017-es tanulmány szerint az MTMT-ben regisztrált hazai szerzők cikkeinek közel harmada olyan folyóiratban jelent meg, amelyet nem követ sem a Web of Science (WoS), sem a Scopus<sup>7</sup>. Egy folyóirat sikeres

4 A Debreceni Egyetem küldetésnyilatkozata. Hozzáférés: 2022.06.19. <https://unideb.hu/debreceni-egyetem-kuldetesnyilatkozata>

5 Debrecenen megjelenő folyóiratok. Utolsó elérés: 2022.06.19. <https://lib.unideb.hu/hu/egyetemi-folyoiratok>

6 Directory of Open Access Journal. Hozzáférés: 2022.06.19. <https://doaj.org/>

7 Holl András – Bilcsi Erika: Nyílt publikálási szoftverek és platformok. In: *Tick József - Kokas Károly - Holl András (szerk.): Workshop 2019: 2019. április 23-26.* Széchenyi István Egyetem, Győr, Hungarnet, 2019. pp. 54-60.

indexelésének (és bent maradásának) magas szakmai és technikai követelményei vannak<sup>1</sup>, amely szempontoknak egyetemi folyóirataink nem minden pontban tudnak eleget tenni.

## 2. Folyóiratkiadás a Debreceni Egyetemen

A Debreceni Egyetemen a [Debreceni Egyetem Kiadói Szabályzata](#)<sup>2</sup> rendelkezik az egyetemen megjelenő folyóiratok, könyvek kiadási jogáról. Az Egyetemen (és jogelőd intézeteiben) hosszú múltra tekint vissza azoknak a tudományos folyóiratoknak, aktáknak a kiadása, amelyek az intézmény egy-egy szervezeti egységéhez kötődnek. Bár az Egyetemen 1995-ben jött létre a Kossuth Egyetemi Kiadó - mely az egyetemi karok integrációja (2000) után Debreceni Egyetem Egyetemi és Nemzeti Könyvtárának (DEENK) részeként [Debreceni Egyetemi Kiadó](#) (DUPress)<sup>3</sup> néven biztosítja az egyetemi jegyzetek, szakkönyvek kiadását - napjainkban is jelentős számú a kari, intézeti kiadású folyóiratok száma.

## 3. Online folyóiratplatform(ok) a DUPress-ben

Jövőképünk szerint, hiszünk a nyílt hozzáférés tudományepeszdtő erejében, a DUPress, a szerzői jogok maradéktalan betartása mellett, minél zöldebb<sup>4</sup> kiadóvá kíván válni. A következő évek célkeresztjébe az egyetem tudományos teljesítményének korszerű és magas színvonalú, hosszú távú lokális és országos megjelenítése mellett a nemzetközi láthatóság kerül, kiadványainkkal növelni kívánjuk az Egyetem nemzetközi rangsorokban elfoglalt helyét.

A Kiadó 2010-ben kezdte el az elektronikus folyóiratkiadást, a tudományos, open access folyóiratok online szerkesztésére és publikálásra alkalmas, [Open Journal System \(OJS\)](#)<sup>5</sup> alapokra épülő, saját fejlesztésű, magyar nyelvű platform, a Métisz használatával. A Métisz célja a magyar nyelvű egyetemi folyóiratok online megjelenésének biztosítása volt.

A Métisz által támogatott online szolgáltatási környezetet 7 – elsősorban az Egyetemi Kiadónál hagyományos formában is megjelenő - egyetemi folyóirat használta. A folyóiratok többsége továbbra is az intézményi honlapokon való elhelyezést választotta online megjelenési formaként, ami a gyakorlatban a legtöbb lap számára a tanszéki vagy intézeti honlapon megjelenő folyóirat tartalomjegyzékének megjelentetését jelentette, néhány esetben az évfolyamokat letölthető formátumban is elérhetővé tették. A honlapokon azonban nem tüntették fel a szerző(k) felhasználási jogait, mint ahogyan a felhasználási licenc megadása is

---

1 DOAJ kritériumok: <https://doaj.org/apply/guide/>

Scopus kritériumok: <https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection>

Web of Science kritériumok: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/core-collection/editorial-selection-process/>

2 Debreceni Egyetem Kiadói Szabályzat. Hozzáférés: 2022.06.19.

<https://mad-hatter.it.unideb.hu/portal/displayDocument/id/2012825>

3 <https://dupress.unideb.hu/hu/fooldal/>

4 Metz, Katja – Seadle, Michel (2012): Green publishing with green technologies. *Library Hi Tech*. Vol.30., Issue 3, pp. 381-383. <https://doi.org/10.1108/07378831211266528>

5 Az tudományos folyóiratok online közzétételét és a szerkesztőségi folyamatok támogatását biztosító nyílt forráskódú szoftvert 2001-ben alkották meg a University of British Columbián, egy projekt -az open access mozgalom szellemiségét támogató - Public Knowledge Project (PKP) projekt keretében. Saját fejlesztésre a magyar nyelvű platform és útmutató hiánya és az egyetemi folyóiratok sajátos szerkesztőségi folyamatai miatt volt szükség.



elmaradt. A folyóiratok adatainak rendszeres frissítését, a cikkek állandó, tartós elérését, a hosszú távú megőrzését a szerkesztőségek nem tudták biztosítani<sup>6</sup>.

#### 4. A Debreceni Egyetem Publikációs Platformja

A DEENK az egyetemi tudásvagyon megőrzése mellett az egyetemi publikációk, így az egyetemi folyóiratok láthatóságának növelését is célul tűzte ki. Ez nem egyedi kezdeményezés, az ország más könyvtárai olykor az egyetemi kiadóval karonfogva, is ebbe az irányba indultak el<sup>7</sup>.

Megvizsgálva a tudományos folyóiratokat szolgáltató nemzetközi adatbázisok kritériumrendszerét, elsődleges feladatnak tekintettük olyan online megjelenést is biztosító platform biztosítását, amely automatizmusokon keresztül lehetővé teszi az adatbázisok közötti átjárhatóságot, átláthatóvá teszi a szerkesztőségi munkafolyamatokat, és biztosítja az archiválást, a hosszú távú megőrzést. A Métisz ezeket az elvárásokat részben a fejlesztések hiányosságai, részben pedig a rendszer egyedi sajátosságai miatt nem tudta teljesíteni, így könyvtár stratégiai vezetése úgy döntött, hogy a nemzetközi tudományos folyóiratok körében már nagy népszerűséggel és elismeréssel bíró, és a hazai folyóiratkiadási környezetben is használt Open Journal System 3.1-es verziójával folytatja a folyóiratok kiadását.

2016-ban alakítottuk ki a [Debreceni Egyetem Publikációs Platformját](#)<sup>8</sup>, ami egyrészt a szerkesztőségi munkafolyamatok teljes menedzselését, az egyes munkafolyamatok pontos nyilvántartását és automatizálását is lehetővé teszi; másrészt az OJS-sel a tudományos folyóiratoktól elvárt, szabványos megjelenési felületet biztosít a rendszerhez csatlakozó folyóiratok számára. A szolgáltatás igénybevételével a kiadványok egy önálló webcímmel rendelkező kész honlapot kapnak, amely olvasói felülete teljesen reszponzív, azaz a legkülönbözőbb eszközökön is optimális megjelenést, könnyű olvashatóságot biztosít. Az OJS ezen verziója már biztosítja a lapszámok és cikkek folyamatos elérését, megosztását a különböző közösségi platformokon, kompatibilis a hivatkozáskezelő alkalmazásokkal<sup>9</sup>. A program által biztosított automatizmusoknak köszönhetően a Google Scholar, a Crossref és az ORCID kapcsolat is biztosított. A könyvtár a platform térítésmentes működtetésével, a rendszer használatával és a folyóiratok formai megjelenésével kapcsolatos technikai és módszertani tanácsokkal segíti a szerkesztőségeket.

Aszabványoknak és a nemzetközi elvárásoknak megfelelő formai megjelenés után támogatást nyújtunk a folyóiratoknak a különböző nemzetközi adatbázisokban való regisztrációhoz is. (Ezt a segítséget megkapják azok a folyóiratok is, amelyek a közös egyetemi felületet nem használják.)

A módszertani támogatás első lépéseként útmutatókat készítettünk a szerkesztőségek számára, az open access szerkesztőségi folyamatok, a bírálati űrlapok, a copyright és a licenc-ek meghatározásán át a folyóiratok online felületének kialakításáig.

A legnagyobb és legfontosabb feladatunknak a folyóirat nemzetközi adatbázisokba való indexeltetését tekintjük. A DEENK kezdeményezését (a közös egyetemi platform kialakítását) kedvezően fogadták az egyetemi karok vezetői és a folyóiratok szerkesztőbizottságai. Míg 2018 végén csupán 3 folyóiratot menedzselünk a platformon, 2019-ben már 15 folyóirat

6 Holl András (2022): A hazai tudományos eredmények láthatóvá tétele, kiaknázása és megőrzése modern eszközökkel. *Magyar Tudomány*. 183. évf. 1. sz. pp. 69–78. <https://doi.org/10.1556/2065.183.2022.1.6>

7 Kilián Zsolt – Margit István (2021): Open Access kiadás könyvtári támogatása egyetemi környezetben. *Tudományos és műszaki tájékoztatás*, 68. évf. 7. sz. <https://tmt.omikk.bme.hu/tmt/article/view/13063>

8 Debreceni Egyetem Publikációs Platform. <https://ojs.lib.unideb.hu/>

9 Páyer Barbara (2017): A megújuló online folyóirat-kiadás. *Tudományos és műszaki Tájékoztatás*. 64. évf. 11.sz. pp. 548-559. <http://real.mtak.hu/70283/>

jelent meg ezen a felületen, mára pedig több mint 30 egyetemi folyóirat megjelentetését és elérhetőségét biztosítjuk.

## 5. Támogatási program az egyetemi folyóiratok Scopus-ba kerüléséhez

Az egyetem nemzetközi rangsorokban való megjelenésével, az adatok elemzésével, a rangsorokban való előrelépést segítő akciótervek kidolgozásával a *Debreceni Egyetem Nemzetközi Oktatási Kordinációs Központja (NOKK)* foglalkozik. A NOKK látva az DEENK egyetemi folyóiratok nemzetközi láthatóságával kapcsolatos törekvéseit, olyan program kidolgozására kérte fel a könyvtárat, amelynek segítségével belátható időn belül megemelkedik a debreceni folyóiratok nemzetközi láthatósága. A könyvtár a program kialakításához az Utrechti Egyetemi Könyvtár open access folyóiratkiadásra kidolgozott projektjét vette példaként<sup>10</sup>. A holland inkubátorprogram egy hatéves időszakot határozott meg arra, hogy az open access folyóiratok tudományos láthatóságának növelésével egyidőben folyóirataik számára fenntartható pénzügyi modellt dolgozzanak ki. A több évre szóló akciótervben könyvtárunk az utrechti példa mellett az egyetemi rangsorokat készítő vállalatoknak az egyetem publikációs adatairól elemzést adó Scopus, valamint az open access folyóiratokat minősítő DOAJ kritériumait alapul véve egy 8 mérföldkőből álló keretrendszert készített, a NOKK pedig a programban részt vevő folyóiratok DEENK által javasolt feladatainak teljesítéséhez pénzügyi támogatását, biztosítását vállalta.



**DEBRECENI  
EGYETEM**

**DEBRECENI EGYETEM  
EGYETEMI ÉS NEMZETI KÖNYVTÁR**  
H-4002 Debrecen, Egyetem tér 1, Pf.: 400  
Tel.: 52/410-443, e-mail: office@lib.unideb.hu

### A DE FOLYÓIRATTÁMOGATÁSI PROGRAMJÁNAK MÉRFÖLDKÖVEI

- LÉPÉS**  
Folyóiratok döntése és **szándéknyilatkozat aláírása a főszerkesztők és a kari dékánok által** a Scopus-ba való bekerülést célzó támogatási programban való részvételtől és a DE OJS (Open Journal System - ojs.lib.unideb.hu) platformra való áttérésről.  
**ÉVES TÁMOGATÁS**  
Felhasználási terület: személyi költségek (lektor, szerkesztő, korrektor, törődő munkadíj), arculati elemek tervezése
- LÉPÉS – HATÁRIDŐ: 6. HÓNAP**  
A DEENK valamennyi, a programban résztvevő folyóiratról **készít egy részletes helyzetelemzést**, amely többek között kitér a következőkre:
  - folyóirat Scopus-os hivatkozásainak száma
  - DOI szám megléte
  - Open Access-e a folyóirat
  - Indexelve van-e valahol
  - MTMT-ben rendben van-e
  - MTA osztály jegyzékén szerepel-e
  - Scopus által támasztott minimum elvárások teljesítésével hogyan áll - egyedi ütemterv az összes elvárás teljesítéséhez
- LÉPÉS – HATÁRIDŐ: 12. HÓNAP**  
Részvevő folyóiratok **DE OJS-re (Open Journal System - ojs.lib.unideb.hu) való áttérése** - DEENK biztosítja a felületet és a szakmai útmutatást, de a folyóiratok szerkesztői töltik fel a tartalmakat és kezelik a későbbiekben is a saját oldalukat.  
**EGYSZERI BÓNUSZ**
- LÉPÉS – HATÁRIDŐ: 12. HÓNAP**  
**Lektorált, angol nyelvű absztrakttal rendelkező, minimum 9 cikket tartalmazó, külföldi szerző(k) bevonásával készült lap megjelenése az év alatt vállalt darabszámban (rendszerességgel)**  
**EGYSZERI BÓNUSZ**
- LÉPÉS – HATÁRIDŐ: 12. HÓNAP**  
A DEENK által leírt **helyzetelemzéshez kapcsolódó, folyóiratonként eltérő további feladatok teljesítése a 12. hónap végére.**  
**EGYSZERI BÓNUSZ**
- LÉPÉS – HATÁRIDŐ: 24. HÓNAP**  
Az egy évre **vállalt rendszerességgel** való megjelenés (lektorált, angol nyelvű absztrakttal rendelkező, minimum 9 cikket tartalmazó, külföldi szerző(k) bevonásával készült) + a **szerkesztők minimum 40%-a külföldi** (legalább egy Európán kívüli) és **minimum 40%-nyi külföldi szerzős cikk** minden megjelenet lapszámban.  
**EGYSZERI BÓNUSZ**
- LÉPÉS – HATÁRIDŐ: 24. HÓNAP**  
A DEENK által leírt **helyzetelemzéshez kapcsolódó, folyóiratonként eltérő további feladatok teljesítése a 24. hónap végére** - Scopus-ba való jelentkezés elindítása (először „pre evaluation”, majd a javaslatok végre hajtása után a tényleges jelentkezés elindítása).  
**EGYSZERI BÓNUSZ**
- LÉPÉS – HATÁRIDŐ: 36. HÓNAP**  
**Scopus-ba való bekerülés.**  
**EGYSZERI BÓNUSZ**

### 1. ábra DETamfi mérföldkövek

<sup>10</sup> Sondervan, Jeroen – Stigter, Fleur (2018): Sustainable open access for scholarly journals in 6 years—the incubator model at Utrecht University Library Open Access Journals. *Learned Publishing*. Vol. 31. Issue 3. <https://doi.org/10.1002/leap.1151>





A mérföldkövek tartalmazzák a folyóirat – lehetőleg kettős vak – lektorálási szempontjait, a rendszeres megjelenést, a kiadói etikai nyilatkozat meglétét, nem angol nyelvű cikkek esetén az angol nyelvű absztraktot és a tárgyszavakat. Mivel a folyóirat Scopusba való bekerülésének fontos szempontja a szerkesztők és a szerkesztőbizottság szakmai tekintélye és nemzetközi összetétele<sup>11</sup>, ennek biztosítása a szerkesztőbizottságok kiemelt feladata lett. A NOKK kérésére a könyvtár a mérföldkövek teljesítését 3 éves időkereten belül határozta meg.

A mérföldkövek a folyóiratszerkesztőségek által is elfogadott feladatok teljesítését az DEENK felügyeli, a programban részt vevő folyóiratok eredményeinek függvényében a NOKK a szerkesztőségek számára pénzügyi keretet biztosít, amit szabadon felhasználhatnak olyan tevékenységekre, amelyek előmozdíthatják az online folyóirat formai, minőségi és tartalmi megjelenését. Ilyen feladat pl. az OJS platformra való átköltözés a teljes szerkesztőségi munkafolyamat implementálásával, az archív számok feltöltése, a folyóirat designjának kialakítása vagy a lektorok, technikai munkatársak, fordítók munkadíjának kifizetése.

A program indulása előtt megvizsgáltuk a folyóiratok megjelenésének gyakoriságát és rendszerességét, az angol nyelvű absztraktok és kulcsszavak meglétét, a szerzők nemzetközi összetételét, és a folyóiraatra történő Scopus hivatkozások számát. A folyóiratszerkesztőségek és a karok vezetői megkapták a támogatási program feltételeit és a programban előírt feladatokat, elvárásokat. Az elemzés eredményei alapján a kari vezetőségek döntöttek el, melyik folyóirat fenntarthatóságát tudják a 3 éves program ideje alatt kari szinten is biztosan támogatni. Ezek után (2019 őszén) 25 folyóirattal indult el a DEENK és a NOKK első közös projektje.

Az első év mérföldkövei közé tartozott a folyóirat honlapjának kialakítása vagy megújítása. A honlapokon a folyóiratok küldetésnyilatkozata mellé felkerültek az open access nyilatkozatok, a copyright politikák és a licencek. Az egyetemi folyóiratok közül egyre többen kezdték el használni az OJS-t. A DEENK a folyóiratoknak online ISSN számot és DOI-t regisztrált. A folyóirat szerkesztőségek számára a rendszeres megjelenés mellett az első év kiemelt feladata volt a szerzők és a szerkesztőbizottsági tagok nemzetköziesítése, minél több külföldi szerző és szerkesztőbizottsági tag bevonása.

Az első év végén megvizsgáltuk a folyóiratok „teljesítményét”, az első évre meghatározott elvárások teljesítését. Az egy éves projektidőszakban 7 folyóirat nem tudta a rendszeres online megjelenést biztosítani, így a NOKK további pénzügyi finanszírozási szakaszaiban már nem vesznek részt. A második évtől a támogatási program 17 folyóirattal<sup>12</sup> folytatódott. A második, és a jelenlegi harmadik évben a szerkesztőségek kiemelt feladata a DOAJ és a Scopus regisztrációs folyamatok előkészítése, az ezzel kapcsolatos folyamatok felülvizsgálata és a regisztráció elindítása.

11 Dudás Anikó (2017): Folyóirat – ismertség, elismertség, indexelés. *Tudományos és Műszaki Tájékoztatás*. 62. évf. 7-8. sz. pp. 251-274.

12 Acta Agraria Debreceniensis (MÉK), Acta Classica Universitatis Debreceniensis (BTK), Abstract: Applied Studies in Agribusiness and Commerce (GTK), Central European Journal of Educational Research (BTK), Ethnographica et folkloristica Carpathica (BTK), Gerundium (ÁOK), Hungarian Journal of English and American Studies (BTK), International Journal of Horticultural Science (MÉK), International Review of Applied Sciences and Engineering (MK), Italianistica Debreceniensis (BTK), Kitaibelia (TTK), Különleges Bánásmód (GYGYK), Magyar Nyelvjárások (BTK), Metszetek (BTK), Publicationes Mathematicae Debrecen (TTK), Slavica (BTK), Stadium



## 5.1 Tapasztalatok

A folyóiratszerkesztőségekben szakmájuk és hivatásuk elismert tekintélyei kitartó, de a tudományos ranglétrán való előmenetelük szempontjából kevésbe elismert munkával működtetnek egy-egy folyóiratot. Az OJS-be költöző folyóiratok szerkesztőségeire háruló napi feladatok ugyan jelentősen optimalizálódtak, ám maga az online platform-váltás (nem egy esetben első online megjelenés) valamint a Scopus szempontrendszere szerinti formai, leíró elemek, a különböző nyilatkozatok (az indexelés oldaláról kiemelt fontosságú a publikációs és etikai nyilatkozat) és tájékoztatók megírása jelentős plusz terhet rótt a szerkesztőségi tagokra, nem egy esetben alapjaiban változtatta meg a folyóirat működését a megjelenési periódustól a bírálati mérföldkövek ütemezésén át a cikkek szerkezetéig és formai megjelenéséig.

A könnyen ütemezhető és teljesíthető elvárásokon (pl. főszerkesztő presztízsének igazolása; angol absztrakt és kulcsszavak; a szerző affiliációjának kötelező feltüntetése a cikkekben; a nyílt hozzáférésről, archiválásról nyilatkozat; átlátható és kiszámítható bírálati folyamat) túl, hosszú távon a rendszeres megjelenés (az egyes számokba megfelelő számú és minőségű cikk) valamint a szerzők és lektorok folyamatos toborzása jelent kihívást.

Könyvtárunk nemcsak a célokat és feladatokat határozta meg, hanem a fent említett részfeladatokhoz egyfajta dentritként közvetítette a különböző szerkesztőségeik által újonnan kialakított jó megoldásokat. Ízelítő a praktikákból:

- a leíró részek előzetes megírása;
- a leíró részek megírása munkamegosztással;
- nemzetközi folyóiratok honlapjain elhelyezett nyilatkozatminták elemzése, implementálása saját környezetbe (NEM másolása);
- ütemezett, gyakori és rendszeres szerkesztőségi ülések;
- az adminisztrációs felülettel lassabban barátkozó szerzők, lektorok mentorálása;
- közösségi média tervezett és rendszeres használata;
- a Scopus indexelés előtt pl. a DOAJ regisztrációval tesztelni nemzetközi adatbázisokba való bekerülési feltételeknek való megfelelést.

## 5.2 Javaslatok

A program a végéhez közeledik, ám az egyetemi kutató- és oktató munka, annak disszeminációja, nem véges folyamat, ahogy az egyetemi könyvtár kutatástámogató küldetése sem. A könyvtár oldaláról folyamatos volt és lesz az eddig vázolt módszertani támogatás, emellett több irányt is látunk, amivel folyóiratainak láthatóságát növelni lehetne:

- tudományterületét ismerő szaknyelvi lektor alkalmazása; esetleg ilyen jellegű online szolgáltatás előfizetése;
- (web)grafikus, tördelő alkalmazása;
- minőségi tördelő, képszerkesztő szoftverek használatával magasabb formai és tartalmi szintre emelhető a cikkek, folyóiratok tudományos megjelenése;
- a folyóiratok arculatkialakításához az üzleti világban is használt webergonómiai módszerek alkalmazása;
- a tudományos írást, publikálást, lektorálási feladatokat segítő kurzusok indítása;
- marketinget segítő jó gyakorlatok megosztása.



## 6. Eredmények és zárszó

A program jelenleg a 3. év közepén tart. Ez idő alatt:

- kialakítottuk a folyóiratok teljesítményelemző rendszerét;
- évente többször a folyóiratokra szabott helyzetelemzéseket készítettünk és pontos menetrendeket adtunk a további feladatok sikeres teljesítéséhez;
- workshop-okat és online szakmai előadásokat szerveztünk és tartottunk;
- a NOKK-kal közös menetrendet alakítottunk ki az támogatás folyamatos biztosítása érdekében.

Napjainkra a programban részt vevő valamennyi folyóirat rendelkezik DOI azonosítóval, 13-an az OJS platformot használják a folyóirat megjelenítésére, minden folyóirat „nemzetköziesítette” a polgárságát, azaz határon túli szerzők és szerkesztőbizottsági tagjainak száma emelkedett, a szerkesztőségek elkészítették a szerzők felhasználási szerződéseit, és a cikk elején felsorolt további nyilatkozatokat.

A DEENK módszertani támogatásának, a NOKK motiválásának és a szerkesztőségek kitartó munkájának sikereként könyveljük el, hogy 2016-tól 10 egyetemi folyóirat került be a DOAJ adatbázisába. A Scopus-ban korábban már indexelt folyóirat (Publicationes Mathematicae Debrecen (DE Természettudományi Kar) és a támogatási program indításával egyidőben bekerülő, a Debreceni Egyetem Műszaki Kara és az Akadémia Kiadó közös kiadásában megjelenő International Review of Applied Sciences and Engineering folyóiratok mellé a támogatási programnak köszönhetően újabb 2 egyetemi folyóirat indexelését kezdte meg a Scopus 2022 elején<sup>13</sup>.

Bízunk benne, hogy a jelenleg a Scopus minősítési eljárás utolsó szakaszában lévő folyóirataink<sup>14</sup> hasonlóan kedvező elbírálást kapnak.

### Felhasznált irodalom

1. A Debreceni Egyetem küldetésnyilatkozata. <https://unideb.hu/debreceni-egyetem-kuldetesnyilatkozata> Hozzáférés: 2022.06.19.
2. DUDÁS Anikó (2017): Folyóirat – ismertség, elismertség, indexelés. *Tudományos és műszaki tájékoztatás*. 62. évf. 7-8. sz. pp. 251-274.
3. HOLL András (2022): A hazai tudományos eredmények láthatóvá tétele, kiaknázása és megőrzése modern eszközökkel. *Magyar tudomány*. 183. évf. 1. sz. pp. 69-78. <https://doi.org/10.1556/2065.183.2022.1.6> Hozzáférés: 2022.06.01.
4. HOLL András – BILICSI Erika (2019): Nyílt közzevetési szoftverek és platformok. In: *Tick József - Kokas Károly - Holl András (szerk.): Networkshop 2019: 2019. április 23-26. Széchenyi István Egyetem. Győr, Hungarnet*, pp. 54-60. <https://doi.org/10.31915/NWS.2019.7>
5. KILIÁN Zsolt – MARGIT István (2021): Open Access kiadás könyvtári támogatása egyetemi környezetben. *Tudományos és műszaki tájékoztatás*, 68. évf. 7. sz. <https://tmt.omikk.bme.hu/tmt/article/view/13063> Hozzáférés: 2022.06.09.
6. METZ, Katja – SEADLE, Michel (2012): Green publishing with green technologies. *Library Hi Tech*. Vol.30. Issue 3, pp. 381-383. <https://doi.org/10.1108/07378831211266528> Hozzáférés: 2022.06.12.

13 The Hungarian Journal of English and American Studies (DE Bölcsészettudományi Kar) és a Kitaibelia (DE Természettudományi Kar).

14 Acta Classica Universitatis Scientiarum Debreceniensis (DE Bölcsészettudományi Kar); The International Journal of Horticultural Science (DE Mezőgazdaság-, Élelmiszertudományi és Környezetgazdálkodási Kar), valamint a Slavica (DE Bölcsészettudományi Kar)

7. PÁYER Barbara (2017): A megújuló online folyóirat-kiadás. *Tudományos és műszaki tájékoztatás*. 64. évf. 11. sz. pp. 548-559. <http://real.mtak.hu/70283/> Hozzáférés: 2022.05.27.
8. SONDERVAN, Jerien – STIGTER, Fleur (2018): Sustainable open access for scholarly journals in 6 years–the incubator model at Utrecht University Library Open Access Journals. *Learned Publishing*. Vol. 31. Issue 3. <https://doi.org/10.1002/leap.1151> Hozzáférés: 2022.06.07.

## Nemzetközi kiadványazonosítók és kötelezpéldányok kezelése az OSZK OKP (Országos Könyvtári Platform) rendszerében

Csirmazné Rezi Éva  
Országos Széchényi Könyvtár  
[rezi.eva@oszk.hu](mailto:rezi.eva@oszk.hu)

**Absztrakt:** Régi vágya az Országos Széchényi Könyvtárnak, hogy a nemzeti könyvtári feladatok ellátására fejlesztett különböző adatbázisok és nyilvántartások egy rendszerben legyenek kezelve. Különös tekintettel arra a törekvésre, hogy a már egyszer valahol rögzített kiadványadat vagy kiadói információ ne vesszen el, hanem kísérje végig a kiadványt a feldolgozási munkafolyamaton.

Az Országos Könyvtári Platform úgynevezett KAPU moduljának azon részét mutatom be, amely az OSZK Kötelezpéldány- és Kiadványazonosító-kezelő Osztály munkafolyamatait érinti. A KAPU modulnak – többek között – feladata a nemzetközi kiadványazonosítók (ISBN, ISMN, ISSN) kiutalása és a kapcsolódó kiadói adatok biztonságos kezelése, a kiadványazonosítók felhasználásának ellenőrzése, továbbá a kötelezpéldányok érkeztetési és szétosztási adatainak nyilvántartása.

**Kulcsszavak:** Országos Könyvtári Platform, Országos Széchényi Könyvtár, nemzetközi kiadványazonosítók (ISBN, ISMN, ISSN), kötelezpéldány

### Bevezető

Az Országos Széchényi Könyvtárban már korábban is készültek különböző tervek és specifikációk arra vonatkozóan, hogy a nemzetközi kiadványazonosítók kezelését, hogyan lehetne összekapcsolni a kötelezpéldányok érkeztetésével, illetve az így létrejött rekordkezdeményeket, milyen módon lehetne áttölteni a nemzeti könyvtár integrált könyvtári rendszerébe.

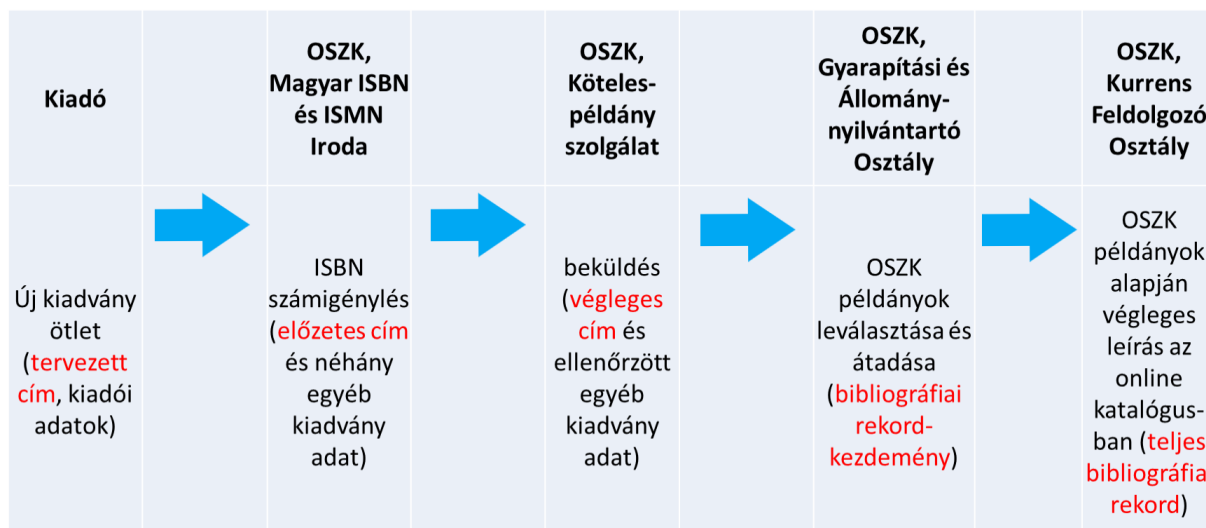
Az Országos Könyvtári Platform (a továbbiakban OKP) egy olyan új szemléletű, rugalmas könyvtári rendszer, amely bővíti az integrált könyvtári rendszerek funkcióit, megvalósítja a nemzetközi kiadványazonosítók (ISBN, ISMN, ISSN) és kötelezpéldányok kezelését.

### Mi volt a cél a kiadványazonosítók és a kötelezpéldányok nyilvántartásának megújításával?

- Nyilvántartási rendszereink korszerűsítése (több adat tárolásának lehetősége és azok több szempontú visszakereshetőségének biztosítása).
- Partnereink (testületi- és magánkiadók, kötelezpéldányból részesülő könyvtárak) munkájának hatékonyabb támogatása.
- A könyvtári feldolgozó munkafolyamatok felgyorsítása.
- Az információk (dokumentumok bibliográfiai metaadatai, kiadványok kiadáshoz és kötelezpéldány elosztáshoz kapcsolódó adatok) egyszeri rögzítésének biztosítása és a párhuzamos nyilvántartások és munkafolyamatok megszüntetése.
- A különböző nyilvántartások összekapcsolása oly módon, hogy az OSZK integrált rendszerének is részévé váljanak.

Hogyan tudjuk biztosítani az egyszeri adatrögzítést és az információk akadálymentes továbbítását? Hogyan tudjuk megakadályozni a munkafolyamatok ismétlődését?

Legrövidebben a következőképp válaszolhatunk erre: az eddig különálló nyilvántartásainkat összekapcsoljuk és egy rendszerben kezeljük az adatokat. Ez így nagyon egyszerűen hangzik, de a megvalósítás során sok problémába ütköztünk. Az 1. számú ábrán nyomon követhető, hogy az összekapcsolt és újragondolt nyilvántartásunkban mit is jelent az információáramlás egy adott mű kiadása során. Vagyis, hogyan jutunk el egy kiadói ötlettől a könyvtári rendszerben létrejövő bibliográfiai rekordhoz?



1. ábra Hogyan lesz egy kiadói ötletből bibliográfiai rekord?

Természetesen nem csak kiadványadatokat kezelünk az új rendszerben, hanem partneri, pl. kiadói adatokat is, amelyek a szigorú GDPR-szabályozás miatt fokozott figyelmet igényeltek a rendszer kialakítása során.

Miben jelent újdonságot az OKP rendszer Kapu modulja az osztályunk munkafolyamatait illetően?

Elsősorban abban, hogy az OSZK integrált rendszerének részeként, egy adatbázisban kezeli mind a nemzetközi azonosítók számkialakítását, mind a számkialakításhoz kapcsolódó kiadók nyilvántartását, illetve a kötelesempéldányok érkezését és szétosztását is. A kötelesempéldány-érkezéskor létrejött bibliográfiai rekordkezdemény pedig jó alapot nyújt a feldolgozó könyvtáros kollégák számára a teljes bibliográfiai rekord elkészítéséhez.

A nemzetközi kiadványazonosító számok (ISBN, ISSN, ISMN) magyarországi bevezetése különböző időpontokban történt. Ezért is kapcsolódtak különböző nyilvántartások, majd később adatbázisok a számkialakításokhoz. Hazánkban 1974-ben kezdték el alkalmazni az ISBN- (International Standard Book Number), majd 1977-től az ISSN- (International Standard Serial Number) azonosítókat. A nemzetközi gyakorlatban az ISBN-t már 1970-től, míg az ISSN-t, illetve annak elődjét, az ISDS-azonosítót 1972-től használták. Az ISMN- (International Standard Music Number) azonosítók alkalmazása 1993-ban indult el, mely rendszerhez 2001-ben csatlakozott Magyarország. A kiadványazonosítók használatát nemzetközi szabványok (lásd 2. ábra) írják elő, ugyanakkor hazánkban jogszabály nem teszi kötelezővé ezeknek az alkalmazását.

Azonosító rövid név	Azonosító teljes név	Milyen kiadványtípusra alkalmazzák?	Nemzetközi alkalmazás kezdete	Magyarországi alkalmazás kezdete	Szabvány
ISBN	International Standard Book Number	könyv, térkép (nyomtatott és elektronikus)	1970	1974	MS ISO 2108
ISSN	International Standard Serial Number	időszaki kiadványok, integráló kiadványok (nyomtatott és elektronikus)	1972	1977	MS ISO 3297
ISMN	International Standard Music Number	kották, zeneművek (nyomtatott és elektronikus)	1993	2001	MS ISO 10957

2. ábra Nemzetközi kiadványazonosítók

A kiadói partneri nyilvántartás az ISBN-számkiutalásokkal egyidőben kezdett el épülni azoknak a kiadóknak az adataiból, amelyek bármikor is ISBN-számigényléssel fordultak a Magyar ISBN Nemzeti Központoz. A jelenleg hatályos 717/2020. (XII. 30.) Korm. rendeleten<sup>1</sup> alapuló kötelezpéldányokból való részesülés pedig a nemzeti könyvtár alapításától (1802-től) fontos eszköze az állománygyarapításnak. Természetesen a kötelezpéldányként beküldött kiadványok nem minden példányát a nemzeti könyvtár őrzi, hiszen a kormányrendelet alapján más országos szakkönyvtárak és egyes egyetemi könyvtárak is részesülnek a kötelezpéldányokból. Többek között ezért volt fontos, hogy a kötelezpéldány-nyilvántartás is része legyen az OKP Kapu moduljának.

### Milyen követelményeknek kellett megfelelnie az új rendszernek?

A fejlesztő munkatársak nem voltak egyszerű helyzetben. Nemcsak az eddig jól működő számkiutaló rendszereket, a kötelezpéldány-érkeztetést és elosztást kellett összekapcsolniuk, hanem új elvárásoknak is meg kellett felelni. A legtöbb fejtörést a kiadványazonosító számok nem megfelelő használata és a hiányos impresszumú kiadványok kezelése okozta.

### Az OKP ISBN/ISMN-kezelő része felé támasztott alapkövetelmények

Mivel az ISBN- és az ISMN-azonosítók felépítése hasonló, így a számkezelésekhez kapcsolódó követelmények is szinte teljesen megegyeztek:

- ISBN/ISMN-számkiutalás kezelése előzetes adatok alapján. A kiadók által megadott kiadói és kiadvány adatok alapján kiadói rekordok és előzetes kiadványrekordok létrehozása.
- ISBN/ISMN-számellenőrzés biztosítása a beérkezett kiadványok alapján. A kötelezpéldányként vagy egyéb forrásból beszerzett kiadványok alapján a felhasznált ISBN- és ISMN-számok ellenőrzése, jóváhagyása, javítása vagy törlése, az előzetes rekordok véglegesítése.
- Az ISBN-azonosító elemeinek önálló kezelése. Ezek a következők: előtag (978), országhód (963) vagy (615), kiadói kód vagy gyűjtőszám kód, kiadványelem, ellenőrző számjegy. Az új ISBN-azonosító elemeinek automatikus generálása vagy irányított létrehozása és kezelése.

1 717/2020. (XII. 30.) Korm. rendelet a kiadványok kötelezpéldányainak szolgáltatásáról, megőrzéséről és hasznosításáról: <https://njt.hu/jogszabaly/2020-717-20-22> (Hozzáférés: 2022.04.20.)



ISBN 978-963-200-100-5



3. ábra ISBN-szám vonalkóddal

- Az ISMN-azonosító elemeinek önálló kezelése. Ezek a következők: előtag 979-0, országkód *nincs*, kiadói kód vagy gyűjtőszám kód, kiadványelem, ellenőrző számjegy. Az új ISMN-azonosító elemeinek automatikus generálása vagy irányított létrehozása és kezelése.

ISMN 979-0-2600-0043-8



4. ábra ISMN-szám vonalkóddal

- **10 és 13 jegyű ISBN/ISMN-azonosítók párhuzamos kezelése.** Az azonosító számok nagymennyiségű kiutalása miatt 2007-ben a Nemzetközi ISBN Ügynökség, illetve 2008-ban a Nemzetközi ISMN Ügynökség is átállt a 13 számjegyű azonosító számokra. Az új rendszernek mind a 10, mind a 13 jegyű kiadványazonosító számokat kell tudnia kezelni.
- **Többkötetes műveknél az össz-ISBN/ISMN kezelése.** Többkötetes műveknél mindig egy ISBN-nel, illetve ISMN-nel többet utalunk ki a kiadványokra, mint ahány kötetből áll a mű. Ez a plusz azonosító az úgynevezett össz-ISBN, illetve össz-ISMN, melyet minden kötetben fel kell tüntetni.
- **Sorozati művek kezelése.** A sorozatok előre meg nem határozott időtartamra tervezett kiadványok, amelyek egymást követő részegységekből állnak. Lehetnek számozott vagy számozatlan sorozatok. A könyv- vagy kottasorozatok egésze ISSN-t, míg egyes kötetei ISBN-t, illetve ISMN-t kapnak. A sorozatokat meg kell különböztetni azoktól a többkötetes könyvektől és többkötetes kottáktól, melyek meghatározott számú kötet megjelenésével válnak teljessé, így ISSN-szám helyett össz-ISBN-számot kapnak.
- **Közös kiadások kezelése.** Ha több kiadó közösen jelentet meg egy könyvet, általában annak a kiadónak az ISBN-je szerepel a könyvön, amelyik kiadó a terjesztésre jogosult. Lehetőség van arra is, hogy minden résztvevő kiadó feltüntesse a saját ISBN-jét; a kiadványban, ebben az esetben jelölni kell azt is, hogy melyik szám melyik kiadóhoz tartozik.



## Az OKP ISSN-kezelő része felé támasztott alapkövetelmények

- ISSN-számigényléshez kapcsolódó adatbekérő felület kialakítása kiadványtípusok szerint. A bekért kiadványadatok alapján döntést kell hozni, hogy időszaki kiadványnak vagy folyamatosan frissülő integráló dokumentumnak minősíthető-e egy adott kiadvány. Ezért olyan adatbekérő felületet kellett kialakítani, amely minden olyan információt tartalmaz, amely alapján a Magyar ISSN Nemzeti Központ jól minősítheti a kiadásra szánt dokumentumot.
- ISSN azonosítóra nem jogosult kiadványok kiszűrése. Az időszaki kiadványok és folytatódó dokumentumok nem minden típusát látja el ISSN-azonosítóval a Magyar ISSN Nemzeti Központ, ezért a lehető legpontosabb adatbekérő űrlap elkészítésére volt szükség.
- Előzetes és utólagos ISSN-számkiutalás biztosítása. Az ISSN nyolc számjegyből álló numerikus kód, amelynek elemei – a könyveket azonosító ISBN-től eltérően – semmiféle jelentést nem hordoznak, csupán azonosító funkciót töltenek be. A nyolcadik számjegy ellenőrző számjegy, mely az előző hétből számítható ki, és 0-tól 10-ig terjedhet. A 10-es számjegyet római szám helyettesíti, azaz: X. Az ISSN feltüntetése két, egyenként négy számjegyet tartalmazó adatcsoportban történik, amelyeket kötőjel kapcsol össze, és az ISSN betűkód előz meg.



5. ábra ISSN-szám vonalkóddal

Az ISBN és ISMN számkiutalással ellentétben ISSN-azonosítót a már korábban megjelent kiadványra is lehet igényelni, illetve kiutalni. Ezért a fejlesztés során meg kellett oldani, hogy a modulban a már megjelent és a még megjelenés előtt álló kiadványra egyaránt lehessen ISSN számot igényelni/kiutalni.

- ISSN-számok kiutalásáról értesítés küldése. A kiadóktól, szerkesztőségektől beérkezett kiadványadatok, illetve a megküldött kiadványtervek alapján megszületik a döntés az ISSN-szám kiutalásáról vagy a kérés elutasításáról. Ez az információ a modul partneri felületén keresztül jut el az igénylőhöz.
- Későbbiekben a rendszer közvetlen összekapcsolása a nemzetközi ISSN-regiszterrel. A jelenlegi rendszer úgy lett kialakítva, hogy az ISSN-igényléshez szükséges adatok bekérése és a számigénylők értesítése nincs közvetlenül összekapcsolva a nemzetközi ISSN Regiszterrel, ahol az időszaki kiadványok és a folyamatosan frissülő integráló dokumentumok adatai rögzítésre kerülnek. Ennek oka, hogy az ISSN Regiszter jelenleg megújítás alatt áll.

## Az OKP kötelezpéldány-kezelő része felé támasztott alapkövetelmények

- Feleljen meg a hatályos kötelezpéldányról szóló kormányrendeletnek. 2021. január 1-én lépett hatályba a 717/2020. (XII.30.) Korm. rendelet a kiadványok kötelezpéldányainak szolgáltatásáról, megőrzéséről és használatáról. A hatályos rendelet a korábbi jogszabálynál részletesebben taglalja az elektronikus formában megjelent kiadványok kezelését, a kötelezpéldányok elosztását, az 50 példányszámnál kevesebb példányban készült kiadványok beküldését és a nyomtatott kiadvánnyal megegyező elektronikus változat beküldését. Az új rendszert az eddigi kötelezpéldány-kezelés mellett ezekre az új funkciókra is fel kellett készíteni.
- Kezelje a nyomtatott és elektronikus formában is megjelenő kiadványok kötelezpéldányainak érkeztetését és elosztását. A fejlesztők számára nagy kihívást jelentett a rendszer kialakítása során, hogy az azonos művek különböző megjelenési formáira (pl. nyomtatott kötött, nyomtatott fűzött, elektronikus e-pub, elektronikus mobi, hangoskönyv) kiutalt kiadványazonosító számokat gyakran felcserélik a kiadók és ezek kötelezpéldányai nem is feltétlenül egy időben érkeznek be a nemzeti könyvtárba.
- A kötelezpéldányok érkeztetésénél eddig nem használt új adatmezőket is tudjon kezelni (pl. oldalszám, kötés típusa, illusztráltság, várható megjelenés dátuma, évjelzés típusa, gerincméret, reklamáció jelzése, melléklet jelölése).
- Egy adott kiadvány többszöri beérkezését is tudja kezelni. Egy kiadványból utánnnyomások készülhetnek, amelyekből beérkeznek a kötelezpéldányok, ezeket újból és újból érkeztetnünk kell.
- Azonos kiadványazonosítóval, de különböző kiadásjelzéssel megjelenő kiadványok érkeztetése. Sajnos nem minden kiadó jár el szabályszerűen akkor, amikor nem igényel új kiadványazonosító számot a korábban megjelent kötetek újabb kiadásaira. Így előfordul, hogy azonos ISBN- vagy ISMN-számhoz több kiadás és így több kötelezpéldány beküldés is tartozik.
- Azonos kiadványazonosítóval megjelenő kiadásváltozatok (nyomtatott, elektronikus, kötött- fűzött stb.) érkeztetése. A különböző kiadásjelzéssel megjelenő kiadványokhoz hasonlóan néhány kiadó elfelejt a kiadásváltozatokra is külön azonosító számot igényelni és egy ISBN, ISMN, vagy ISSN számon jelenti meg a nyomtatott és elektronikus változatokat is.
- A számkiutaláskor megadott kiadó megváltozásának kezelése. Előfordul, hogy az előzetes számkiutaláskor más kiadót ad meg a számigénylő, mint ami végül a kiadványban szerepel. Ez leginkább az ISBN- és ISMN-számkiutalásnál jelent nagyobb problémát, hiszen egyes kiadók rendelkeznek saját számtartománnyal, amiből történik részükre a számkiutalás.
- Az előzetes kiadványazonosítóval rendelkező kiadványok, ISBN, illetve ISMN nélküli megjelenésének kezelése. A kötelezpéldány érkeztetés során számos esetben találkozunk olyan kiadvánnyal, melyhez igényelt a kiadó kiadványazonosító számot, de végül nem tüntette fel azt a megjelent kiadványban.
- Egy kiadványazonosító két különböző kiadványra történő felhasználásának kezelése. Előfordul, hogy tévedésből egy kiadványazonosító számot két különböző kiadványra is felhasznál egy kiadó, legtöbbször kiadói, vagy nyomdai adminisztrációs hiba miatt.
- A kiadványban és a vonalkódban feltüntetett eltérő azonosító kezelése. Gyakori hiba, hogy a kiadványban szereplő ISBN, ISMN vagy ISSN eltér a vonalkódban feltüntetett azonosítótól. Általában a vonalkódban szerepel a jó azonosító szám.



- Kevesebb vagy több példányban, utólag beérkezett kiadványok kezelése. Sajnos nagyon ritka az olyan nyomdai vagy kiadói szállítmány, ahol ne lenne probléma a beküldött példányszámokkal. Szerencsére partnereink mindig megküldik a hiányzó példányokat, de azok érkeztetése később történik, mint az alappéldányoké.
- Melléklet nélkül érkezett kiadványok kezelése. Előfordul, hogy a nyomdából érkező kiadványokból hiányoznak a mellékletek, melyeket a kiadóktól kell utólag bekérnünk.
- Az előzetestől eltérő című kiadványok érkeztetése. Naponta találkozunk olyan kötelezpéldányként beérkezett kiadványokkal, amelyekre egészen más előzetes címen kértek kiadványazonosító számot.

### Partnerek kezelése

Mind a kiadványazonosító számok kiutalása során, mind a kötelezpéldányok kezelése folyamán különböző külső partnerekkel kerül kapcsolatba a nemzeti könyvtár. Ezeket a partnereket három nagyobb csoportba sorolhatjuk:

- kiadók (testületi és magánkiadók)
- kötelezpéldány beszállítók (nyomdák, kiadók, szerzők, közreműködők, stb.)
- Az OSZK-nak beszállítandó kötelezpéldányokból részesülő könyvtárak (nemzeti könyvtárak, országos szakkönyvtárak, egyes egyetemi könyvtárak)

Jellemzően egy intézmény csak egy típusba sorolható, de vannak olyan partnereink, akik akár több szerepkörrel is rendelkeznek. Előfordul például, hogy egy könyvtár saját kiadásában jelentet meg műveket, amelyeknek kötelezpéldányait ő maga szállítja be a nemzeti könyvtárba, emellett a 717/2020. (XII.30.) Korm. rendelet alapján kötelezpéldányválogató-joggal rendelkező intézmény is. Ebben az esetben az adott könyvtárnál mindhárom szerepkör teljesül.

A partneri rendszer kialakítása során az alábbiakra kellett figyelniük:

- Testületi partnerek és magánszemélyek megkülönböztetése. Elsődlegesen a GDPR-előírások betartása miatt van erre szükség.
- Több szerepkörben is aktív partnerek definiálása.
- Partneri névváltozatok kezelése.
- Partneri átalakulások (pl. jogutód nélküli megszűnés, jogutódlás, összeolvadás kezelése)
- Gyűjtőszámos kiadói partnerek kezelése. A saját kiadói azonosítóval nem rendelkező alkalmi kiadók leválasztása oly módon, hogy az esetlegesen megnövekvő kiadványszám tükrében önálló, saját kiadói azonosítóval rendelkező kiadóvá válhassanak.

## Mindezek tükrében hogy áll az új rendszer?

Hosszas tesztelési időszakot követően jelenleg éles adatokkal dolgozunk mind a partneri nyilvántartás (6-7. ábra), mind az ISBN és ISMN azonosító kiutalás és ellenőrzés (8-9-10. ábra), mind a kötelepéldány-érkeztetés és szétoztás (11-15. ábra) során.

Partner azonosítója	Partner típusa	Partner neve	Partner székhely címe	Partner e-mail címe és telefonszáma	Státusz	Partner fő kapcsolattartó neve	Partner fő kapcsolattartó e-mail címe	Partner jogelődje/jogutódja
eddc8738-44ae-4e5f-b41b-c6cd706530be	kiadó (fő tevékenységi körben kiadással foglalkozó gazdasági társaság)	Országos Széchényi Könyvtár	1827 Budapest Budavári Palota F. ép.	+36 1/224-3700	Élő			

6. ábra Partnerek keresése/listázása

Partner típusa: kiadó (fő tevékenységi körben kiadással foglalkozó gazdasági társaság)

Partner típusa: Testületi

Partner neve: Országos Széchényi Könyvtár

Státusz: Élő

Létrehozás dátuma: 2022.02.26

Partnerhez tartozó funkciócsoportok: Kiadó, Szállító

Szállító azonosító: 2002623

Pályázati azonosító:

Tudományterület:

Névváltozat(ok): Névváltozat OSZK

Mégjegyzés:

7. ábra Partneradatok szerkesztése

Partner\* (20)

Számartomány típusa\* (ISBN)

Kiadás típusa\* (Egyszeri (számkiosztás 1 kiadványra))

Mű jellege kötetek számossága szerint\*

Megjelenési forma\*

8. ábra ISBN-számkiutalás. Alapadatok

OKP Dokumentumfogadó kapu

Új ISBN szám kiutalása - Schnitta Sámuel Magyar Éttermi Kultúráért Egyesület

ALAPADATOK SZÁMTARTOMÁNY **SZÁM KIOSZTÁS ALAPADATOK** DOKUMENTUM ADATOK

**Nyomatott szám kiosztás alapadatok**

Beérkezettég\*

Szám megadás módja

Dedikált számtartomány

Gyűjtő számtartomány

Számtartomány kiválasztása\* (0)

Szám kiosztása automatikusan

**Mellékletek**

Számigénylésre vonatkozó megjegyzés

9. ábra ISBN-számkiutalás. Számkiosztási adatok

OKP Dokumentumfogadó kapu

Új ISBN szám kiutalása - Schnitta Sámuel Magyar Éttermi Kultúráért Egyesület

ALAPADATOK SZÁMTARTOMÁNY SZÁM KIOSZTÁS ALAPADATOK **DOKUMENTUM ADATOK**

Dokumentum jellege\*

Sorozat-e?

Fordítás alapján készült?

Várható ISBN szám(ok)

Cím\*

Alcím

Szerző, vagy szerző hiányában egyéb közreműködő\*

Közreműködés jellege

Társszerző, egyéb közreműködő

Társkiadók

50 példány alatti megjelenés

10. ábra Számkiosztási adatok. Dokumentumadatok

OKP Dokumentumfogadó kapu

Nyomatott kötelepéldány érkeztetése

ISBN ISMN ISSN

Keresési kifejezés: 978-963-200-720-5

Keresési paraméter: Kiutalt ISBN/ISMN/ISSN szám

OSZLÓP PARAMÉTEREZÉS MÓDOSÍTÁSA

Partner/kiadó megnevezése	Kiutalt szám típusa	Kiutalt ISBN/ISMN/ISSN szám	Cím	Szerző	Dokumentumtípus	Nyomatott/e-dokumentum	Kötés típusa	Sorozat címe	Beadás dátuma	Lajstromszám	Beérkezettég	Beérkezés dátuma	Műveletek
Országos Széchényi Könyvtár	ISBN	978-963-200-720-5	Ribizli na kraji svitu	nincs adat	Nyomatott nyelvi anyag (pl. könyv)	nyomatott dokumentum	-	-	2022.03.10	-	Előzetes	-	

Sorok száma: 10 / 1-1 / 1

11. ábra Kötelepéldány érkeztetés. Kiadvány kiválasztása



OKP Dokumentumfogadó kapu

Nyomatott kötelesspéldány érkeztetése

Kötelesspéldány beadás - Országos Széchényi Könyvtár

ALAPADATOK   DOKUMENTUM METAADATOK   JOGKEZELÉSI ADATOK   FELDOLGOZÁSI ADATOK   SZÁLLÍTÁS ALAPADATAI   KAPCSOLÓDÓ PÉLDÁNYOK   REKLAMÁCIÓK

Beadó partner megnevezése\*  
Országos Széchényi Könyvtár (4)

Beadó partner kapcsolattartójának megnevezése

Beküldetés\*  
Előzetes

Szám típusa\*  
ISBN

Szám nélküli

Csím  
978-963-200-720-5

Számküldés dátuma  
2022. 03. 10.

Számgigijésre vonatkozó megjegyzés

Hibás számok:

KÖVETKEZŐ   MENTÉS   MENTÉS ÉS ELFOGADÁS

12. ábra Kötelesspéldány érkeztetés. Kiadvány alapadatai

Nyomatott kötelesspéldány érkeztetése

Kötelesspéldány beadás - Országos Széchényi Könyvtár

ALAPADATOK   **DOKUMENTUM METAADATOK**   JOGKEZELÉSI ADATOK   FELDOLGOZÁSI ADATOK   SZÁLLÍTÁS ALAPADATAI   KAPCSOLÓDÓ PÉLDÁNYOK   REKLAMÁCIÓK

A dokumentumhoz tartozó metaadatokat űrlapon keresztül szeretné megadni, vagy metaadat rekordot szeretne az e-dokumentumhoz feltölteni? Az elfogadott metaadat rekord formátumok: MARC21 leírási szabványt használó .mrc kiterjesztés, MARC21 leírási szabványt használó MARCXML fájl, qDC leírási szabványt használó XML fájl.

Űrlappal metaadat megadás vagy metaadat rekord feltöltés  
Űrlap kitöltés

Dokumentum jellege\*  
Nyomatott nyelvi anyag (pl. könyv)

Tudományterület

Sorozat-e?

Fordítás alapján készült?

Cím\*  
Ribizli na krau svitu

Alcím

Szerző, vagy szerző nényében egyéb közreműködő?  
nincs adat

Közreműködés jellege

VISSZA   KÖVETKEZŐ   MENTÉS   MENTÉS ÉS ELFOGADÁS

13. ábra Kötelesspéldány érkeztetés. Dokumentumok metaadatai.

The screenshot shows a web form for submitting a book sample. The 'SZÁLLÍTÁS ALAPADATAI' (Shipping Basic Data) tab is active. Fields include: 'Szállító partner megnevezése\*' (20 characters), 'Szállító partner kapcsolattartójának megnevezése' (100 characters), 'Szállító partner kapcsolattartó elérhetőség', 'Szállítólevél száma\*', 'Vonalkód', and 'Melléletek' (Attachments) with a '+3' button. A 'Beadom példányok száma\*' field is set to 6. There is a checkbox for 'Plusz példány beküldése OSZK digitalizálásra' and a checked checkbox for 'Automatikus érkeztetés'. At the bottom, there are buttons for 'VISSZA', 'KÖVETKEZŐ', 'MENTÉS', and 'MENTÉS ÉS ELFOGADÁS'.

14. ábra Kötelezpéldány-érkeztetés. Szállítás alapadatai

The screenshot shows the 'Elosztás' (Distribution) form. It includes a header with 'Cím:', 'Kiutalt szám:', and 'Benyújtott példányok száma Tudományterület'. Below this is a table of distribution items, each with a code, a description, and a quantity. The items are: 1/B1 (OSZK nyomtatott használati...), 1/B1 (OSZK Nyomatott archiv pél...), 1/B1 (OSZK ún. EV-s nyomtatott p...), 2/D1 (Debreceni Egyetem. Egyete...), 3/B9 (KSH Kvt), 6/SZ1 (Szegeődi Tudományegyetem...), 7/P1 (Pécsi Tudományegyetem Eg...), 10/B2 (ELTE Egyetemi Kvt. és Levél...), 11/B10 (Fővárosi Szabó Ervin Könyvt...), 12/B8 (Országgyűlési Kvt), 15/B12 (Budapesti Corvinus Egyete...), 16/B7 (Országos Pedag. Kvt. és Mű...), 17/B1700 (Kvt. Int. / Könyvtártörténeti...), 19/B11 (BME-OMIKK), and 22/B5 (Magyar Mezőgazdasági Mű...). At the bottom right, it says 'Maradék: 4' and has 'MENTÉS' and 'MENTÉS ÉS ELFOGADÁS' buttons.

15. ábra Kötelezpéldány-elosztás

### Milyen feladatok várnak még ránk?

- **Partneri felület indítása.** Reményeink szerint a partneri felület elsősorban a kiadók részére fog óriási segítséget nyújtani, hiszen egy felületen keresztül fognak tudni mind ISBN, ISMN és ISSN azonosítót igényelni és kiadói adataikat módosítani. Nyomon tudják a jövőben követni az összes részükre kiutalt nemzetközi azonosító számot és az azokhoz tartozó kiadványok adatait. Ellenőrizhetik, hogy saját kiadványaikból beérkeztek-e a kötelezpéldányok.
- **Pályázati igazolások kezelésnek tesztelése.** A modul úgy lett kialakítva, hogy a pályázati támogatás révén megjelent és beszolgáltatásra került kötelezpéldányokról igazolásokat állítson ki a pályázatkezelő részére.

- Az előforduló adatmigrációs hibák javítása. Az adatmigráció több rendszerből egy rendszerbe történt, ezért keletkeztek duplikált rekordok, amelyeket ki kell szűrni néhány egyéb adatmigrációs hiba javítása mellett.
- A katalogizáló és gyarapítási modulhoz kapcsolódás ellenőrzése. Amint az OKP katalogizáló és gyarapítási modulja is éles adatokkal kezd majd dolgozni, ellenőrizni kell a kiadványrekordok hibátlan áttöltését.

## Források

1. Országos Könyvtári Platform honlapja. <https://okp.oszk.hu/> (Hozzáférés: 2022.04.20.)
2. Országos Széchényi Könyvtár honlapja. ISBN, ISMN menüpont. Kapcsolódó tartalmak <http://www.oszk.hu/sites/default/files/ISBN-utmutato-%2020121220.pdf> (Hozzáférés: 2022.04.20.)
3. Országos Széchényi Könyvtár honlapja. ISBN, ISMN menüpont. Kapcsolódó tartalmak [http://www.oszk.hu/sites/default/files/ISMN\\_utmutato.pdf](http://www.oszk.hu/sites/default/files/ISMN_utmutato.pdf) (Hozzáférés: 2022.04.20.)
4. Országos Széchényi Könyvtár honlapja. ISSN menüpont. Kapcsolódó tartalmak <http://www.oszk.hu/sites/default/files/ISSN%20utmutato-2012.pdf> (Hozzáférés: 2022.04.20.)
5. Országos Széchényi Könyvtár honlapja. Kötelesepéldányok menüpont. Kapcsolódó tartalmak <http://www.oszk.hu/sites/default/files/magyar-kozlony-koteles-rendelet-2021.pdf> (Hozzáférés: 2022.04.20.)

## Digitálisan született kutatási anyagok megőrzése: a relációs adatbázis mint born-digital objektum

Alföldi István

Digitális Örökség Nemzeti Laboratórium  
[alfi@poliphon.hu](mailto:alfi@poliphon.hu)

Dióssy Anna Laura

Digitális Örökség Nemzeti Laboratórium  
[diossy.panka@btk.elte.hu](mailto:diossy.panka@btk.elte.hu)

Hungarian cultural heritage experiences a significant loss of valuable research data due to a lack of knowledge, infrastructure, and often even plans by academic institutions to preserve born-digital information. National Laboratory for Digital Heritage (DH-Lab) has launched a project, supporting the open academic research initiative, to archive and provide open access to born-digital research data. DH-Lab plans to achieve a set of digital archiving pilot scenarios starting with a database archiving pilot. This presentation provides a summary of the pilot activities and results.

In the scope of the pilot, three research databases created by the Institute for Literary Studies of Hungarian Academy of Sciences have been archived using the information package specifications and database archiving tools of the European E-ARK program. E-ARK aims to impact the development of internationally accessible archives through the provision of technical specifications and tools. E-ARK components comply with the Open Archival Information System (OAIS) reference model for digital archiving. All three databases are archived as SIARD files (a format designed specifically for database archiving), packaged in a standard E-ARK SIP package and ingested in the RODA OAIS repository. Three different image archiving strategies have been tested in order to find an optimal solution for storing images (or theoretically any other larger objects) accompanying the databases.

### 1. Bevezetés

A Digitális Örökség Nemzeti Laboratórium 2020. szeptemberében azzal a céllal jött létre, hogy intézményközi összefogásban kidolgozza a nemzeti kulturális örökség MI alapú feldolgozásának, kutatásának és oktatásának, valamint a lehető legszélesebb körű közzétételének módszertanát. A Digitális Örökség Nemzeti Laboratórium egyik alprojektje egy born-digital labor létrehozása és működtetése, mivel a magyar kulturális örökség nagy mennyiségű adatvesztést szenved el azáltal, hogy korlátozott az infrastruktúra és kevés a szakember. Nagy szükség van tehát jó-gyakorlatok kialakítására, hogy a különféle, adott esetben elavult hordozókon tárolt és/vagy elavult formátumú anyagok speciális szaktudással tudjuk kezelni.<sup>1</sup> A laboratórium kiemelt figyelmet szentel ennek a projektnek ugyanis az anyaország és a határon túli magyar közösség born-digital anyagainak archiválása és közzététele egy rendkívül értékes és ugyanakkor veszélyeztetett szelétét menti meg a magyar nemzeti kultúrának.

1 Patrik Svensson: *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. University of Michigan Press, 2016, 2. o. <https://doi.org/10.2307/j.ctv65sx0t.5>. (utolsó letöltés dátuma: 2022. június 16.)

Aborn-digital anyagok kezelése sok szempontból kihívást jelent, felmerülnek ugyanis technikai és etikai kérdések is, kezdve a hozzáférhetőség problematikájával, a megbízhatóságon és hitelességen, a megrongálódott adatok helyreállíthatóságán, egészen a módosítások nyomon követhetőségéig. Mindezekre nyújthatnak megoldást a digitális nyomrögzítési módszerek, melyek sok esetben szinte analóg módon alkalmazhatók a kulturális örökség anyagainak archiválásakor.<sup>2</sup> Ugyan a digitális kriminalisztika és a humántudományok egymástól távoli területeknek tűnhetnek, azonban a törvényszéki szakértők által kifejlesztett módszerek és eszközök hasznosnak bizonyulhatnak a közgyűjtemények és kutatóhelyek számára.<sup>3</sup> Nincs ugyanis nagy különbség például egy bűnöző számítógépének átvizsgálása vagy egy írói hagyaték részét képező számítógép dokumentumainak vizsgálata között. Mindössze néhány példát említve, mindkét esetben biztosítani kell a digitális tartalom integritását, elérhetővé kell tenni az eredeti állapot bit szintű képét, valamint nyomon követhetővé kell tenni az esetleges módosításokat.

A bölcsészettudományi kutatások során gyakran használt eszköz a relációs adatbázis, melyre born-digital anyagként kell tekintenünk. A Digitális Örökség Nemzeti Laboratórium egyik pilot projektje egy olyan módszertan kidolgozása, amely az ilyen adatbázisok hosszútávú megőrzését célozza. A pilot projekthez a Bölcsészettudományi Kutatócsoport Irodalomtudományi Intézetének három adatbázisát használtuk. Mindhárom esetben még nem lezárt kutatások adatbázisairól van szó, melyek a mai napig folyamatosan bővülnek. A Lendület Nyugat-magyarországi irodalom 1770–1820 Kutatócsoport Csörsz-Rumen István vezetésével 2017-ben azzal a céllal jött létre, hogy feltárja a nyugat-magyarországi irodalom 1770-1820 közötti történetét. Az első adatbázis a *Magyarországi populáris nyomtatványok (17–19. század)* nevet viseli, ez az első magyar kísérleti jellegű ponyva- és kalendárium-adatbázis, mely kora újkori és újkori magyarországi, nyomtatott füzetként, zsebkönyvként, esetenként röplapként terjedő irodalmi kiadványok adatait teszi hozzáférhetővé. A második adatbázis fő célja, hogy kutatási segédletet nyújtson a felvilágosodás kori magyar levélkorpuszok vizsgálatához, és egy helyen tegye áttekinthetővé a kevésbé feltárt, magyar vonatkozású, többnyelvű irodalmi és tudományos levelezések fő adatait. Ezen adatbázis az *Írói és tudóslevelezés* nevet viseli. A harmadik adatbázis, melynek címe *Regények Magyarországon (1730–1836)*, a napjainkban ismeretes és fellelhető, 1730 és 1836 között keletkezett regények adatait teszi hozzáférhetővé és kutathatóvá.

## 2. Az adatbázis archiválás kihívásai

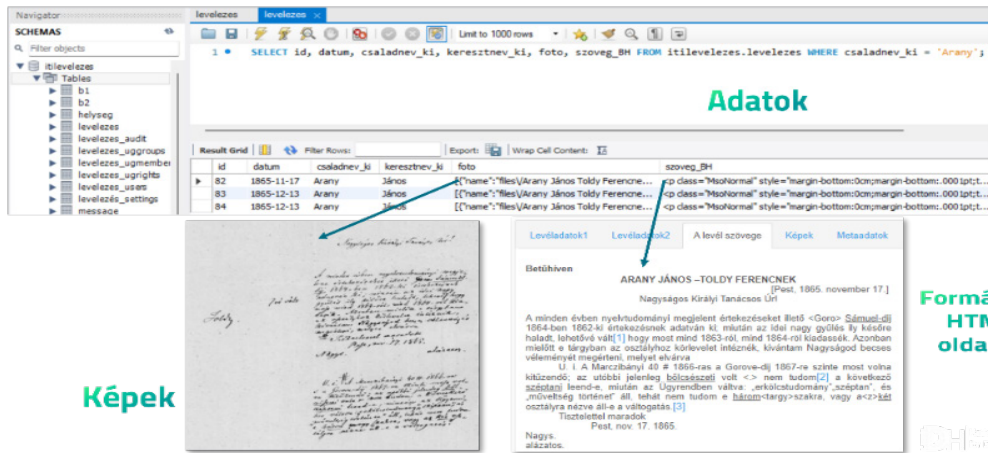
Az archiválás megkezdése előtt számos kérdést érdemes tisztázni. Például élő (aktív használatban lévő, bővülő, módosuló) vagy lezárt adatbázisról van szó? Milyen nem-triviális mezőtípusokat tartalmaz az adatbázis? Hogyan használták (használják) az adatokat? Milyen üzleti, közgyűjteményi, tudományos kontextusban értelmezhetők?

A korai adatbázisok lényegében egyszerű adatrekordokból álltak, amelyek csak az alap adattípusokat tartalmzták. A későbbiekben az adatbáziskezelők lehetővé tették az alap típusok mellett nagyobb objektumok (képek, szövegfájlok) tárolását is az adatbázison belül. Ez megnyitotta az utat a komplex adatobjektumok (HTML vagy JSON állományok, weblinkek, stb.) alkalmazása felé. Speciális adattípusok közé tartoznak továbbá a szöveggént tárolt külső hivatkozások (linkek). A pilothoz választott adatbázisok mindegyik itt felsorolt nem-triviális adattípust tartalmazzák.

---

2 Matthew G. Kirschenbaum és mások: *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington, DC, Council on Library and Information Resources, 2010, 1. sk. o. <https://www.clir.org/pubs/reports/pub149/>. (utolsó letöltés dátuma. 2022. június 16.)

3 Uo.

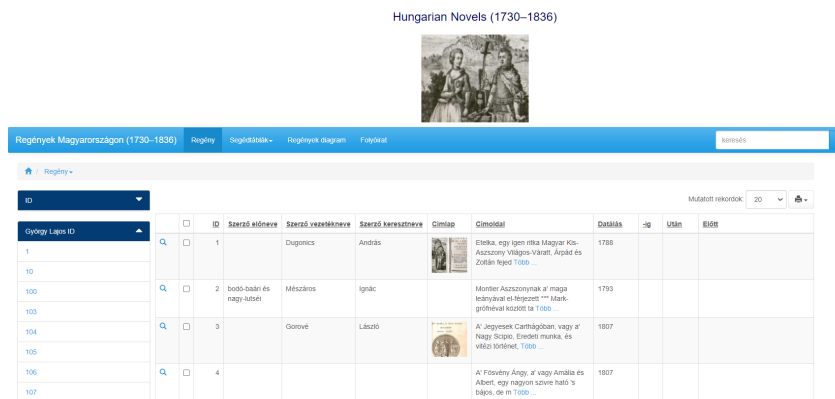


1. ábra. Képek és HTML mezők az adatbázisban

Az adatbázisok használati módja szintén kihívások elé állította a labort. Az Irodalomtudományi Intézettől kapott adatbázisok mind élő, ma is aktív használatban lévő adatbázisok. Az élő adatbázisok hosszútávú megőrzésének elvei és gyakorlata jelentősen eltérhet a már lezárt, nem módosuló adatbázisok archiválásától. (Pl. az archiválás során is folyamatosan biztosítani kell az elfogadható válaszidejű hozzáférést az adatokhoz, és nem egyszeri, hanem rendszeres mentési stratégiát kell kidolgozni.) Végül az a döntés született, hogy a pilot szempontjából lezártnak tekintjük a kapott adatbázisokat, és az egyszeri adatbázis exporttal készült pillanatfelvételeket tekintjük az archiválandó adatbázisoknak.

A kapott adatbázisokat a kutatók és az érdeklődők, az Irodalomtudományi Intézet honlapjáról elérhető, PHP nyelven megírt felületeken keresztül használják. Az egyedi felületek forráskódja, kinézete és működése nem archiválható a sztenderd adatbázis-archiválási eszközökkel.

### Regények Magyarországon (1730–1836)



2. ábra. Egyedi fejlesztésű PHP felület

Külön kihívást jelent, ha az adatbázisok nehezen interpretálható mezőket is tartalmaznak. Az archiválás célja nem pusztán a megőrzés, hanem a hozzáférés biztosítása az archivált információhoz. Azokat az információ darabokat, amelyeket kizárólag az eredeti adatbázishoz külön fejlesztett felülettel lehet interpretálni, valamilyen módon elérhetővé kell tenni az archív állomány felhasználói számára is.

A fentiekből jól látható, hogy az adatbázis-archiválás koránt sem automatikus tevékenység, így azt az adatbázisok alapos elemzése és a stratégiai kérdések tisztázása nélkül aligha lehet sikeresen végrehajtani.



### 3. Az adatbázis-archiválás eszközei

Az adatbázisok archiválására rendelkezésre álló eszközök közül jelen pilothoz az Európai Bizottság által finanszírozott E-ARK és eArchiving projektek eredményeként fejlesztett (vagy továbbfejlesztett) komponenseket használtuk:

- ✓ SIARD 2.1 formátum – adatok tárolása
- ✓ Database Preservation Toolkit – SIARD fájl előállítás és metaadat-hozzárendelés
- ✓ RODA-in – információs csomag (SIP) előállítására
- ✓ RODA Repository – hosszútávú megőrzést biztosító OAIS

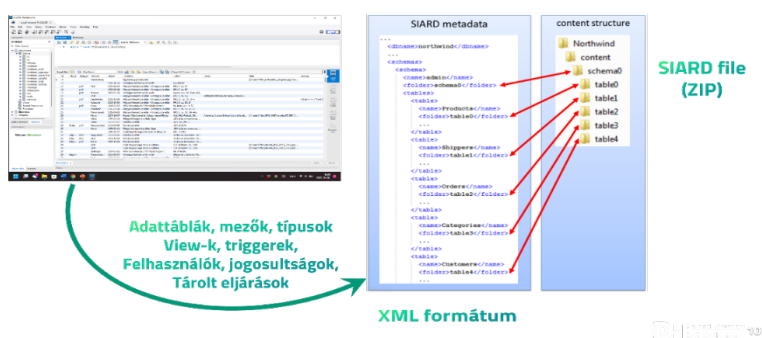
### E-ARK program

Az Európai Bizottság által finanszírozott E-ARK program célja, hogy az európai tapasztalatok és jó gyakorlatok alapján harmonizálja a digitális archiválási eszközöket és módszereket. Szigorúan követi az OAIS referenciamodell logikáját és követelményeit, de ezen túlmenően támogatást nyújt az OAIS modell által le nem fedett területeken is. Az OAIS definiálja a különböző archív csomagokat, mint a beadási vagy a disszeminációs csomag, azonban nem adja meg ezek struktúráját. Az E-ARK program részletes, alaposan átgondolt SIP, AIP, DIP specifikációkat nyújt, valamint számos eszközt fejlesztett az archív csomagok kezelésére, és az OAIS folyamatok kiszolgálására. A projekt során ezen kívül elkészültek az OAIS folyamatok nagyvonalú folyamatmodelljei mind BPMN 2.0, mind ArchiMate specifikációk szerint.<sup>4</sup>

### SIARD

Az adatbázis archiválás de facto sztenderdje a SIARD formátum. Az eredetileg a svájci szövetségi levéltár által fejlesztett, majd az E-ARK projektek által továbbfejlesztett SIARD formátum a 2.2 verzióán tart. A pilothoz a SIARD 2.1 specifikációnak megfelelő formátumot használtuk. Az adatbázisból exportált SIARD fájl az adatokat egy XML fájlban, a nagy objektumokat (LOBs) pedig opcionálisan egy meghatározott könyvtárstruktúrában tárolja. A SIARD formátum nem csak az adattáblák adatait, hanem lényegében minden fontos információt el tud tárolni.<sup>5</sup>

### Adatbázis archiválás – SIARD



3. ábra. SIARD XML és könyvtárstruktúra

4 <https://eakr.online/>

5 <https://dilcis.eu/content-types/siard>



## Database Preservation Toolkit

A SIARD fájl előállításához a Database Preservation Toolkit (DBPTK) nevű eszközt használtuk. A portugál KEEP Solutions cég által fejlesztett, és az E-ARK projektek keretében folyamatosan továbbfejlesztett DBPTK a legtöbb relációs adatbázisból képes kinyerni a szükséges információt, valamint előállítani a megfelelő SIARD fájlt.<sup>6</sup>

## RODA-in

A RODA Repository-hoz készült RODA-in program különböző formátumú szabványos beadási információs csomagokat (SIP) állít elő. Jelen projektben az E-ARK SIP formátumot használtuk. A pilotban létrehozott SIP csomagok megfelelnek az E-ARK program által az adatbázisok archiválására kialakított CITS-SIARD tartalomspecifikus csomag specifikációnak.<sup>7</sup>

## RODA Repository

A RODA egy OAIS szabványnak megfelelő open-source archívum, mely digitális állományok hosszútávú megőrzésére alkalmas. A RODA, mint a pilotnál használt összes eszköz, teljesen E-ARK kompatibilis, verziói követik az E-ARK specifikáció verziókat.<sup>8</sup>

## 4. Pilot tervezés

A pilot tervezése során felmértük a Lendület Nyugat-magyarországi irodalom 1770–1820 Kutatócsoporttól kapott három adatbázist, és kialakítottuk az archiválási stratégiát.

### Az adatbázisok felmérése

Az adatbázisok felmérése során össze kellett gyűjteni minden olyan problémát, amelyek az archiválási stratégia kialakítása előtt döntést igényelnek. A felmérés célja, hogy pontos képet kapjunk az adatbázisokban tárolt adatok mennyiségéről, a használt adattípusokról, ezen belül a nem-triviális adatokról (képek, szövegfájlok, kódolt állományok, linkek, stb.), a használat módjáról, a meglévő dokumentációról, stb.

### Archiválási stratégia

Az archiválási stratégia kialakításakor a következő kérdéseket kellett megválaszolnunk:

- Hogyan archiváljuk a PHP nyelvű, egyedi fejlesztésű felületeket?
- Hogyan archiváljuk a nehezen interpretálható adatokat (HTML oldalak, JSON állományok)?
- Hogyan archiváljuk a képeket?
- Milyen adatbázis módosításokat kell elvégezni és hogyan biztosítsuk az eredeti adatok sérthetetlenségét?

A kapott adatbázisokat a kutatók és az érdeklődők, az Irodalomtudományi Intézet honlapjáról elérhető, PHP nyelven megírt egyedi felületeken keresztül érik el. Sajnos az eredeti PHP

6 <https://database-preservation.com/>

7 <https://rodain.roda-community.org/>

8 <https://roda-community.org>

felület archiválására nincsen készen használható megoldás. Természetesen megtehetnénk, hogy elmentjük a PHP kódot (amennyiben ez rendelkezésre áll), vagy emuláljuk az egész PHP környezetet, de ez az adatbázis archiválásnál jóval nagyobb és komplexebb projekt lenne. Mindenképpen túlmutatna a jelenlegi pilot céljain és lehetőségein. Az általános gyakorlat az, hogy az adatbázisokban tárolt primer adatokat archiválják és a kontextust, használati módot, egyedileg fejlesztett kiegészítéseket legfeljebb dokumentációként tárolják. Mi is ezt a módszert választottuk.

Az ITI három adatbázisában többféle, módosítás nélkül nehezen interpretálható információ található. Egyes szöveges állományok formázott HTML kódként vannak tárolva. Ez praktikus a PHP megjelenítő számára, de nehezen értelmezhető a laikus kutatónak, aki csak az archív állományt látja, valamint szinte lehetlenné teszi a szöveges keresést a rekordokban. Ugyanez a helyzet a Jason állományként kódolt fájlhivatkozásokkal és linkekkel. Úgy döntöttünk, hogy (az eredeti mezők megtartása mellett) létrehozunk konvertált mezőket, amelyekbe egyszerű, kereshető szöveggé tároljuk az információt.

Ezek a konverziók értelemszerűen adatbázis-módosítással járnak. Annak érdekében, hogy a SIARD állomány egyben lássa az adatokat, elengedhetetlen, hogy ugyanabba az adatbázisba kerüljenek a konvertált mezők is. Ez ezért is fontos, mert a Database Preservation Toolkit számos keresőfunkciót biztosít a SIARD fájlban anélkül, hogy a vissza kelljen állítani az adatbázist a SIARD állományból. Így viszont fokozottan kell ügyelni arra, hogy az eredeti adatok jól láthatóan elkülönüljenek az archiválás során előállított adatoktól. (Azt is megtehetjük, hogy az eredeti adatbázisról is készítünk SIARD mentést, és azt is eltesszük a SIP csomagban referenciának.)

**Új táblák (archiv\_...\_2022)**

- Az eredeti adatbázis táblák nem változnak
- Külön tábla a fotóknak, html oldalaknak és linkeknek

id	ponyvakataszter_id	img_order	img_name	img
694	54	0	files/IMGP9658_#_b6f9qc3.jpg	BLOB
695	54	1	files/IMGP8863_Kio_1vzkivuu.jpg	BLOB
696	81	0	files/B1850_2.02_rvqdhqio.jpg	BLOB
697	81	1	files/B1850_2.04_8ts4gjl.jpg	BLOB
698	81	2	files/B1850_2.03_4zyzcju.jpg	BLOB
699	224	0	files/IMGP9221_#_4ug1h0m.jpg	BLOB

id	ponyvakataszter_id	nyomatvany_leirasa_text	nyomatvany_leirasa_html	kapcsolat_text
7	245	Az ismeretlen nyomdahelelyű füzet Nr. 13 számot...		
8	246	Arany János autográf levele (Szemere Bertalan...		
9	247	Az OSZK 821.132-821.158 jelzetű, jórészt 181...		Arany a sorozat 2. füzetéből is mellékelt
10	248	A British Library (11585.aa.52 jelzetű) példány...		A Barna Péter-ponyvák repertoriája nem

4. ábra. Adatbázis módosítások

A módosított adatok számára tehát külön adattáblákat hoztunk létre, amelyek neve az archiv\_ előtaggal kezdődik. Minden módosítástípushoz külön tábla jött létre, hogy jól követhető legyen, hogy az értelmezhetőség és kereshetőség érdekében milyen módosított mezőket hoztunk létre.

Az adatbázismezőkhöz tartozó képek archiválására a SIARD 2.1-es verziója már többféle megoldást ajánl, attól függően, hogy az eredeti képek az adatbázisban, vagy hivatkozásként az adatbázis mellett vannak tárolva. Amennyiben a képek eredetileg az adatbázisban, BLOB mezőben vannak tárolva, elmenthetjük a képeket magában az adatbázisban. Ilyenkor egyetlen (de gyakran nagy méretű) SIARD fájl készül. A második lehetőség, hogy az eredetileg az adatbázisban tárolt képeket a SIARD fájlban kívül archiváljuk. Ekkor a DBPTK program létrehoz egy szabványos könyvtárstruktúrát, amelyben a képfájlok szabványos nevekké szerepelnek, a SIARD fájlba pedig csak a megfelelő hivatkozások kerülnek. Ilyenkor a SIARD

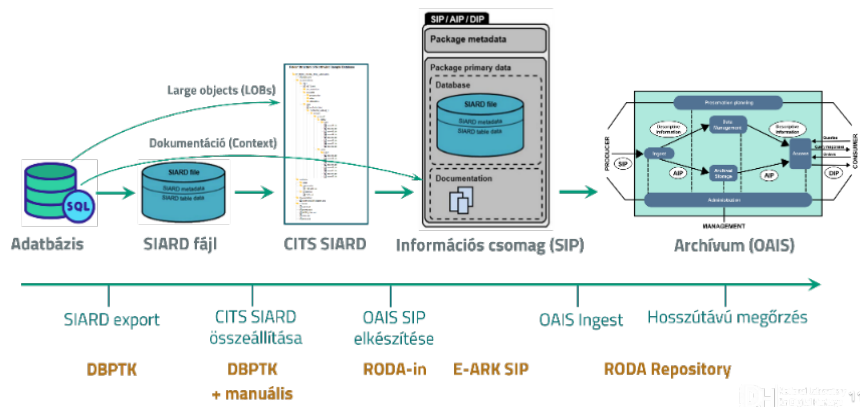
fájl sok kép esetén is kezelhető marad, de figyelni kell rá, hogy a képek a SIARD fájlal mindig együtt mozogjanak, és ne változzanak a könyvtár és fájl elnevezések. A harmadik megoldás szerint az eredetileg is külön tárolt képeket a SIARD mellett archiváljuk. Ez szabványos Data Link mezők esetén automatikusan történik, de az adott MySQL verzió sajnos nem támogatja ezt a mezőtípust, így a linkeket nekünk kellett létrehozni.

		Adatok	Képek	HTML oldalak
Populáris nyomtatványok		SIARD	Adatbázisban LOB mezőben SIARD-on belül	Adatbázisban LOB mezőben SIARD-on belül
Írói levelezés		SIARD	Adatbázisban SIARD-on kívül	Adatbázisban SIARD-on kívül
Regény		SIARD	Adatbázison kívül SIARD-on kívül	Adatbázison kívül SIARD-on kívül



5. ábra. Alkalmazott adatbázis archiválási stratégiák

## 5. Pilot megvalósítás



6. ábra. Az archiválási folyamat

7. ábra. Képek és HTML mezők a SIARD fájlban

A megvalósítás folyamatát a fenti ábra szemlélteti. Az első lépések még az adatbázisban történtek.

Adatbázis módosítások:

- ✓ A kapott adatbázis exportokat betöltöttük egy MsSQL adatbázisba
- ✓ Az 1. és 2. pilotok esetében „archiv\_” táblákat hoztunk létre a képek számára és BLOB mezőkbe töltöttük a képeket. Adatbázisonként több ilyen tábla is létrejöhetett, amennyiben a rekordok több szintjéhez is tartoztak képek.
- ✓ A HTML mezők és Jason linkek számára is létrehoztunk „archiv\_” táblákat. Az eredeti kódolt mezőket text formátumra konvertáltuk.

A SIARD fájlok létrehozása:

- ✓ A Database Preservation Toolkit programmal be kell jelentkezni az alkalmazásba. Ezért fontos, hogy teljes, legalább read-only hozzáférésünk legyen az adatbázishoz.
- ✓ Megadtuk a szükséges metaadatokat.
- ✓ Itt lehet beállítani az ellenőrző hash típusát, és itt kell megadni, hogy a nagy objektumokat (LOBs) a SIARD-on kívülre vagy belülré tegye. Az 1. pilot esetében belülré a 2., 3. pilotoknál kívülre tettük a LOB-okat.
- ✓ A generálás futtatása után előálltak a SIARD fájlok és a megfelelő LOB könyvtárstruktúra.

SIP csomag készítése:

- ✓ A SIP csomagot a RODA-in eszközzel készítettük.
- ✓ A SIARD fájlt és – amennyiben tartozott hozzá – a könyvtárstruktúrát tettük be elsődleges tartalomként.
- ✓ A dokumentáció könyvtárba kerültek az adatbázisok eredeti leírásai és a PHP felületről készült képernyőképek.
- ✓ SIP formátumnak az E-ARK2-t választottuk, ez az E-ARK Common IP Specification második generációjába tartozó SIP specifikációnak megfelelő beadási csomagot jelent.
- ✓ A metaadatokat az E-ARK CITS SIARD tartalomspecifikus specifikáció szerint adtuk meg.
- ✓ Végül legeneráltuk a SIP-et.

OAIS Ingest:

- ✓ OAIS komponensnek a RODA Repository-t választottuk. A RODA nem csak az OAIS szabvánnyal de az E-ARK program eredményeivel is kompatibilis, így az OAIS ingest (befogadás) folyamatban az E-ARK SIP csomag is minden további nélkül feltölthető és befogadható volt.
- ✓ Az adatbázisokat tartalmazó csomagok megjelentek a RODA Repository katalógusában, és az access (hozzáférés) OAIS folyamattal E-ARK DIP csomaggént letölthetővé váltak



## 6. Hivatkozások

### SIARD 2.1 specifikáció

(<https://github.com/DILCISBoard/SIARD/tree/master/SIARD%202.1.1/format/2019-05-15>)

### E-ARK CITS-SIARD specifikáció

([https://citssiard.dilcis.eu/specification/CITS\\_SIARD\\_version1\\_0\\_0.pdf](https://citssiard.dilcis.eu/specification/CITS_SIARD_version1_0_0.pdf))

### E-ARK SIP specifikáció

([https://earksip.dilcis.eu/archive/v2\\_0/eark-sip-v2-0-4.pdf](https://earksip.dilcis.eu/archive/v2_0/eark-sip-v2-0-4.pdf))

International case studies by the Relational DataBase Archiving Interest Group of the DILCIS Board:

- Case Study 1 ([https://dilcis.eu/images/2020review/9\\_Draft\\_SIARD\\_Case\\_Study\\_1.pdf](https://dilcis.eu/images/2020review/9_Draft_SIARD_Case_Study_1.pdf))
- Case Study 2 ([https://dilcis.eu/images/2020review/10\\_Draft\\_SIARD\\_Case\\_Study\\_2.pdf](https://dilcis.eu/images/2020review/10_Draft_SIARD_Case_Study_2.pdf))

### Szakirodalom

- Matthew G. Kirschenbaum és mások: *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington, DC, Council on Library and Information Resources, 2010, <https://www.clir.org/pubs/reports/pub149/>. (utolsó letöltés dátuma. 2022. június 16.)
- Patrik Svensson: *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. University of Michigan Press, 2016, <https://doi.org/10.2307/j.ctv65sx0t.5>.



## HTR-modellépítés és kézírásfelismerés nagyméretű, többszerzős szövegkorpuszon A Transkribus alkalmazása az Arany János hivatali iratokon<sup>1</sup>

Fekete Norbert  
Miskolci Egyetem-DigiPhil  
[feketenorbert1986@gmail.com](mailto:feketenorbert1986@gmail.com)

Creation of HTR models and handwritten text recognition on a large, multi-author text corpus

Using Transkribus on documents from the office of János Arany

In our study, we show how Transkribus has been used on a multi-author text corpus to create a HTR model that can process documents from the office of Arany efficiently. The test results so far are encouraging, and allow the features of model creation to be applied to other archives. To do this, it is worth creating HTR models based on one or two dominant handwritings first, and then other handwritings can be added to increase efficiency.

**keywords:** handwritten text recognition (HTR), Transkribus, János Arany

**kulcsszavak:** kézírásfelismerés (HTR), Transkribus, Arany János

Tanulmányunkban azt vizsgáljuk, hogy a Transkribus kézírásfelismerő (HTR) program hogyan használható egy több ember kézírását tartalmazó, eltérő dokumentumtípusokból álló szövegkorpuszon. A vizsgálathoz szükséges korpuszt az Arany János akadémiai főtőkárságának ideje alatt (1865–1877) keletkezett hivatali iratok szolgáltatták. Az Arany műveiből készülő kritikai kiadás új sorozatának előkészítő munkálatai során eddig feltérképezett korpusz kb. 9200 dokumentumot, azaz mintegy 30000 képfájlt jelent. Ennek a nagyméretű irategyüttesnek a kézírásról szöveggé történő átírása megoldhatónak tűnik a Transkribus segítségével. Ehhez első lépésként egy olyan HTR-modellt szükséges létrehozni, amely képes kellő hatékonysággal megküzdeni a korpuszt létrehozó számos kézírással, ugyanis a vizsgált anyag nemcsak az Arany által írt dokumentumokat, hanem más, az MTA-val kapcsolatba kerülő szerzők iratait is tartalmazza. A kutatócsoport tagjai az eddigi digitális filológiai kutatások során Arany kézírásával kapcsolatban két modellt dolgoztak ki.<sup>2</sup> Alkalmazkodva a korpusz sajátosságaihoz egy általánosabb harmadik modell létrehozására törekedtünk, amelynek eredményeit a következőkben szeretnénk bemutatni.

---

<sup>1</sup> A tanulmány a Digitális Örökség Nemzeti Laboratórium támogatásával valósult meg.

<sup>2</sup> Bobák Barbara, Gábori Kovács József, „Kézírásfelismerés Arany János levelein”, In: *Networkshop 2019*, szerkesztette Tick József, Kokas Károly, Holl András, 38-44, Budapest: Hungarnet Egyesület, 2019.



## A hivatali iratok és jelentőségük

Arany 1859-től az MTA rendes tagja volt, majd 1865-ben főtitkárrá választották, amely pozíciót egészen 1877-ig töltötte be. Arany tizenkét éven keresztül gyakorlatilag az Akadémia harmadik, az elnöki és az alelnöki pozíciót követően legfontosabb tisztségét birtokolta. Az általa végzett hivatali munka jelentős, mivel főtitkársága alatt kialakította és hosszú évtizedekre meghatározta az MTA, mint a hazai tudományosság egyik legfontosabb intézményének működési rendjét. Munkája során számos napi, a tudományszervezéshez kapcsolódó adminisztratív teendőt látott el: bejegyezte az üléseken tartott előadásokat, szerkesztette a társaság folyóiratait, intézte a kiadványok ügyeit, felbontotta és megválaszolta a postát, elkészítette az MTA éves beszámolóit, vezette az ülések jegyzőkönyveit, illetve intézte a pályázatok ügyeit. Akadémiai munkáját a kezdetektől egy öt fős titkárság segítette.<sup>3</sup> Az Arany-érában keletkezett iratok vizsgálata és digitális kritikai kiadásának létrehozása nemcsak a költő főtitkári pályájának mélyebb megértéséhez vezet el, hanem rámutat az MTA történetére, illetve az 1860-as és 1870-es évek tudományos, kulturális és művészeti életére is. Egyrészt a korpusz közzététele azért jelentős, mert számos kutatás tudja majd hasznosítani a jövőben; másrészt a kiadás egy lényeges, megoldandó digitális filológiai probléma, amely a hazai archívumok feldolgozására kínálhat lehetőséget.

## A feldolgozás mérföldkövei a Transkribus előtt

A hivatali iratok feldolgozása és közzététele az Arany János összes művei kritikai kiadássorozat 14. kötetében történt meg először. A sajtó alá rendező Gergely Pál elsődleges célja az volt, hogy a költő hivatali működését leíró legfontosabb iratokat és dokumentumtípusokat bemutassa. Az általa feltárt mintegy 2500 dokumentumból mindössze 717-et tett közzé, melyek között természetesen nem csak Arany kezétől származó iratokat találunk.<sup>4</sup> A kritikai kiadás új sorozatának munkálataihoz kapcsolódva 2014-ben folytatódott az iratokat feldolgozó munka. Ennek keretében feltérképezték és digitalizálták a Magyar Tudományos Akadémia Könyvtár és Információs Központ Kézirattár Régi Akadémiai Levéltárában (MTA KIK, Kt, RAL) található gyűjteményt. Ez összesen kb. 9200 dokumentumot, vagyis kb. 30000 fényképfelvételt jelent. Ebből a korpuszból kb. 2000 dokumentum szövegének átírását végezte el Gábori Kovács József, Szlanyiszló Lilla, Antal Alexandra és Fekete Norbert. A kutatók a szövegeket kézzel, wordben, illetve a Hegedüs Béla által létrehozott LyX kritikai kiadás készítésére alkalmassá tett programban végezték. A kiadás koncepciója többször változott: a jelenlegi cél, hogy a főtitkár kezén átment teljes anyag kiadása megvalósuljon, mely lehetővé teszi a teljes főtitkári munka feltárását.<sup>5</sup> A RAL feltárt állománya sem tekinthető azonban teljesnek, melyet a digitális kiadás előkészületeihez tartozó munkálatok során magunk is tapasztaltunk: Arany autográf kézírásával készült számos olyan további dokumentum található, melyek a költő hivatali tevékenységéhez köthetők. Ezek alapján látható, hogy a korpusz a digitalizált állomány feldolgozását követően tovább fog bővülni.

3 Gábori Kovács József, „Arany János, a hivatalnok”, In: *Önarckép álarckokban: Kiállításkatalógus*, szerkesztette Vaderna Gábor, Kalla Zsuzsa, Kaszap-Asztalos Emese, Sidó Anna, 350–366, Budapest: Petőfi Irodalmi Múzeum, 2018, 357–361.

4 Arany János, *Hivatali iratok 2.: Akadémiai évek (1859–77)*, szerkesztette, a jegyzeteket készítette Gergely Pál, Budapest: Akadémiai Könyvkiadó, 1964. (Arany János összes művei, 14.)

5 Gábori Kovács József, „Arany János hivatali levelei az Akadémián: A kritikai (újra)kiadás szükségessége és lehetőségei”, In: „... és palota épül a pusztá beszédből”: Akadémiai tudományos ülésszakok a 200 éves Arany Jánosról, szerkesztette Gábori Kovács József, Major Ágnes, 212–221, Budapest: reciti Könyvkiadó, 2017, 217–218; Bobák Barbara, Gábori Kovács József, „Kézírásfelismerés...”, 5.

## A feldolgozás lehetőségei és eredményei Transkribusban

A hivatali iratok feldolgozásához az Innsbrucki Egyetem által 2016-ban kifejlesztett Transkribus programot használtuk fel, amelynek nagy előnye, hogy nyílt forráskódú, felhasználóbarát felülettel bír, és egyszerre rendelkezik OCR és HTR modullal, így a nyomtatott és a kéziratos források feldolgozására is alkalmas. Ennek köszönhetően nagy mértékben csökkenthető vele a digitális szövegkiadások elkészítésének ideje, hiszen lerövidíti a szövegek átírására fordított időt. Hátránya azonban, hogy nem ingyenes.<sup>6</sup> Az eredményes kézírásfelismeréshez szükséges egy tanító modell összeállítása, amelynek létrehozásához egy legalább 15000 szóból álló korpusznak kell rendelkezésre állnia. A DigiPhil csoport és az ELKH Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetének közös kutatása során eddig két Arany-modell készült, melyeket olyan szövegekből hoztak létre, amelyek csak Arany kézírását tartalmazzák. A két modell között lényeges eltérés figyelhető meg: egyrészt bővült a modellekhez kiválogatott és átírt szövegek mennyisége, illetve a második modell hatékonysága jelentősen javult az elsőhöz képest. A Digitális Örökség Nemzeti Laboratórium, az ELTE, az ELKH BTK Irodalomtudományi Intézet és a Miskolci Egyetem kutatóiból álló kutatócsoport a hivatali iratok feldolgozásához egy új, harmadik modell létrehozását tűzte ki célul, amely már nemcsak Aranytól, hanem más szerzőktől is tartalmaz szövegeket. Távlati cél az, hogy az új modell képes legyen bármilyen 19. századi magyar nyelvű kézírás kiolvasására a lehető legkisebb hibaarányal.

A harmadik modell kialakításához kapcsolódóan több megfontolást is figyelembe kellett venni. Először át kellett tekintenünk a korpusz egy részét: erre azért volt szükség, mivel nem állt rendelkezésünkre olyan átfogó összeírás, amely tartalmazná a hivatali iratok összességét, ezért ezt párhuzamosan kellett elvégeznünk. A korpusz részleges áttekintése lehetővé tette, hogy megalapozzuk a harmadik modellhez tartozó tanítóanyag létrehozását. A források közül döntően magyar nyelvű, lehetőleg teljes dokumentumokat tartalmazó képeket válogattunk ki. Fontos szempont volt, hogy a képeken ne csak pár sornyi szöveg legyen, és az egyszerűség kedvéért kerültük a táblázatos források kiválasztását. A korpusz egyrészt az 1865-ben keletkezett és az első két modellhez használt képeket, valamint 1868-as és 1869-es dokumentumokat tartalmazza. A harmadik modell létrehozásához összesen 974 képfájl használtunk fel, melynek megoszlását az 1. táblázat mutatja.

Évkör	Korpusz	Mennyiség
1865	1. modellhez	100
1865	2. modellhez	200
1865	Ring Adorján	101
1868-1869	3. modellhez	573

1. táblázat A modellépítéshez használt Arany korpuszok

Az 1865-ös anyagból látható, hogy az egyik alkorpusz Ring Adorjának, Arany egyik írnokának a szövegeit tartalmazza, aki pozíciójából kifolyólag nagy számú hivatali szöveget hozott létre. Ebből következően a hozzá kötődő korpusz jelentősen tudja javítani egy többszerzős modell létrehozását.

Az új Arany-modell rendkívül heterogén, ugyanis kb. 30 kéztől tartalmaz különböző iratokat. A korpusz 75%-át teszik ki Arany és a Főtitkári Hivatal dolgozóinak iratai, míg 15%-át adja az MTA többi meghatározó személyisége (elnök, osztályelnökök stb.), akik pozíciójukból kifolyólag gyakran leveleztek a főtitkárral. Emellett mintegy 10%-a származik azoktól a levélíróktól, akik csupán néhány levélváltás erejéig szerepelnek a korpuszban.

<sup>6</sup> Transkribus <https://transkribus.eu/Transkribus/> (hozzáférés: 2022. 06. 07.)

A modellkészítés során az anyagot a második modell segítségével ismertettük fel, majd ezt kézzel javítottuk. A harmadik modellhez kapcsolódó átírások javítását összesen tizenegy fő végezte, míg a második modellhez készült szövegeket korábban Gábori Kovács József ellenőrizte.<sup>7</sup> A harmadik modellhez kapcsolódó textusok javítására fordított idő függött az írás terjedelmétől, az íráskép olvashatóságától. Arany kézírását a Transkribus a második modell alapján jó hibahatárral ismerte fel, így az ő szövegeinek javítása más szerzőkével szemben problémamentes volt. A legrövidebb javításra fordított idő kb. 5 perc, míg a leghosszabb akár 60–70 percet is igénybe vehetett. Átlagosan 15–20 percet fordítottunk egy átírat átnézésére.

A harmadik modell eredményei jócskán felülmúlták az előzőekét, amely leginkább a Character Error Rate (CER on Validation set) értéken mérhető le. Ez alapján megállapítható, hogy a program milyen hibaszázalékkal képes elvégezni a kézírásfelismerést, vagyis képenként a felismert karakterek hány százalékát rontja el. A korábbi eredményekhez képest (lásd 2. táblázat) a CER Validation értéke csökkent, miközben a korpuszt alkotó szószám jelentősen növekedett. A CER Validation értéke az első modell esetében 10,42% volt, míg a másodikonál 8,08%-ra csökkent, viszont a harmadik esetében 9,3%-ra növekedett. Ez utóbbi azonban jó aránynak tekinthető, főleg egy ennyire heterogén korpusz esetében.

	CER Train	CER Validation	Szavak száma
1. modell	5,73%	10,42%	17636
2. modell	2,71%	8,08%	29498
3. modell	5,30 %	9,30%	124070

2. táblázat Az Arany-modellek CER mutatói

A harmadik modell eredményeit a Bobák Barbara és Gábori Kovács József által alkalmazott vizsgálat módszerét figyelembe véve is felmértük.<sup>8</sup> Ennek alapja az ún. token, vagyis a szóköztől szóközиг terjedő karaktersorban ejtett hibák száma, melyhez kapcsolódva a szerzőpáros felmérte a karakterekben ejtett hibákat is, vagyis azt, hogy hány karaktert tévesztett a program az átírás elkészítésekor. A harmadik modell eredményeit a 3. táblázatban foglaltuk össze, melyben szerepeltettük Bobák és Gábori Kovács (BB-GKJ) adatait is. A teszthez két darab, egy-egy képre kiterjedő dokumentumot jelöltünk ki, melyeken csak egy kéztől származó kézírás szerepelt. Az egyik kép Arany Jánosé (AJ) és két levél piszkozatát tartalmazza, míg a másik képen található egyetlen levél Szabenyi Ferencről (SzF), a pesti Királyi Magyar Egyetemi Nyomda gondnokától származik.<sup>9</sup> Mind a két választott képen lefuttattuk mindhárom modellt, melynek eredményeképpen megállapítható, hogy az első két modell nagyobb sikerrel ismerte fel a karaktereket és kevesebb tokenben hibázott az Aranytól származó szöveg esetében. Ez többek között annak is köszönhető, hogy ezek Arany szövegeit felhasználva készültek. A választott Arany-leveleken végzett tesztelés közel hasonló eredményeket hozott mint a Bobák és Gábori Kovács által végzett teszteké. A harmadik, heterogén modell azonban sikeresen ismerte fel mind a két képen található szöveget, és már csak minimális hibák mutatkoztak a tokenek és a karakterek szintjén egyaránt. A tesztek más példák esetében eltérő számokat mutathatnak, de az arányok jól szemléltetik, hogy a nagyobb korpuszon alapuló harmadik modell közel azonos hatékonysággal képes felismerni

7 A harmadik modellhez kapcsolódó szövegek átírásában közreműködtek: Bajzát Tímea Borbála, Bobák Barbara, Fazekas Júlia, Fekete Norbert, Fellegi Zsófia, Hernády Judit, Horváth Anna, Jakab Éva, Szolnoki Anna, Vámos Violetta, Vétek Bence.

8 Gábori Kovács, Bobák, „Kézírásfelismerés...” 6.

9 Arany János Radnótfáy Sámuelnek, Pest, 1868. jan. 8., MTA KIK, Kt, RAL, 10/868; Arany János Karvassy Ágostonnak, Pest, 1868. jan., MTA KIK, Kt, RAL, 10/868; Szabenyi Ferenc a m. kir. vallási és közoktatási ministeriumnak, Buda, 1869. máj. 18., MTA KIK, Kt, RAL, 1869/1246.

az Aranyétól eltérő kézírásokat is, így a modell gyakorlatilag alapjául szolgálhat a 19. századi kézírásokat jó hatásfokkal felismerni képes modellhez.

	Token BB-GKJ	Karakter BB-GKJ	Token AJ	Karakter AJ	Token SzF	Karakter SzF
1. modell	128/200	276/954	69/133	121/814	109/127	436/793
	64%	29%	52%	15%	86%	55%
2. modell	48/200	116/954	34/133	51/814	79/127	102/793
	24%	12%	26%	6%	62%	13%
3. modell	-	-	7/133	9/814	7/127	9/793
	-	-	5%	1%	6%	1%

3. táblázat Az Arany-modellek teszteredményei

## Összegzés

Tanulmányunkban azt mutattuk be, hogy miként sikerült egy egyre bővülő és több szerzői kéztől származó szövegtörzset betanítani a Transkribus és létrehozni egy olyan HTR-modellt, amely hatékonyan tudja kezelni a hivatali iratokat. Az eddigi teszteredmények bizakodásra adnak okot, és valószínűleg lehetővé teszik, hogy a modellalkotás sajátosságait más hasonló archívumok esetében is alkalmazni lehessen. Ehhez először is érdemes olyan modelleket létrehozni, amelyek egy-két meghatározó kézen alapulnak, majd újabb kezek hozzáadásával a hatékonyság tovább növelhető.

## Bibliográfia

Arany János Radnótfáy Sámuelnek, Pest, 1868. jan. 8., MTA KIK, Kt, RAL, 10/868.

Arany János Karvassy Ágostonnak, Pest, 1868. jan., MTA KIK, Kt, RAL, 10/868.

Arany János, Hivatali iratok 2.: Akadémiai évek (1859–77). szerkesztette, a jegyzeteket készítette Gergely Pál, Budapest: Akadémiai Könyvkiadó, 1964. (Arany János összes művei, 14.)

Bobák Barbara, Gábori Kovács József, „Kézírásfelismerés Arany János levelein”, In: *Networkshop 2019*, szerkesztette Tick József, Kokas Károly, Holl András, 38-44, Budapest: Hungarnet Egyesület, 2019. <https://doi.org/10.31915/NWS.2019.5>

Gábori Kovács József, „Arany János, a hivatalnok”, In: *Önarckép álarckokban: Kiállításkatalógus*, szerkesztette Vaderna Gábor, Kalla Zsuzsa, Kaszap-Asztalos Emese, Sidó Anna, 350–366, Budapest: Petőfi Irodalmi Múzeum, 2018

Gábori Kovács József, „Arany János hivatali levelei az Akadémián: A kritikai (újra)kiadás szükségessége és lehetőségei, In „... és palota épül a pusztából”: Akadémiai tudományos ülésszakok a 200 éves Arany Jánosról, szerkesztette Gábori Kovács József, Major Ágnes, 212–221, Budapest: reciti Könyvkiadó, 2017.

Szebenyi Ferenc a m. kir. vallási és közoktatási ministeriumnak, Buda, 1869. máj. 18., MTA KIK, KT, RAL, 1869/1246.

Transkribus <https://transkribus.eu/Transkribus/> (hozzáférés: 2022. 06. 07.)

## ELTE Népdalkorpusz – magyar népdalok gépileg annotált adatbázisa

Horváth Péter

*Eötvös Loránd Tudományegyetem, Digitális Örökség Nemzeti Laboratórium*  
[horvath.peeteer@gmail.com](mailto:horvath.peeteer@gmail.com)

Kundráth Péter

[peter.kundrath@gmail.com](mailto:peter.kundrath@gmail.com)

Palkó Gábor

*Eötvös Loránd Tudományegyetem, Digitális Örökség Nemzeti Laboratórium*  
[palko.gabor@btk.elte.hu](mailto:palko.gabor@btk.elte.hu)

ELTE Folk Song Corpus is a database that stores Hungarian folk songs with automatically generated annotations of the folk songs' structural units, grammatical features and sound devices. In the annotation process we followed the workflow developed for the ELTE Poetry Corpus. The corpus has an open access online query tool with several search functions. Besides the annotation process and the query tool of the corpus, the paper presents a quantitative comparison of the ELTE Folk Song Corpus and the ELTE Poetry corpus on the basis of lexical features, rhyme patterns and metrical properties.

**Keywords:** folk songs, corpus, automatic annotation, ELTE Poetry Corpus

### 1. A korpusz főbb jellemzői és létrehozásának lépései

Az ELTE Népdalkorpusz az ELTE Verskorpusz mintájára létrehozott, magyar népdalokat tartalmazó, automatikusan annotált adatbázis. A korpusz forrását az Ortutay Gyula szerkesztésében és Katona Imre válogatásával megjelent Magyar népdalok című gyűjteményes mű második, 1976-os kiadása adta, amely megtalálható a Magyar Elektronikus Könyvtár adatbázisában. A korpuszban szereplő népdalok száma 2390, a korpusz szavainak száma 110 ezer, a tokenek száma pedig 150 ezer. A korpusz a népdalok szövegei mellett három annotációs réteget tartalmaz: annotáltuk a népdalok szerkezeti egységeit, a szavak grammatikai tulajdonságait, valamint a hangzásjellemzőkhöz kapcsolódó poétikai jellemzők bizonyos típusait.

A korpusz létrehozása során az ELTE Verskorpusz létrehozásához kidolgozott módszertant követtük (Horváth et al. 2022).<sup>1</sup> Első lépésben egy szkript segítségével a Magyar Elektronikus Könyvtár oldaláról letöltött HTML fájlokat átalakítottuk olyan TEI XML fájllokká, amelyek tartalmazzák a szövegek szerkezeti egységeinek, azaz a címeknek, a strófáknak és a soroknak az annotációit. Ezt követően az e-magyar emtsv változatával tokenizáltuk a szövegeket, és annotáltuk a szavak szótári alakját, szófaját és morfoszintaktikai jellemzőit (Váradi et al. 2017, Indig et al. 2019). A korpusz létrehozásának harmadik lépése a hangzásjellemzőkhöz kapcsolódó poétikai tulajdonságok automatikus annotálása volt, amelyhez az ELTE Verskorpusz hangzásjellemzőinek az annotálására fejlesztett programot használtuk (Horváth 2020). A hangzásjellemzők annotálása a versszakok rímképletének, a rímpárt alkotó szavaknak, a sorok időmértékes ritmusának, az alliterációknak és a szavak fontosabb fonológiai

1 <https://github.com/ELTE-DH/poetry-corpus>



jellemzőinek, a szótagszámnak, a hangrendnek és a fonológiai szerkezetnek a felismertetésére terjedt ki. Természetesen az időmértékes ritmusnak népdalok esetében nincs túl nagy relevanciája. Az erre vonatkozó annotációkat azért nem töröltük, mert egyelőre nem akartunk eltérni az ELTE Verskorpusz esetében alkalmazott formátumtól. Végezetül egy XSLT stíluslap segítségével elvégeztünk a TEI XML fájlokban egy formátumátalakítást, azaz bizonyos XML-elemeket és -attribútumokat áthelyeztünk, illetve átneveztünk, valamint további, nagyrészt a különböző szerkezeti egységek szó- és szótagszáma vonatkozó annotációkkal bővítettük a fájlokat. Az így előállt XML-fájlok bár TEI-közeliek, de nem felelnek meg a TEI által megadott szabványnak. A formátumátalakítással és az annotációk bővítésével az volt a célunk, hogy a korpuszhoz a lekérdezéseket egyszerűbben meg lehessen írni, és gyorsabban le lehessen futtatni.

A fent bemutatott lépések során előálló, egyre több annotációs réteget tartalmazó különböző verziók a korpusz github oldaláról<sup>2</sup> szabadon letölthetők. A github oldalon és az ELTE Verskorpust bemutató tanulmányban (Horváth et al. 2022) részletes leírás olvasható a korpusz egyes verzióiban szereplő XML-elemekről és -attribútumokról.

## 2. A korpusz lekérdezőfelülete

A korpuszban való keresésekhez egy szabadon elérhető online lekérdezőeszközt fejlesztettünk a Verskorpusz lekérdezőeszközének mintájára. A korpusz szövegeit és annotációit tartalmazó XML-fájlokból egy MariaDB-alapú SQL-adatbázist hoztunk létre. Ebben keres a <https://verskorpusz.elte-dh.hu/nepdal> címen elérhető lekérdezőeszköz. A lekérdezőeszköz felületén számos keresési lehetőség közül választhatunk. Kereshetünk szóalakokra, lemmákra, morfoszintaktikai jellemzőkre, szótagszáma, hangrendre, szótagok hosszúságára, fonológiai szerkezetre, valamint ezek tetszőleges kombinációira. Úgyszintén kereshetünk ugyanezen jellemzők alapján több szóból álló szerkezetekre. Generálhatunk gyakorisági listákat is szavakra vagy szó szerkezetekre vonatkozóan, szóalakok vagy lemmák formájában, valamint szűrhetjük a népdalokat rímképletek alapján. A keresési eredményeket letölthetjük TSV-formátumban, amely a legtöbb táblázatkezelő programban megnyitható. Ha megnyitunk egy népdalt, akkor a felületen láthatjuk a népdal összes annotált jellemzőjét is.

## 3. A Népdalkorpusz és a Verskorpusz néhány lexikai jellemzője

Az alábbiakban a Népdalkorpusz és a Verskorpusz különböző lexikai jellemzőinek a kvantitatív összevetését mutatjuk be, érzékeltetve a Népdalkorpusz vizsgálatában rejlő lehetőségeket. Hangsúlyozandó, hogy míg az ELTE Verskorpusz a 20. század első feléig bezárólag tartalmazza a magyar kanonikus költészet legnagyobb részét, addig az ELTE Népdalkorpusz forrásául szolgáló gyűjteményes kötet anyaga egy önkényes, szubjektív szempontokat érvényesítő válogatás eredményeként jött létre. Mindez azt jelenti, hogy a Népdalkorpuszból kinyert kvantitatív adatokon alapuló következtetéseket inkább erős hipotézisekként érdemes kezelni. Az 1. táblázat a Verskorpusz és a Népdalkorpusz leggyakoribb tíz főnévi és igei lemmáját mutatja be az előfordulási számokkal. Félkövérrel emeltük ki azokat a lemmákat, amelyek az első tízben csak az egyik korpuszban fordulnak elő. A gyakorisági listákat a lekérdezőfelülettel generáltuk.

---

2 <https://github.com/ELTE-DH/folk-song-corpus>

	FŐNEVEK				IGÉK			
	Verskorpusz (2,7 millió szó)		Népdalkorpusz (114 ezer szó)		Verskorpusz (2,7 millió szó)		Népdalkorpusz (114 ezer szó)	
1	szív	8235	isten	506	van	28515	van	1750
2	lélek	7461	rózsa	450	lesz	10940	lesz	582
3	szem	7204	szív	387	lát	7985	ad	481
4	isten	6610	baba	356	tud	6706	nincs	408
5	élet	6599	úr	304	él	5346	megy	407
6	ég	6251	anya	286	nincs	5213	szeret	370
7	világ	5773	lány	277	néz	5112	jár	365
8	föld	5544	legény	271	megy	4819	jön	344
9	nap	5520	ló	266	jön	4796	tud	323
10	szó	5000	nap	242	mond	4734	kell	318

1. táblázat. A Verskorpusz és a Népdalkorpusz leggyakoribb főnévi és igei lemmái

Atáblázatból látható, hogy a Verskorpusz leggyakoribb főnevei között nagyrészt absztraktabb, azaz nem tapintható, és sok esetben nem is látható entitásokra vonatkozó szavak jelennek meg, míg a Népdalkorpusz esetében előtérbe kerülnek a konkrétabb jelentésű főnevek. Ez utóbbiak sok esetben személyek által betöltött, a legalapvetőbb emberi kapcsolatok kontextusában értelmezhető szerepre vonatkoznak (*baba*, *anya*, *lány*, *legény*). Érdekes, hogy a Verskorpusz esetében egy személyre vonatkozó főnév sem szerepel a leggyakoribb tíz között. Az igék esetében a legszembevetőbb különbség az, hogy míg a Verskorpusz esetében két perceptuális jelentésű, látáshoz kapcsolódó ige is megjelenik (*lát*, *néz*), addig a Népdalkorpusz esetében egy ilyet sem találunk a leggyakoribb tíz ige között. Mindebből arra következtethetünk, hogy a magyar kanonikus költészetben a megfigyelő, szemlélődő attitűd erőteljesebben jelenik meg, mint a magyar népdalokban.

A két korpusz szókincsét nemcsak gyakorisági listák alapján, hanem kulcsszavak alapján is összevethetjük. Kulcsszavak azok a szavak, amelyek egy adott korpuszban nagyobb arányban fordulnak elő, mint egy másik korpuszban. A kulcsszavak tehát olyan szavak, amelyek egy adott korpust a legjobban jellemeznek egy másik korpuzshoz képest. Kulcsszavak kinyerésére számos számítás létezik, mi a Kilgarriff-féle *simple math* nevű eljárást implementáltuk. Az eredeti módszertől annyiban eltértünk, hogy a lemmák relatív gyakoriságát nem az alapján számoltuk, hogy egy szó összesen hányszor szerepel a két korpuszban, hanem az alapján, hogy azok hány darab versben, illetve népdalban fordulnak elő. A számítást olyan módon alkalmaztuk, hogy a nagyobb gyakorisággal megjelenő szavakra fókuszálva kérdezhessünk le kulcsszavakat (Kilgarriff 2009).<sup>3</sup> A 2. táblázat mutatja be az alkalmazott számítás alapján a Verskorpusz és a Népdalkorpusz tíz legnagyobb kulcsszóértékkel rendelkező szavát, vagyis a nagyobb gyakoriságú szavak közül azokat, amelyek a Verskorpuszban arányaiban jelentősen több szövegben fordulnak elő, mint a Népdalkorpuszban, illetve amelyek a Népdalkorpuszban arányaiban jelentősen több szövegben fordulnak elő, mint a Verskorpuszban. A kulcsszavak mellett feltüntettük, hogy azok hány versben, illetve népdalban fordulnak elő.

<sup>3</sup> A kulcsszavak számítása során 1000 versenkénti előfordulásra normalizáltuk a szóelőfordulásokat. A *simple math* paramétert 100 értékben határoztuk meg, aminek köszönhetően a nagyobb gyakoriságú, azaz a sok versben és népdalban megjelenő szavakra fókuszálva tudtuk elvégezni a kulcsszavak kinyerését.

	Verskorpusz (13 064 vers)		Népdalkorpusz (2390 népdal)	
1	s	10429	baba	247
2	és	9113	édesanya	169
3	amely	3418	legény	208
4	e	3932	kislány	105
5	lélek	3642	rózsa	295
6	mint	6225	szerető	174
7	ég	3966	ló	170
8	élet	3671	katona	108
9	minden	4478	mög	44
10	álom	2118	Kossuth <sup>4</sup>	47

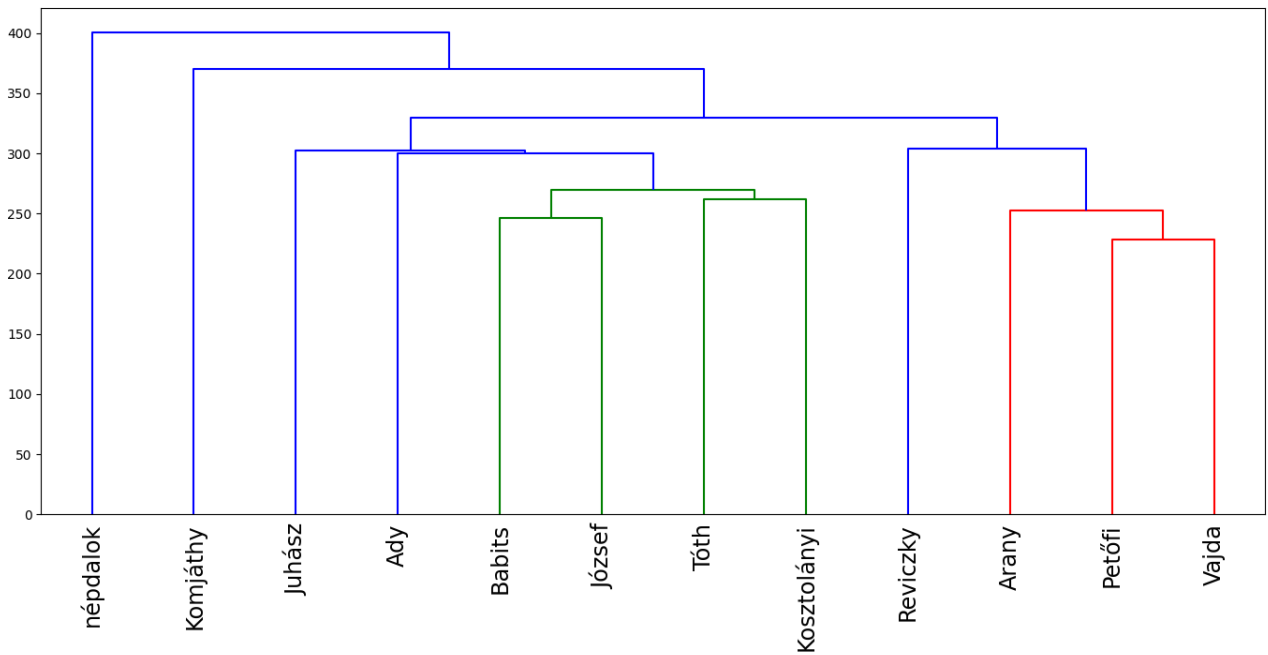
2. táblázat. A Verskorpusz és a Népdalkorpusz kulcsszavai

A két korpusz tíz legnagyobb kulcsszóértékével rendelkező szavának a listája hasonló tendenciát mutat, mint a főnevek gyakorisági listái. A táblázatból látható, hogy míg a Verskorpusz kulcsszavai között megjelenő főnevek (*lélek*, *ég*, *élet*, *álom*) absztrakt jelentésűek, addig a Népdalkorpusz első tíz kulcsszava nagyrészt konkrét jelentésű szó, amelyek többnyire valamilyen személy által betölthető szerepre utalnak. A táblázatból azt is láthatjuk, hogy a Verskorpuszban számos grammatikai szó is kulcsszóként jelenik meg, amelyek utalhatnak az összetettebb, illetve retorikusabb szövegszerkesztésre, például a mondaton belüli felsorolások és mellérendelő tagmondatkapcsolatok (*s*, *és*), az alárendelő tagmondatok (*amely*), a retorikai kérdések és rámutatások (*e*) vagy a hasonlatok (*mint*) nagyobb mértékű jelenlétére.

Érdekes kérdés az is, hogy a népdalok szókinccse és bizonyos költők szókinccse hogyan viszonyul egymáshoz, vajon vannak-e olyan költők, akiknek a szókinccse közelebb áll a magyar népdalokéhoz, mint más költőkéhez. Ehhez elvégeztünk egy agglomeratív hierarchikus klaszterelemzést, amelyhez alkorpuszként a Népdalkorpuszt, valamint a Verskorpusból kiszedett 11 költő összes versét használtuk, akik a következők voltak: Arany János, Petőfi Sándor, Vajda János, Reviczky Gyula, Komjáthy Jenő, Ady Endre, Juhász Gyula, Babits Mihály, Kosztolányi Dezső, Tóth Árpád, József Attila. A klaszterezéshez egy saját szkriptet használtunk, amely a leggyakoribb 1000 lemma alapján, a Burrows' delta eljárást alkalmazva adta meg az 1. ábrán látható kimenetet (Burrows 2002).<sup>5</sup>

4 A *Kossuth* tulajdonnév azért szerepel ennyi szövegben, mert a korpusz forrásául szolgáló gyűjteményes kötet tartalmaz egy „48-as dalok” tematikájú részt.

5 A 12 alkorpusz távolságának a kiszámításához használt Burrows' delta eljárás a szavak (jelen esetben lemmák) normalizált előfordulásainak standardizált értékei által kijelölt, az alkorpuszokat reprezentáló pontok közötti Manhattan-távolságon alapul. A klaszterek nagyobb klaszterekbe való sorolásához a klasztereket alkotó adatpontok közötti átlagos távolságot használtuk.



1. ábra A Népdalkorpusz és a Verskorpusz 11 költőjének hierarchikus agglomeratív klaszterelemzése

Az 1. ábrán látható, hogy szemben a megfogalmazott hipotézissel, egyik költő sem került a népdalokkal egy közös, alacsonyabb szintű klaszterbe. A klaszterelemzés eredményéből úgy tűnik, hogy a népdalok és a magyar kanonikus költészet szókincsé markánsan elválnak egymástól. Még az olyan, népies megszólalásmódokkal kísérletező költők, mint Petőfi vagy Arany szókincsé is jobban hasonlít más költők szókincsére, mint a Népdalkorpusz szókincsére. Megjegyzendő, hogy a klaszterelemzés Komjáthyt leszámítva jól el tudta különíteni egymástól a 19. és a 20. századi költőket.

#### 4. A Népdalkorpusz és a Verskorpusz néhány hangzásjellemzője

Mivel a Népdalkorpuszban és a Verskorpuszban nemcsak a szavak grammatikai jellemzőit, hanem a szövegek bizonyos hangzásjellemzőit is annotáltuk, ez utóbbiak alapján is összevethető a két korpusz. A 3. táblázat a két korpusz tíz leggyakoribb rímképletét mutatja be. Egy adott rímképlet előfordulásába csak azokat a verseket, illetve népdalokat számoltuk bele, amelyeknek mindegyik strófája ugyanazzal a rímképlettel lett annotálva.

	Verskorpusz (13 064 vers)		Népdalkorpusz (2390 népdal)	
	Rímképlet	Előford.	Rímképlet	Előford.
1	abcb	520	aabb	411
2	abab	458	aa	100
3	aa	305	aaaa	87
4	aabb	300	aabc	86
5	ab	209	abcc	70
6	abcd	186	aaba	27
7	aba	115	abcb	27
8	aab	74	aabbcc	24
9	aaaa	66	aaa	15
10	abba	65	aabbc	15
Összes		4959		1318

3. táblázat. A Verskorpusz és a Népdalkorpusz leggyakoribb rímképletei

A 3. táblázatból látható, hogy míg a Verskorpusz első két leggyakoribb rímképlete a félrím és a keresztrím, addig a Népdalkorpusz esetében a páros- és a bokorrímek vannak a lista élén, és csak a lista 6. helyén jelenik meg az első olyan rímképlet, ahol az egymással rímelő sorok közé beékelődik egy azokkal nem rímelő sor (aaba). A kapott eredmények megerősítik azt a már Arany János által is hangoztatott meglátást, miszerint „[o]da törekedvén a népköltészet, hogy az értelemnek mielőbb teljességet adjon, nem fűzi hosszan gondolatját, hanem siet azt befejezni, mi külsőleg a második rím által történik. Ezért nem kaphatott lábra népdalainkban a váltogató, vagy keresztrím, mint a mely hosszabb elnyújtását a gondolatnak engedi meg” (Arany 2012, 315).

A Verskorpusz és a Népdalkorpusz jelenlegi verziója csak az időmértékes ritmusra, azaz a hosszú és rövid szótagokra vonatkozó annotációkat tartalmaz, metrumra vonatkozó annotációkat nem. Az utóbbi időben azonban elkészült egy program, amely képes az alapvetőbb metrumok automatikus felismerésére (Horváth 2021). E program futtatásával például megtudhatjuk, hogy melyek a leggyakoribb ütemhangsúlyos metrumok a Verskorpuszban és a Népdalkorpuszban. Ezt a 4. táblázat mutatja be.<sup>6</sup>

	Verskorpusz (13 064 vers)		Népdalkorpusz (2390 népdal)	
	Ritmus	Előford.	Ritmus	Előford.
1	6   6	622	4   4	270
2	4   4	305	6	160
3	6	229	4   4   3	150
4	5   5	185	4   3	105
5	5   6	166	6   6	91
6	4   6	138	2   2   4	61
7	4   4 és 4   3	120	4   6	52
8	5   4	79	2   2   3	45
9	5   3	64	4   2   2	30
10	5   6 és 5   5	57	4   1   3	27

4. táblázat. A Verskorpusz és a Népdalkorpusz leggyakoribb ütemhangsúlyos metrumai

<sup>6</sup> A programot a következő beállításokkal futtattuk: megegyező szótagszámú sorok és ütemkezdetre eső szó eleji szótagok minimális aránya: 0,75; ütemek maximális szótagszáma: 6.



A 4. táblázatból látható, hogy a Verskorpusz esetében a leggyakoribb ütemhangsúlyos metrum a felező tizenkettes, a második leggyakoribb metrum pedig a felező nyolcas. A Népdalkorpusz esetében a leggyakoribb ütemhangsúlyos metrum a felező nyolcas, a második leggyakoribb metrumot pedig azok a szövegek adják, amelyek hat szótagos sorokból állnak, és a sorok további ütemekre nem oszthatók. Megjegyzendő, hogy ez utóbbi eset felező tizenkettesnek is tekinthető. A két korpusz leggyakoribb metrumai között két feltűnőbb különbséget találhatunk. Egyrészt a Verskorpusz esetében a hetedik és a tizedik helyen is megjelenik olyan metrum, ahol a páratlan és a páros sorokban eltérő ütemosztás található. Például a hetedik helyen szereplő 4 | 4 és 4 | 3 azt jelenti, hogy a versek páratlan sorai felező nyolcasok a páros sorok pedig kétütemű hetesek. Ilyen típusú, a páratlan és a páros sorok eltérő ütemezésére épülő metrumok a Népdalkorpusz esetében egyáltalán nem fordulnak elő a leggyakoribb tíz metrum között. Ez a különbség természetesen szorosán összefügg azzal a rímképletek kapcsán említett különbséggel, hogy a népdalok monolitikusabb rímelésével szemben a Verskorpuszra nagymértékben jellemzőek a félrímes, illetve keresztrímes megoldások. A másik szembetűnő különbség az, hogy a Népdalkorpusz esetében a leggyakoribb tíz metrumból öt három ütemből áll. A Verskorpusz esetében ilyen típusú metrum egyáltalán nem jelenik meg a leggyakoribb tíz között.

## 5. Összefoglalás

Az ELTE Népdalkorpusz egy olyan automatikusan annotált, szabadon elérhető adatbázis, amely reményeink szerint mind a magyar népdalokra, illetve a magyar költészetre vonatkozó kutatásokban, mind a közoktatásban hasznosítható lesz. A korpuszhoz fejlesztett lekérdezőfelület lehetővé teszi, hogy a felhasználó mélyebb informatikai tudás nélkül is különböző, a korpuszt jellemző kvantitatív adatokhoz jusson. A népdalok szövegei és az annotációkat tartalmazó XML-fájlok szabadon letölthetőek kutatás céljából a korpusz github oldaláról. A tanulmányban a Népdalkorpuszt a szókincs, valamint a rím és a metrum különböző kvantitatív jellemzői alapján vetettük össze az ELTE Verskorpusszal, rámutatva a két szövegtípus néhány különbségére és a Népdalkorpusz használatának a lehetőségeire. Az ELTE Népdalkorpusz nem egy lezárt projekt, a jövőben szeretnénk az adatbázist további népdalokkal, valamint további annotációs rétegekkel bővíteni.

## Bibliográfia

- Arany János: A magyar nemzeti vers-idomról. In Arany János: *Tanulmányok és kritikák I.* vál., szerk., az utószót és a jegyzeteket írta S. Varga Pál. Második, javított kiadás. Debrecen, 2012, Debreceni Egyetemi Kiadó. 288–320.
- Burrows, John: 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 2002. vol. 17. No. 3. 267–287. <https://doi.org/10.1093/lc/17.3.267>
- Horváth Péter: A vershangzás jellemzőinek automatikus feltárása József Attila verseiben. *Digitális Bölcsészettudomány*, 2020. 3. sz. M:3–M:27. <https://doi.org/10.31400/dh-hun.2020.3.422>
- Horváth Péter: Két eljárás magyar nyelvű versek metrumának gépi felismertetéséhez. *Digitális Bölcsészettudomány*, 2021. 4. sz. T:79–T:103. <https://doi.org/10.31400/dh-hun.2021.4.2361>



- Horváth Péter, Kundráth Péter, Indig Balázs, Fellegi Zsófia, Szláwich Eszter, Bajzát Tímea Borbála, Sárközi-Lindner Zsófia, Vida Bence, Karabulut Aslihan, Timári Mária, Palkó Gábor: ELTE Verskorpusz – a magyar kanonikus költészet gépileg annotált adatbázisa. In *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette Berend Gábor, Gosztolya Gábor, Vincze Veronika. Szeged, 2022, Szegedi Tudományegyetem TTIK, Informatikai Intézet. 375–388.
- 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. Szeged, 2019, Szegedi Tudományegyetem TTIK, Informatikai Intézet. 235–247.
- Kilgarriff, Adam: Simple math for keywords. In *Proceedings of the Corpus Linguistics Conference 2009 (CL2009). Held at the University of Liverpool, UK, 20-23 July 2009*, edited by Michaela Mahlberg, Victorina González-Díaz, Catherine Smith. [https://ucrel.lancs.ac.uk/publications/cl2009/171\\_FullPaper.doc](https://ucrel.lancs.ac.uk/publications/cl2009/171_FullPaper.doc)
- Váradi Tamás, Simon Eszter, Sass Bálint, Gerőcs Mátyás, Mittelholtz Iván, Novák Attila, Indig Balázs, Prószéky Gábor, Farkas Richárd, Vincze Veronika: Az e-magyar digitális nyelvfeldolgozó rendszer. In *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerkesztette Vincze Veronika. Szeged, 2017, Szegedi Tudományegyetem, Informatikai Intézet. 49–60.

## IKT eszközök alkalmazása az alsó tagozatos környezetismeret órákon

Nagy György  
Tokaj-Hegyalja Egyetem  
Comenius Intézet, Természettudományi Tanszék  
[nagy.gyorgy@unithe.hu](mailto:nagy.gyorgy@unithe.hu)

### Application of ICT tools in primary school environmental education classes

#### Abstract

Education in the 21st century is unimaginable without the innovative application of technological developments. The way of cognition, understanding and learning affects the future applicability of knowledge, competences. In primary schools, the environment-related content may be, used most effectively when all senses are involved in the acquisition according to the content of the knowledge and it reaches the most efficient channel or combination of channels until logical processing.

**Keywords:** environment, ICT tools, constructivism

#### Bevezetés

A XXI. század technológiai fejlődése által indukált kihívások jelentős feladat elé állítják az oktatási rendszert. Nagy mennyiségben jelentek meg az átlagdiák tudástárában olyan ismeretek, melyeknek nem az iskola volt a forrása. Ennek tartalmi elemei nagyrészt az információs-kommunikációs technológiához köthetők. A COVID fertőzés elterjedése még jobban indokoltá tette az IKT tudás kiszélesítését, alkalmazhatóságának növelését. A hirtelen jött változás, és a hozzá néhol adaptálódni nehezen tudó oktatási rendszer tehetetlensége okán lassan történtek a változások. Sok helyen kellő figyelemmel és fegyelemmel tértek át a digitális technológiák alkalmazására. Azonban nagyon sok helyen ez lassan, és akadozva történt. Széchényi szavai rég voltak olyan aktuálisak, mint a COVID járvány idején: „Merjünk nagyok lenni, s valóban nem olyan nehéz, de legyünk egyszersmint bölcsek is”. A megoldási sémákat sok esetben a körülmények határozták meg, és nem a hatékony magvalósítási formák. Alkalmazkodni kellett a meglévő szoftverekhez, hardverekhez és a diákok és pedagógusok IKT hozzáértéséhez. A kidolgozott magvalósítási folyamatokat általánosították minden területre, néha nem kellő bölcsességgel. Azonban a tanulók által szerzett érdemjegyek alig térnek el a COVID időszak előttiektől, az oktatási folyamat tehát sikeres volt. A nevelési folyamat egészéhez kaptak a pedagógusok adalékokat a tanulókkal folytatott interakciók során, azonban ez nem ad teljes képet a folyamatról. A kompetenciák fejlődésében pedig értékelhető eredményeket a 2022-es Pisa felmérést után mondhatunk. A járvány időszak megmutatta, hogy milyen az, amikor nagyrészt IKT eszközökön keresztül folyik a nevelés és az oktatás.

Ma már nem az a kérdés, hogy van-e helye az oktatásban az IKT eszközöknek és az alkalmazásukhoz szükséges tudásnak, hanem az, hogy hol és milyen mélységig lehet, vagy kell használnunk a maximális hatékonyság érdekében. Sarkalatos pontja az oktatási rendszernek a természettudományok oktatása. A jelenlegi NAT a kisiskolás korosztály számára nem ír elő kötelező Környezetismeret órát, hanem „a természettudományos gondolkodás megalapozása az alapfokú képzés első szakaszában a magyar nyelv és irodalom tanulási területének tudásbővítést és olvasásfejlődést segítő olvasmányai (1–2. évfolyam) ágyazva kezdődik.”[1] A természettudományos megismerés módjait és lépéseit (megfigyelés,

kísérlet, mérés) tanórai formában a 3. osztálytól kezdik megismerni, elsajátítani a tanulók. A természetismeret óra a legtöbb esetben a gyerekek számára élménygazdag módszereket kínál a megismerésre és a készségek kialakítása játékos formában történik a lehető legtöbb érzékszerv bevonásával. Az átadásra váró ismeretek mennyisége azonban nem csökken, inkább növekvő tendenciát mutat. Sajnos a rendelkezésre álló óraszám nem elegendő arra, hogy minden ismeretet tanórai körülmények között sajátítsa el a tanuló. Arra kell törekednie a pedagógusnak, hogy megtanítsa a gyereket az önálló ismeretszerzésre, az ismeretek logikai feldolgozására, értelmezésére és alkalmazására. Azok a tanulók, akik hatékonyan alkalmazzák a természettudományos megismerés módszereit és az életkori sajátosságoknak megfelelő tanulási stratégiákat, jól motiváltak, nem igénylik a tanár előadását, inkább mentorként tartanak rá igényt. A tanóra mozzanatai is transzformálódnak, már nem motivációról és célkitűzésről, hanem ráhangolódásról, nem ismeretszerzésről, hanem jelentésteremtésről, és nem rögzítésről és alkalmazásról, hanem reflektálásról beszélünk (RJR módszer) [9]. A továbbiakban arra keresem a választ, hogy az IKT mennyiben és hol tud hatékony segítséget nyújtani ebben a folyamatban.

### Problémafelvetés

A legutóbbi Pisa felmérés (2018) eredményei a természettudomány területén [2] nem mutatnak javulást az előző méréshez képest. Ismét az átlag alá került Magyarország, olyan országokkal mutat szignifikáns egyezést a pontszám, mint Spanyolország, Litvánia, Oroszország, vagy Luxemburg. Az eredmények ismeretében tükrében elmondható, hogy az iskolákban alkalmazott módszertannal nem lehet szintet lépni. A nevelés –oktatás történeti előzményeit áttekintve ilyenkor jelentős változtatásra, több esetben paradigmaváltásra volt szükség, és a teljesen újszerű ötletek hozták el a várva várt sikert. Ilyen paradigmaváltás volt az első tanóraba szervezett előadás, vagy a szemléletesség pedagógiájának bevezetése Comenius által. Az IKT bevezetése valószínűleg nem fog ilyen erejű változást hozni a természettudományos megismerés területén, de jelentősen hozzájárulhat annak sikerességéhez. Lehetséges-e a segítségével motiváltabb, a konstruktivista tanulásfelfogást alkalmazó, egy konyhát is laborrá változtató diákok kinevelése? A kisiskolások esetében az életkori sajátosságok figyelembe vételével a következő megállapításokat vizsgálom:

- 1) Az IKT alkalmazása motiváltabbá teheti a tanulókat és a pedagógusokat a természettudományos megismerés területén.
- 2) A tanóratípusok is jelentősen kihatnak az IKT alkalmazhatóságának hatáira.
- 3) A különböző típusú tanórák momentumaiban eltérő hatékonysággal alkalmazható az IKT.
- 4) A IKT alkalmazása félresöpörheti a megismerési folyamatban a maximálisan bevonható érzékszervek elvét, csak vizuális és akusztikus információkon keresztül szolgáltat ismereteket.
- 5) A természettudományos megismerés gyakorlatában alkalmazható mozgásműveletek nem sajátíthatók el digitális platformokon.

### Történeti előzmények

Az őskorban a környezetet az élő természet jelentette, és az ismeretek utánpótlás során kerültek át nemzedékről nemzedékre. A túléléshez minden érzékszerv bevonására szükség volt, a számonkérés eredménye pedig maga az élet vagy a halál volt. Az ókorban megnövekedett a tanult emberek iránti igény, így áthelyeződött a helyszín, már mesterséges lett a környezet, és a módszertant a szóbeli közlés uralta, hallás után tanultak



a diákok. Comenius volt az első, több száz év után, aki ismét a természetes környezetet tette tanteremmé, és a legtöbb érzékszerv bevonását ösztönözte a megismerési folyamat során (érezékszervek általi tanulás). „Semmi sincs meg a képzetekben, ami ne lett volna meg az érzetekben” –vallja [3]. Az önálló tanulói cselekvés még nem jelent meg, a pedagógus fegyelmet követelt, és minden iskolást a maga képére kívánt formálni. Az újkor tanulója már a természetből, vagy annak rajzolt másából tanulhatott (illusztrált könyvek, pl. „Orbis pictus”). A legújabb kor már engedékenyebb a tanulói cselekvéssel szemben, engedélyezi azt, az ismeret forrása lehet a természet is, de minden a pedagógus irányításával, központi szerepével történik. Az eddigi passzív befogadó – a tanuló – most aktív részese lett az eseményeknek, az ismeretek értelmi feldolgozása a személyes tapasztalaton, cselekvésen alapul [4]. A globális problémák korával eljött a kontrollált cselekvés általi tanulás ideje, ahol a természet, vagy a tanterem adott otthont az új ismeretek megalkotásának, megkonstruálásának a tanulóknak által. A jó pedagógus csak irányít, és hagyja, hogy magától törjön elő a tudás. A módszer egyszerre deduktív és induktív [5], az ismeret közvetítője a tanuló de a folyamat irányítója a pedagógus. A XXI. századdal pedig a környezet mellett megjelent a kibertér [6], annak virtuális és kiterjesztett valósága. Ennek az elterjedt alkalmazása inkább a technológiai háttér hiánya miatt lassú, és nem a jól megfontolt emberi szándék a kerékkötője. A tendenciákat figyelembe véve az ismeretszerzés környezete egyre jobban a kibertér felé tolódik el, évszázadok múlva úgy emlegetik majd mai korunkat, mint a kiberőskort.

## Elemzés

A természettudományok oktatásának a fenntartható fejlődés elérése, a környezettudatos életmód elsajátítása kell, hogy a középpontjában álljon, és ehhez kell eszközként a biztonságos tájékozódás térben és időben, a környezet elemeinek és kölcsönhatásainak ismerete. A kibertér nagyban segítheti ezt a folyamatot, azonban nem tud a természet helyébe lépni, nem tud minden érzékszervet mozgósítani a megismerési folyamatban. [7] A tanítás-tanulási folyamat a megfigyelhető, érzékelhető, észlelhető, vizsgálható, esetleg mérhető természeti jelenségekre, élőlényekre, folyamatokra épül. A természetismeret tárgy tanításának célja az érdeklődés felkeltése a környezetük élő és élettelen világa iránt, annak megismerése, védelme, mindezt eredményező magatartás kialakítása. [8] Az érdeklődés felkeltése pedig a motivációval, annak folyamatos megtartásával történik. Az IKT alkalmazása a motiváció során azért hatékony, mert az eszközök újdonsága, az új technikai megoldások még izgalmasak (pl. virtuális, vagy kiterjesztett valóság). Azonban ezek megismerése után jelentősen visszaesik hatékonyságuk. Az élő bemutatás, vagy interaktív kísérlet a tartalmi elemek folyamatos változása miatt nem veszít hatékonyságából. Azokban az esetekben, ha a téma nem engedi meg ezeket, akkor bátran fordulhatunk az IKT-hoz. Azonban egy vizuális élmény sem pótolja azt a motivációt, amit a tanuló által készített vulkánmodellből előtörő láva látványa okoz. A természetismeret oktatásában alkalmazott óratípusok szintén különböző lehetőségeket biztosítanak az IKT alkalmazására. A gyakorlati óra a meglévő ismeretek begyakorlására, rögzítésére szolgál. A cselekvő tevékenység mellett a feladatlapok kitöltése is része. Ebben az esetben az IKT-s eszközön kitöltött feladatlap azonnal értékelhető és elemezhető a kellő programmal, és azonnal lehet reflektálni az eredményekre, a hiányosságokat kiküszöbölni. A rendszerező, összefoglaló órák esetében a digitális tábla adta lehetőségek segítik a folyamatot, átláthatóvá teszik a tanultak rendszerét, érdekes adalékokat szolgáltathatnak a tananyaghoz. Az ellenőrző órák esetében a digitális platformon történő számonkérés gyors értékelésre ad lehetőséget, a statisztika eszközeivel pedig a tanulási folyamatban jelenlévő tendenciák is nyomon követhetőek.

Az új anyag feldolgozására szolgáló környezetismeret órák az összóraszám 80 %-át teszik ki. A tanóra felépítését az RJR modell alapján mutatom be [9]. Számonkérés során az osztály létszámától függően 1–2 tanuló feleltetése szóban történik, 4–5 tanuló pedig ez alatt írásban felel. Az osztály a szóban felelőt figyeli, ha kell, kiegészíti a feleletet. Az IKT a szóbeli feleletet nem váltja ki, arra szükség van a folyamatos beszéd, a csoport előtti szereplés, a gyors gondolkodás gyakorlása érdekében. Azonban a feladatlapos megoldásra sokkal hatékonyabb megoldást kínál, mint a hagyományos papír alapú forma. Míg azt csak óra után ellenőrizhetjük, időhiány miatt, az IKT eszközön megoldott feladatlap azonnal javítható, ellenőrizhető és a tanulónak nem kell a következő óráig várnia (stresszelnie), hogy megtudja az eredményt. A ráhangolódás első momentuma a motiváció, a következő a célkitűzés, vagyis az óra témájának megfogalmazása, amit általában a motiváció hatására a tanulók tesznek meg. Ezt követi a jelentésteremtés, ami a meglévő ismeretek felidézésével kezdődik, ebben segíthetnek az IKT eszközök, felidézve akár tárgyi ismereteket, óra- vagy kirándulásrészleteket. Ez után következik az új ismeretek nyújtása, ahol a lehető legtöbb érzékszerv bevonása szükséges, és az érzelmi kötődésre is lehet építeni (pl. egy állat bemutatása esetében). Vannak kifejezetten olyan tananyagok, amelyeket IKT-val lehetetlen hatékonyan átadni. Például be lehet mutatni a szőlő képét, fejlődését, de a tapintását, illatát, ízét még nem képes visszaadni a legmodernebb IKT eszköz sem. Arról nem is beszélve, hogy különböző élménypedagógiai elemekkel lehet kombinálni a megismerést (pl. varázsdoboz). A következő mozzanat a megszerzett ismeret logikai feldolgozása, beépítése a meglévő sémába, amit asszimiláció vagy akkumuláció kísér. Az IKT segíthet a tanítónak abban, hogy a tanuló által megkonstruált ismeret minél jobban közelítsen a valósághoz, elkerülje a tudományos tévedéseket. A reflektálás a rögzítés-alkalmazás kombinációja. Itt is sikeresen használhatunk IKT eszközt, abban is segít, hogy takarékosan tudjunk bánni az időkerettel. A munkatankönyv, vagy munkafüzet feladatait digitális platformon megoldhatja a tanuló, azonban ez hatékonyságában messze alulmarad annál, amit egy kísérleti eszközökkel megoldott probléma adhat (pl. homok, fűrészpor és víz szétválasztása). Az eszközhasználat, a műveletekhez tartozó finom-motorika az IKT eszközökkel nem fejleszhető. Bizonyos mozgássorok jól fejleszhetők, pl. konzolok használatával (pl. PS, XBOX), de nem biztos, hogy ez a megtanult mozgássor biztosítja a laboreszközök készségszerű használatát.

Az összefoglalás, ellenőrzés IKT eszközökkel gyorsan és hatékonyan végrehajtható, több információ is kinyerhető belőle, pl. a hozzáadott pedagógiai értékről, ha a hagyományos frontális kérdésfelelet, vagy igaz-hamis állítások helyett jelöléstáblát alkalmazunk. Ennek lényege az, hogy az adott órán átadott konkrét ismerettel kapcsolatban (a tanóra ismeretanyagának mondatba foglalt eleme) a tanuló nyilatkozik arról, hogy ezt eddig is így tudta, nem új ismeret, vagy eddig nem így tudta, most már érti is, vagy eddig nem tudta, és most sem érti, vagy érti, esetleg olyan érdekes volt, hogy otthon is utána néz, vagy elvégzi a kísérletet. Egy ilyen táblázat kitöltése könnyen megtanítható a tanulóknak digitális platformon, ennek a kezelését és ellenőrzését a megfelelő szoftver azonnal elvégezheti, a kellő információkat szolgáltatva a pedagógusnak. Az értékelést természetesen semmilyen körülmények között ne adja át a pedagógus az IKT-nak, a személyes kapcsolat nagyon fontos a gyerek lélektani fejlődésének szempontjából, ezt veszítené el a gyerek, ha mesterséges intelligencia végezné az értékelést.



## Konklúzió

Az IKT eszközök kellő odafigyeléssel jelentősen megnövelhetik a pedagógus munkájának hatékonyságát, azonban ahhoz is vezethet használatuk, hogy a pedagógus elkényelmesedik, és minden ismeretet ezen keresztül juttat el a tanulóhoz, ami nem szolgálja a fejlődésüket, alulmotiváltak lesznek, és elveszítik érdeklődésüket a természettudományok iránt. A feladatunk az, hogy ezeket az eszközöket hasznosan illesszük be az oktatási nevelési folyamatba. Ehhez viszont hozzáértő pedagógusokra van szükség, akik a módszertani tudásuk mellett megfelelő digitális kompetenciával is rendelkeznek. Nem mellékes az sem, hogy az IKT eszközök nem minden iskolában állnak rendelkezésre, így a saját eszközök használatát is fontolóra kell venni egyes esetekben.

Összefoglalva a leírtakat, az IKT alkalmazása motiváltabbá teheti a tanulókat és a pedagógusokat a természettudományos megismerés területén, a tanórátípusoknak és tartalmuknak megfelelően kell alkalmazni őket, sőt a tanórán belüli mozzanatok esetében sem egyforma hatékonysággal alkalmazhatóak. A IKT alkalmazása segíthet egyes érzékszerveink érzékenységének javításában (pl. lassított felvétel egy ütközésről, gyorsított felvétel egy égésről, nagyítás készítése, stb.), azonban nem pótolja minden érzékszervünket. A természettudományos megismerés során alkalmazott műveletek begyakorlására pedig a legtöbb esetben alkalmatlan, csak egyes mozgássorok finom-motorikáját erősíti meg.

## Irodalomjegyzék

1. Nemzeti Alap Tanterv. Magyar Közlöny 2020. 17. szám, 296. oldal
2. [https://www.oktatas.hu/pub\\_bin/dload/kozoktatas/nemzetkozi\\_merese/pisa/PISA2018\\_v6.pdf](https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_merese/pisa/PISA2018_v6.pdf) (letöltés 2022.06.20.)
3. Comenius Ámos János (1896). Nagy oktatástan. (ford. Dezső Lajos). Sárospatak: Steinfeld Jenő
4. Nahalka István (2002). Hogyan alakul ki a tudás a gyerekben: Konstruktivizmus és pedagógia. Budapest: Nemzeti Tankönyvkiadó
5. Jean Piaget (1970). Válogatott tanulmányok. Budapest: Gondolat
6. Komenczi Bertalan (2009). Elektronikus tanulási környezetek. Budapest: Gondolat
7. Nagy György: A technológiai környezet és az alsó tagozatos környezetismeret oktatás. In: Lengyel, Molnár Tünde (szerk.) Agria Media 2017, ICI 15 Információtechnikai és Oktatástechnológiai Konferencia és Kiállítás : Program- és absztrakt füzet. Eger, Magyarország : Eszterházy Károly Egyetem (2017) 106 p. p. 76 <https://doi.org/10.17048/AM.2018.40>
8. Nagy László (1972). Didaktika gyermekfejlődési alapon. In: Nagy László válogatott pedagógiai művei. Budapest: Tankönyvkiadó
9. Nagy, György: Konstruktivista tanulásfelfogás a természettudományok oktatásában. In: Berghauer-Olasz, Emőke; Greba, Ildikó; Hutterer, Éva; Pallay, Katalin (szerk.) Innovatív módszerek a pedagógiai-pszichológiai gyakorlatban Ukrajna európai integrációjának tükrében. Beregszász, Ukrajna : Kálvin Nyomda, II. Rákóczi Ferenc Kárpátaljai Magyar Főiskola (2016) pp. 323–329. , 7 p.



Multimédiás, interaktív és adaptív tananyagok létrehozásának lehetőségei H5P  
keretrendszerrel  
Creating Multimedia, Interactive and Adaptive Learning Materials with H5P Framework

Köpösdi Zsuzsa

*Debreceni Egyetem, Multimédia és E-learning Technikai Központ*

[koposdi.zsuzsa@metk.unideb.hu](mailto:koposdi.zsuzsa@metk.unideb.hu)

ORCID: [0000-0002-2185-4887](https://orcid.org/0000-0002-2185-4887)

Molnár Tamás

*Debreceni Egyetem, Multimédia és E-learning Technikai Központ*

[molnar.tamas@metk.unideb.hu](mailto:molnar.tamas@metk.unideb.hu)

### Abstract

The production of interactive and adaptive learning materials containing multimedia elements usually requires a higher budget and expertise, which greatly hinders and slows down the use and dissemination of this type of materials. While more and more research results show that these types of curricula are more effective in learning processes than curricula that contain only textual and pictorial elements. Using the H5P framework can provide a solution to this problem: the framework is free, open source, user-friendly and requires no special IT expertise. With H5P, we can create interactive and adaptive learning materials relatively easily and quickly. The wide applicability of H5P is also an important advantage: the framework can be used integrated in LMS systems (Moodle, Canvas) as well as in websites (Wordpress, Drupal). With our presentation and publication our goal is to further strengthen the H5P framework and the use of this type of learning material in the Hungarian educational process.

**Keywords:** e-learning, Interactive Learning Materials, Adaptive Learning Materials, H5P, Moodle

### Bevezető

A multimédia elemeket tartalmazó, interaktív és adaptív tananyagok előállítása általában magasabb költségkeretet és szakértelmet igényel, ami az ilyen típusú tananyagok alkalmazását, elterjedést nagymértékben gátolja, lassítja – miközben egyre több kutatási eredmény mutat rá arra, hogy az ilyen típusú tananyagok hatékonyabbak a tanulási folyamatokban a kizárólag szöveges és képi elemeket tartalmazó tananyagokkal szemben. A H5P tartalomfejlesztő keretrendszer alkalmazása megoldást jelenthet erre a problémára: a keretrendszer ingyenes, használatához nem szükséges speciális informatikai szakértelem, könnyen kezelhető és felhasználóbarát. A H5P segítségével viszonylag egyszerűen és gyorsan hozhatunk létre interaktív, illetve adaptív tananyagokat. Célunk, hogy ráirányítsuk a figyelmet a H5P keretprogramra és a használatában rejlő lehetőségekre, hogy a H5P keretrendszer és az ilyen típusú tananyagelemek használata, illetve az ezekről szóló szakmai diskurzus tovább erősödjön a hazai oktatási folyamatokban.



## A H5P keretprogramról röviden

A H5P keretprogram ingyenes és nyílt forráskódú program, amely segítségével akár multimédia elemeket tartalmazó, interaktív és adaptív tananyagokat hozhatunk létre. A program használata egyszerű, nem szükséges hozzá speciális informatikai szakértelem, jól áttekinthető és felhasználóbarát szerkesztőfelületekkel rendelkezik. A segítségével gazdag, a tanulási folyamatokat jól támogató, a tanulók figyelmét megragadó tartalmakat tudunk könnyen és gyorsan előállítani, illetve, már meglévő tananyagainkat ezen irányelvek mentén átalakítani, továbbfejleszteni.

A H5P széleskörű alkalmazhatósága is fontos tényező: a program integráltan is működik bizonyos LMS rendszerekben – például Moodle, Canvas –, illetve honlapszerkesztőkben, tartalomkezelőkben is – például Wordpress vagy Drupal. A H5P-ben létrehozott tananyagokat – .h5p kiterjesztésű fájlban – letölthetjük, exportálhatjuk, illetve megfelelő keretprogramba importálhatjuk, újraszervezhetjük és így újra fel is használhatjuk – amely egyrészt nagyban megnöveli a tananyagkészítés hatékonyságát, másrészt kiszolgálja a mai igényeket a szabványos, újrahasznosítható tartalmak iránt. A H5P-ben létrehozott tananyag reszponzív, mobilbarát.

H5P tartalom többféle módon is létrehozható, ezek közül a két legáltalánosabb módszert emeljük ki. Egyrészt, a H5P hivatalos oldalán<sup>1</sup> tudunk létrehozni H5P tartalmakat, vagy le is tölthetünk onnan már elkészült fájlokat. Másrészt, a Moodle-ben, integráltan és egyszerűen is létrehozhatunk H5P tartalmat<sup>2</sup> – erre a módszerre kitérünk később részletesen, hiszen a Moodle az egyik legszélesebb körben alkalmazott LMS rendszer a magyarországi felsőoktatásban.

## H5P tananyagmodulok, H5P-ben létrehozható tananyagtipusok

A H5P hivatalos weboldalán mindegyik modul, tananyagtípus, kérdéstípus, feladattípus megtalálható részletes leírással és letölthető példákkal.<sup>3</sup> Itt csak néhányat mutatunk be ezek közül részletesebben és az itt következő felsorolás is csak néhány kiemelt példát tartalmaz. Az általunk most kiemelésre érdemesnek vélt modulok a következők:

- interaktív videó (Interactive Video)
- interaktív könyv (Interactive Book)
- idővonal (Timeline)
- képek interaktív összehasonlítása (Image Juxtaposition)
- válasz hangfelismeréssel kérdéstípus (Speak the Words)
- esszé kérdéstípus az értékelés bizonyos fokú automatizálhatóságával (Essay)
- elágazó forgatókönyv (Branching Scenario)
- képek sorba rendezése (Image Sequencing)
- memóriakártya (Memory Game)

A felsoroltak közül tekintsünk át néhány modult, tananyagtypust részletesebben.

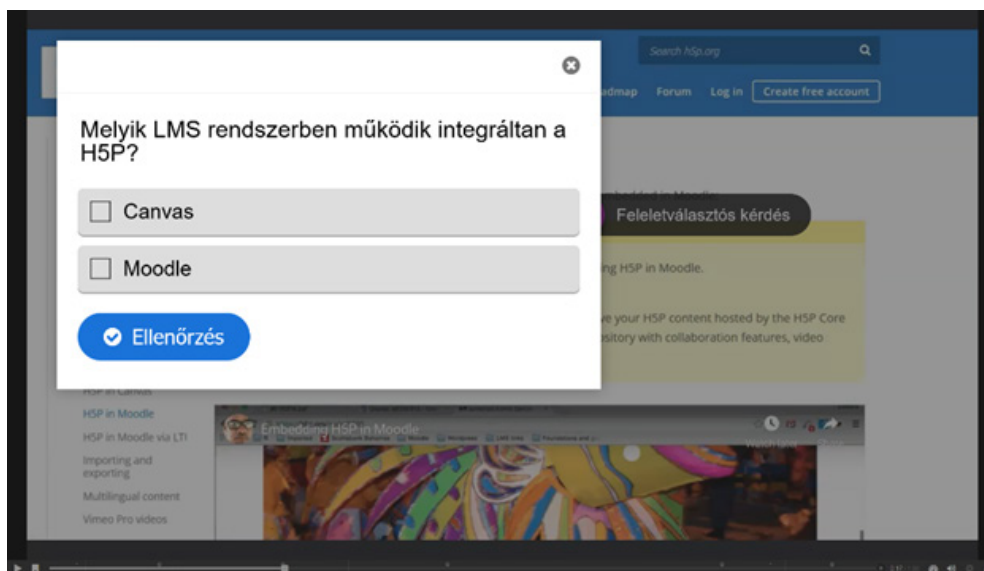
1 H5P Create and Share Rich HTML5 Content and Applications. <https://h5p.org/> Hozzáférés: 2022. június 20.

2 H5P - MoodleDocs. <https://docs.moodle.org/39/en/H5P> Hozzáférés: 2022. június 20.

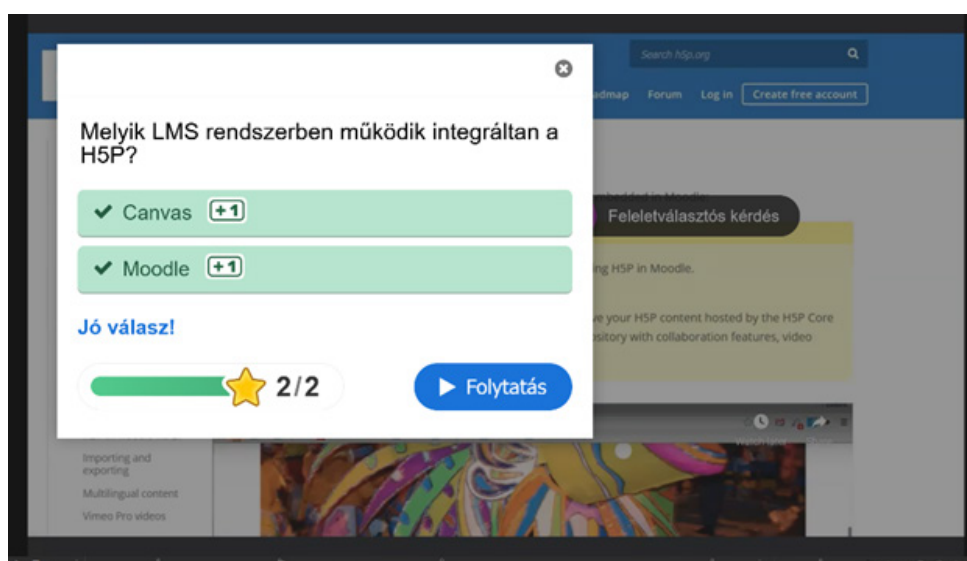
3 Examples and Downloads | H5P. <https://h5p.org/content-types-and-applications> Hozzáférés: 2022. június 20.

Az interaktív videó létrehozásakor az eredeti, alapként használt videofelvételhez különböző interakciókat, tevékenységeket tudunk hozzárendelni. Így a hallgató számára a videó megtekintése közben – az oktató által meghatározott pontokon – különböző tevékenységek, feladatok, kérdések jelennek meg, amelyekkel interakcióba kell lépnie, amelyeket el kell végeznie a videó megtekintésének folytatásához.

Az egyik leggyakrabban alkalmazott gyakorlat az interaktív videóban az önellenőrző kérdések beiktatása a videó bizonyos szakaszai után azzal a céllal, hogy a hallgató ellenőrizhesse, mennyiben értette meg és sajátította el az addig elhangzottakat. Az ilyen típusú, videóban elhelyezett interakciók a videó passzív megtekintése helyett aktív cselekvésre készítik a hallgatót. A megtekintés során folyamatosan megjelenő tevékenységek segítik a hallgatók figyelmének folyamatos fenntartását, amely ma már egyre nehezebben megvalósítható cél az oktatási, tanulási folyamatok során. Mindezek mellett, a kérdések megválaszolása után a hallgató azonnali visszajelzéseket kap a válaszairól – ami szintén fontos elem a tanulási folyamatok támogatásában.



1. ábra Interaktív videóban felugró feleletválasztós kérdés



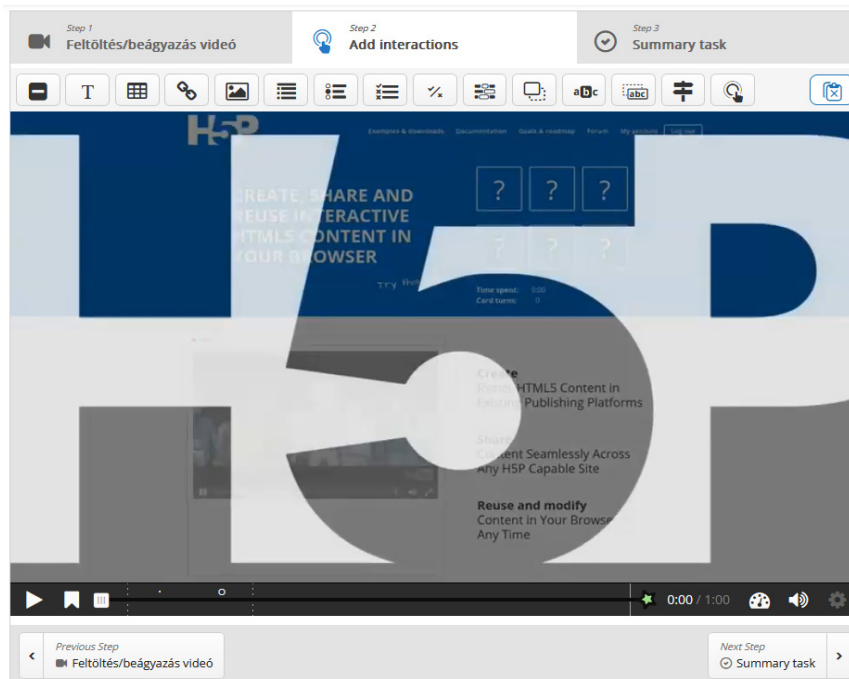
2. ábra A kérdés azonnali értékelése az interaktív videóban

Az interaktív videóban többféle beállítási lehetőségről is dönthetünk kérdés tevékenység esetén: beállítható, hogy többször is próbálkozhat-e egy-egy kérdés megválaszolásával a hallgató, hogy kérheti-e a helyes válasz megtekintését, hogy folytathatja-e a videót rossz válasz esetén is, illetve, hogy a videó mely pontjától folytathatja a megtekintést helyes, illetve mely pontjától helytelen válasz esetén – amellyel tulajdonképpen adaptívvá is tehető a videó.

A különböző kérdéstípusokon túl – mint, igaz-hamis kérdés, egyszeres és többszörös feleletválasztós kérdés, kiegészítendő kérdés stb. –, a következő interaktív tevékenységeket használhatjuk még a videofelvételekhez: beszúrhatunk plusz információkat tartalmazó szöveges vagy képi elemeket, táblázatokat, külső hivatkozásokat, illetve, speciális tevékenységként az ún. elágazási pont (Crossroads) tevékenység is használható. Az elágazási pont tevékenység használatával választási pontot állíthatunk be a hallgató számára, ahol a választása határozza meg, hogy a videó mely pontjától folytatódik számára a lejátszás. Ehhez hasonló a H5P egyik izgalmas modulja az ún. elágazó forgatókönyv (Branching Scenario), ahol szintén ilyen választási pontokat tudunk beállítani, de ott sokkal bonyolultabb forgatókönyv is összeállítható a videót megtekintő hallgató számára. Elágazó forgatókönyv tananyagmodullal bonyolultabb, választási pontok mentén elágazó videót tudunk tehát összeállítani.

Az interaktív videót könyvjelzőkkel is elláthatjuk a jobb áttekinthetőség és a hatékonyabb, kényelmesebb felhasználói élmény érdekében. Illetve, a videó végén egy speciális ún. összefoglaló (Summary) feladat keretében állításokról kell eldöntenie a hallgatónak, hogy azok közül melyik igaz – így egy átfogó feladatban összefoglalva és ellenőrizve a teljes megtekintés során megszerzett tudást.

Az interaktív videó létrehozásához első lépésben az eredeti videót kell feltöltenünk – illetve szerkesztéskor betöltenünk – a H5P szerkesztőbe. Következő lépésként – az interaktív tevékenységek szerkesztő felületén – a videó általunk kiválasztott pontjaihoz tudjuk hozzárendelni a különböző típusú tevékenységeket. A teljes folyamat alatt bármikor elindíthatjuk az eredeti videót és a lejátszás megállításával is kiválaszthatjuk a kívánt pontot a tevékenység beszúrásához. A szerkesztőfelület egyszerű, jól áttekinthető, gyorsan rutin szereshető a használatában.



3. ábra Az interaktív videó modulban az interaktív tevékenységek szerkesztőfelülete

Egy-egy tevékenységnél többféle lehetőségről is dönthetünk, így beállítható többek között, hogy a videó lejátszása folytatódjon-e a tevékenység megjelenésekor vagy álljon meg a videó, hogy a tevékenység gombként jelenjen meg vagy rögtön maga a tevékenység legyen látható a hallgató számára, illetve, hogy pontosan mettől meddig legyen látható a tevékenység.

A H5P keretrendszerben előállítható tananyag típusok közül az egyik leghasznosabb és leginkább előremutató az interaktív videó. Az általánosan ránk zúduló digitális információözmény miatt az e-learning kurzusok résztvevőinek, hallgatóinak is egyre nehezebb a figyelmét viszonylag hosszabb időre lekötöni – ebben jelent komoly támogatást ez a tananyag típus. Illetve, az általa megvalósítható azonnali visszajelzés is a tanulási folyamatokat támogatja.

A H5P modulok közül az idővonal (Timeline) is méltán népszerű a tananyag készítőik körében. Az idővonal szerkesztése szintén nagyon egyszerű: az időpontok, vagy időintervallumok megadása után megadhatjuk azok megnevezését, illetve képet, videót, külső hivatkozást, szöveges tartalmakat tudunk hozzájuk kapcsolni. Ezzel a hallgatók időbeli tájékozódását nagyban elősegítő tartalmakat tudunk előállítani. Az idővonalon megjelenő eseményeket, időpontokat, időintervallumokat, adatokat a hallgató önállóan fedezheti fel.

A képek interaktív összehasonlítása (Image Juxtaposition) is népszerű, és nem csak a tananyagfejlesztők körében, hanem az egyéb internetes tartalmakat előállítók körében is. A modul segítségével két képet tudunk egymásra helyezni úgy, hogy a hallgató egy csúszka segítségével tudja irányítani, hogy melyik kép milyen arányban legyen látható számára. Nagyon hatékony módszer arra, hogy két állapot összehasonlításában elmerüljön a hallgató – az itt látható példában azonos terület fotóit tudja egymásra vetítve összehasonlítani egy folyó áradása előtt és az áradás után. A tananyag létrehozásához csak fel kell töltenünk az adott szerkesztőfelületen a két képet, illetve meg kell adnunk a szükséges szöveges elemeket. Könnyen és gyorsan létrehozhatunk így a hallgatók körében kedvelt, és a figyelmüket megragadó tartalmakat.



4. ábra Példa a képek interaktív összehasonlítása tananyag típusra<sup>4</sup>

4 H5P Image Juxtaposition. <https://h5p.org/image-juxtaposition> Hozzáférés: 2022. június 20.





A képek sorba rendezése (Image Sequencing) típusú feladatban a hallgatónak az oktató által megadott instrukciónak megfelelően kell sorrendbe rendeznie a megjelenő képeket. Ennek a feladatnak az összeállítása is egyszerű és jól áttekinthető szerkesztőfelületen történik. A válasz hangfelismeréssel (Speak the Words) kérdéstípus leginkább a nyelvoktatásban használt kérdéstípus: a hallgató mikrofonba mondott választ hangfelismerés után ellenőrzi a rendszer. Az esszé kérdéstípus az érkelés bizonyos fokú automatizálhatóságával (Essay) a válaszban megjelenő előforduló kulcsszavakra adott pontokkal ad azonnali visszajelzést a hallgatónak.

## H5P tartalom létrehozása Moodle-ben

Ha Moodle kurzusunkban szeretnénk H5P tartalmat létrehozni, azt a következő egyszerű lépésekben tudjuk megtenni a H5P-Moodle integrációnak köszönhetően. Az adott kurzusban a "Tevékenység vagy tananyag beszúrása" gomb kiválasztásával a felugró ablakban az ikonok közül kiválasztjuk a "H5P" ikont. Itt a Moodle-ben már ismert űrlap használatával a tananyaghoz kapcsolódó beállításokat tehetjük meg: többek között a tananyag elérhetőségét határozhatjuk meg, illetve, természetesen a tananyag megnevezését is itt adhatjuk meg. Majd, a Tartalombankba innen belépve, ott a "Hozzáadás" legördülő menüjéből kiválasztjuk a használni kívánt H5P modult. Ezután az adott modul szerkesztőfelületével tudunk majd dolgozni a H5P fájl létrehozásán – például az interaktív videó esetében a fenti leírás és képernyőképek mentén.

The screenshot illustrates the Moodle interface for adding H5P content. It is divided into three main sections:

- Tevékenység vagy tananyag beszúrása:** A search bar at the top is followed by tabs for 'Mind', 'Tevékenységek', and 'Tananyagok'. A grid of activity icons is shown, with 'H5P' highlighted by a red box.
- Új H5P hozzáadása:** A form titled 'Új H5P hozzáadása' with a sub-tab 'Általános'. It includes a 'Név' (Name) field and a 'Csomagfajl' (Package file) section. A red box highlights the 'Tartalombankkal (új ablakban) szeleje H5P-állományait' (Select H5P files from the content bank in a new window) option.
- Tartalombank:** A content bank interface with a search bar and a 'Hozzáadás' (Add) dropdown menu. A red box highlights the 'Interactive Video' option in the dropdown list.

5. ábra H5P tartalom létrehozásának lépései Moodle-ben



## Összegzés

Előadásunk és publikációnk legfontosabb célja, hogy a H5P tartalmak előnyei, illetve előállításának módszerei minél szélesebb körben ismerté váljanak, hiszen a H5P tartalomfejlesztő keretrendszerrel speciális informatikai tudás nélkül is egyszerűen és viszonylag gyorsan hozható létre multimédiás elemeket is tartalmazó interaktív és adaptív tananyag. A H5P előnyei mellett bemutattunk néhány általunk kiemelésre leginkább érdekesnek ítélt H5P modult, tananyagtípust – az interaktív videó elemeit, beállítási lehetőségeit és létrehozásának lépéseit részletezve –, illetve bemutattuk, hogy a Moodle keretrendszerben hogyan tudunk H5P tartalmakat létrehozni. Mindennapi munkák során is fontosnak tartjuk az előremutató és hatékony tananyagtípusok terjedésének elősegítését, emiatt workshopokat is tartunk – az egyetemünk oktatói számára, illetve lehetőség szerint egyéb szakmai rendezvényeken – a H5P tartalomfejlesztő keretrendszerrel. A Debreceni Egyetem oktatói is kezdik egyre inkább felfedezni az ebben rejlő lehetőségeket, és kezdenek egyre gyakrabban H5P tananyagelemet is alkalmazni az e-learning kurzusaikban. Ilyen tananyagelemek például: idővonal és képek sorba rendezése feladat alkalmazása a földtörténeti korok elsajátításának elősegítésére, vagy interaktív videó alkalmazása a videóban elhangzó angol szakmai kifejezések elsajátításának önellenőrzésére. Egyre több szép példát látunk az egyetemi e-learning rendszerben a H5P alkalmazására.

## Irodalomjegyzék

H5P Create and Share Rich HTML5 Content and Applications. <https://h5p.org/> Hozzáférés: 2022. június 20.

H5P - MoodleDocs. <https://docs.moodle.org/39/en/H5P> Hozzáférés: 2022. június 20.

Examples and Downloads | H5P. <https://h5p.org/content-types-and-applications> Hozzáférés: 2022. június 20.

H5P Image Juxtaposition. <https://h5p.org/image-juxtaposition> Hozzáférés: 2022. június 20.

## Munka 4.0 – Ipar 4.0 – Szakképzés 4.0 – : A digitális kompetencia jövőbeni fejlesztési útjai Tamás JANKÓ: Work 4.0 – Industry 4.0 – Vocational Training 4.0 –: Future pathways for the development of digital competence

Jankó Tamás

*Digitális Jólét Nonprofit Kft, Digitális Szakképzési és Felnőttképzési Módszertani Központ  
Digital Success Nonprofit Ltd, Methodological Centre for Digital Vocational Training and Adult Learning*

### Absztrakt

Munka 4.0 – Ipar 4.0 – Szakképzés 4.0 – : A digitális kompetencia jövőbeni fejlesztési útjai

A pandémia alatti kényszerű változások gyors és radikális módon alakították át a munka, a termelés, a szórakozás, és a tanulás világát is. A cikkben a 2020-2021-es friss adatok alapján tekintjük a legújabb trendeket, kihívásokat, és a lehetséges válaszokat is.

**Kulcsszavak:** munka jövője, kompetencia, digitális, világhátrány, munkaerőpiac

### Abstract

Work 4.0 – Industry 4.0 – Vocational Training 4.0 – : Future pathways for the development of digital competence

The forced changes during the pandemic have rapidly and radically transformed the worlds of work, production, entertainment, and learning. In this article, we will look at the latest trends, challenges and possible responses based on recent data for 2020-2021.

**Keywords:** future of work, competences, digital, pandemic, labour market

### I. Ma van az ipar holnapja?

Az elmúlt néhány évben került be az iparági vezetők, HR menedzserek és a képzési-foglalkoztatási szakemberek szótárába az az új talányos kifejezés, amely a „4.0” sorozat egyik első eleme, az „Ipar 4.0” definíció.

Érdeemes visszatekinteni a szókapcsolat születésére, amelyet először a német kormányzat és gazdaság képviselői alakítottak ki és használtak a 2010-es évek elején. A hagyományosan rendkívül erős, német-magyar gazdasági kapcsolatok miatt is lényeges megérteni az Ipar 4.0 kihívás elvárásait, amelyeket a világhátrány közben, és a digitalizáció újabb és újabb hullámhegyei között is szem előtt kell tartani.

Természetesen nem egy véletlennek köszönhetjük azt, hogy a német iparági szereplők fogalmazták meg az alábbi ábra szerinti negyedik ipari forradalom fogalmát, hiszen a világ egyik legfejlettebb, legversenyképesebb gazdasága nem engedhette meg, hogy a globális vezető termelékenységi és innovátori pozíciója megrendüljön.

Az Ipar 4.0 nem önmagában létezik, a megvalósulási helyszíne maga a fejlett termelőüzem (**SMART FACTORY**), amely képes összekapcsolni a szoftverekben megjelenő adatrendszert, amelyek kijelölik például a leendő jármű pontos színét, a fizikai rendszerekkel (pl. egy festőüzemben a tényleges fényező folyamathoz szükséges eszközök, anyagok) és ezek együttesen alkotják a kiber-fizikai rendszereket, (**CYBER -PHYSICAL SYSTEM**) amelyek a termelési folyamatokat megvalósítják. A folyamatok adatain zajlanak, a forgácsoló gépek tengelykopását már nem a TMK csapat vezetője vizsgálja egy kézzel használható tolmérce

segítségével, hanem egy beépített szenzor figyelni ezt, amelyet ipari szintű adathálózatba kapcsolnak, ez az ipari IoT, (INDUSTRIAL IoT) amely a vállalati-üzemi Big Data rendszer alapja.

A hatékony működés érdekében a termelési folyamat több dimenziós voltát, folyamatait, szakaszait, rétegeit felkellett térképezni, a megfelelő számítási, mérési, elemzési automatizációt kialakítani, és a logikai kapcsolatoknak megfelelően, ezt nevezzük megfelelően beépített, beágyazott szemléletű rendszernek (EMBEDDED SYSTEMS), amely a fenti példa szerinti szenzorok segítségével figyelni a rendszer egészét (SMART SENSORS).

Összefoglalva látható, hogy a termelés, az ipar megindult egy rendkívül ígéretes, egyre hatékonyabb működési pályán, a versenyképesség növekedhet, a minőség fejlődik, a hibák csökkennek. Kérdés, hogy ebbe a folyamatba a most dolgozó munkavállalók, az állampolgárok hogyan tudnak részt venni? Milyen feladatokat látnak el az emberek majd 2030-ban?

## II. A munka jövője, a világjárvány és a digitális kompetencia kérdése

A jelenlegi COVID-19 világjárvány által okozott kényszerű lezárások, az emiatt bekövetkező globális recesszió rendkívül bizonytalanná tette a munkaerőpiaci kilátásokat, és felgyorsította az eddig távolinak hitt új trendek megindulását a munka világában. A Világgazdasági Fórum kiadványa, a „Munka Jövője 2020<sup>1</sup>” jelentés a fenti két területet kívánja alaposan feltárni, egyrészt a világjárvány okozta gazdasági törésvonalakat, másrészt a következő öt év tekintetében a technológiai változások új trendjeit, a munkakörök és szükséges kompetenciák változásait kívánja bemutatni.

A bizonytalan körülmények ellenére az elemzés egyedülálló mélységben térképezi fel a jövőbeni munkakörök, készségek, képességek lehetséges változásait, 15 gazdasági szektor és 26 fejlett és fejlődő ország részletes bemutatásával.

A technológiai változás sebessége várhatóan nem fog csökkenni, és néhány területen még növekedhet is. A felhőalapú rendszerek térnyerése, a Big Data és az e-kereskedelem a vezető menedzserek körében továbbra is magas prioritás lesz, hasonlóan a korábbi évekhez. Emellett azonban jelentős növekedés várható a titkosítási rendszerek, a nem-humanoid robotok, és a mesterséges intelligencia területén.

Az automatizáció trendje, a COVID-19 recesszióhoz társulva, „dupla-törést” okozhat a vállalatok és a munkavállalók számára. A cégek át fogják alakítani a feladatrendszerüket, a munkaköröket, a szükséges kompetenciákat 2025-re. A technológiai megfelelés és fejlesztések miatt a munkaerő számának csökkentését tervezi a felmért vállalatok 43%-a, szerződéses alvállalkozókat, munkaerőkölcsönzést kíván igénybe venni további 41%, a fejlett technológiák átvétele miatt a vállalatok 34% kíván új munkaerőt felvenni. 2025-re, a jelenlegi munkafolyamatok időigénye egyenlő módon oszlik majd meg a gépek és a humán erőforrás között.

A munkahelyek számát tekintve, a munkahelyteremtés lassulni fog, míg a munkakörök megszűnése gyorsuló ritmust mutat majd, a munkáltatói becslés alapján 2025-re a felesleges munkakörök száma 15,4%-ról 9%-ra fog csökkenni, és az új trendeknek megfelelő új foglalkozások száma 7,8%-ról 13,5% fog nőni. A bemutatott becslések alapján 85 millió munkahelyet vesznek át a gépek 2025-re, miközben 97 millió új munkakör jön majd létre, amelyek már inkább a gépek-emberek-algoritmusok együttműködése által végrehajtható új típusú elvárásoknak felelnek majd meg.

---

1 <https://www.weforum.org/reports/the-future-of-jobs-report-2020/in-full/country-and-industry-profiles>



A kompetencia hiány továbbra is magas lesz a következő öt év során, a már korábban is ismertek lesznek népszerűek a munkáltatók körében: kritikai gondolkodás, elemzés, probléma megoldás, aktív tanulás, ellenálló képesség mentális téren, stressztűrés, rugalmasság. A képzésekre fordítandó időtartam tekintetében a vállalatok hat hónapos, vagy annál is rövidebb időtartamú képzéseket szerveznének a dolgozók 40%-nak, de általában a munkavállalók 94% -tól várják el, hogy tanuljon, fejlődjön, ami a 2018-as 65% képest meredek növekedés.

A képzés és a kompetenciafejlesztés iránti személyi, szervezeti és szakpolitikai igény nem állítható meg, és rejthető el, az Európai Unió tagállamai által elfogadott Európai Készségfejlesztési Program<sup>2</sup> (2020. július 1.) ambiciózus célokat fogalmazott meg, amely egyaránt hathat a munkaerőpiac képzés mértékére, pótolhatja a hiányzó digitális kompetencia elemeket, és egyben új értelmet adhat a sokak által ismert „Lifelong Learning” kifejezésnek.

Európa nem tud és nem képes fejlődni képzés, továbbképzés és átképzés nélkül, (skilling, upskilling, re-skilling) és ehhez egy új, radikálisan új szemléletre, egy diszruptív megoldásra van szükség, egy sokkal nagyobb mértékű, és intenzívebb képzési rendszerre. Az eredmények ellenőrzése, értékelése és mérése elengedhetetlen, és ehhez a fejlett, digitalizált és digitális eszközök rendelkezésre állnak. Csak ezek a megoldások teszik lehetővé azt, hogy sikeres, versenyképes munkaerő álljon rendelkezésre a dinamikusan fejlődő gazdasági szereplők igényeinek megfelelően.

A Készségfejlesztési Program az alábbi célokat tűzte ki 2025-re, a felnőttkori tanulás elősegítése érdekében, az alacsony képzettségi szint emelése, a digitális kompetencia szint fejlesztése és a munkanélküliség elleni küzdelem támogatása érdekében:

- 2025-re az Unióban 120 millió felnőttnek kell tanulásban részt vennie minden évben. Ez megfelel a felnőtt népesség 50 %-ának, és erre a csoportra vonatkozóan körülbelül 540 millió képzési tevékenységnek az ötéves időszakban.
- 2025-re az Unióban 14 millió alacsony képesítésű felnőttnek kell tanulásban részt vennie minden évben. Ez megfelel a csoport 30 %-ának, és erre a csoportra vonatkozóan körülbelül 60 millió képzési tevékenységnek az ötéves időszakban.
- 2025-re 2 millió álláskeresőnek, vagyis öt álláskeresőből egynek friss tanulási tapasztalattal kell rendelkeznie. Ez erre a csoportra vonatkozóan körülbelül 40 millió tanulási tevékenységnek (40 millió három hónapos kurzus) felel meg az ötéves időszakban.
- 2025-re 230 millió felnőttnek kell rendelkeznie legalább alapfokú digitális készségekkel, ami az Unió felnőtt lakosságának 70 %-át teszi ki.

Mutatók	Céltűzése k 2025-re	Jelenlegi szint (legutóbbi rendelkezésre álló év)	Növekedés (%-ban)
A 25–64 éves felnőttek részvételi aránya a tanulásban az elmúlt 12 hónapban (%-ban) <sup>58</sup>	50 %	38 % (2016)	+32 %
A 25 és 64 év közötti, alacsonyan képzett felnőttek részvétele a tanulásban az elmúlt 12 hónapban (%-ban) <sup>59</sup>	30 %	18 % (2016)	+67 %
A friss tanulási tapasztalattal rendelkező, 25–64 év közötti munkanélküli felnőttek részaránya (%-ban) <sup>60</sup>	20 %	11 % (2019)	+82 %
A legalább alapvető digitális készségekkel rendelkező, 16–74 év közötti felnőttek aránya (%-ban) <sup>61</sup>	70 %	56 % (2019)	+25 %

1. ábra: Az Európai Készségfejlesztési Program (European Skills Agenda) 2025-re kitűzött céljai:

(Forrás: Európai Készségfejlesztési Program,2020:

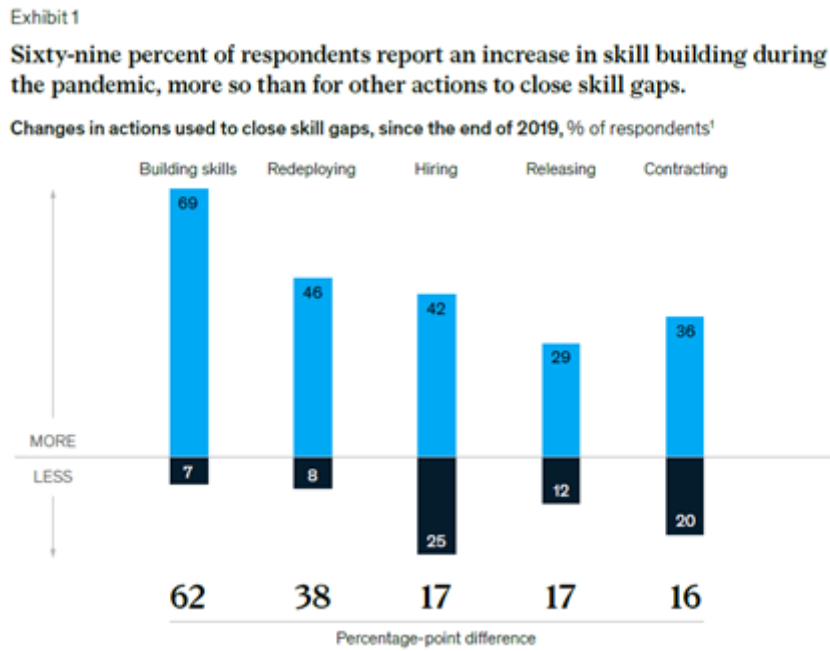
<https://ec.europa.eu/social/main.jsp?catId=1223&langId=en>)

### III. A pandémia által kikényszerített változások a kompetencia elvárások terén

A neves tanácsadó cég, a McKinsey vonatkozó felmérései<sup>3</sup> szerint, a cégvezetők személyügyi-HR politikai reakciói finomodtak, változtak a világjárvány által okozott mély és több ágazatban szinte visszavonhatatlan változásokat előidéző gazdasági válság hatására. A pandémia nyomására új és még újabb trendek, igények, elvárások derültek ki, elég, ha a távmunka által kiváltott új készségekre, a főbb videókonferencia platformok magabiztos ismeretére utalunk. Az alábbi ábra jól érzékelteti, hogy a távmunka, a munkavállalók önállóbb, innovatívabb szerepe és egyben értékei új szemléletet kívánt meg a humán erőforrás fejlesztése terén.

A kényszerű lezárások jelentős változást hoztak a cégvezetők képzésekkel kapcsolatos elképzelései terén, 58% azt állította, hogy a képzési hiányok kezelése fontosabb lett a lezárások közben, és után, mint korábban. A munkahelyi képzések megvalósítása esetében 69% -uk állította, hogy ténylegesen több képzést szerveznek, mint a kényszerű lezárások előtt, összevetve a többi HR kezelési eszközzel (öt HR beavatkozás: munkaerő kölcsönzés, szerződtesítés, újra aktivizálás, elbocsátás, továbbképzés):

3 McKinsey & Company: Building workforce skills at scale to thrive during—and after—the COVID-19 crisis, 2021. április: <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis>



2. ábra: McKinsey-felmérés eredményei (Forrás: McKinsey Global Survey on Reskilling, 2020: <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis>)

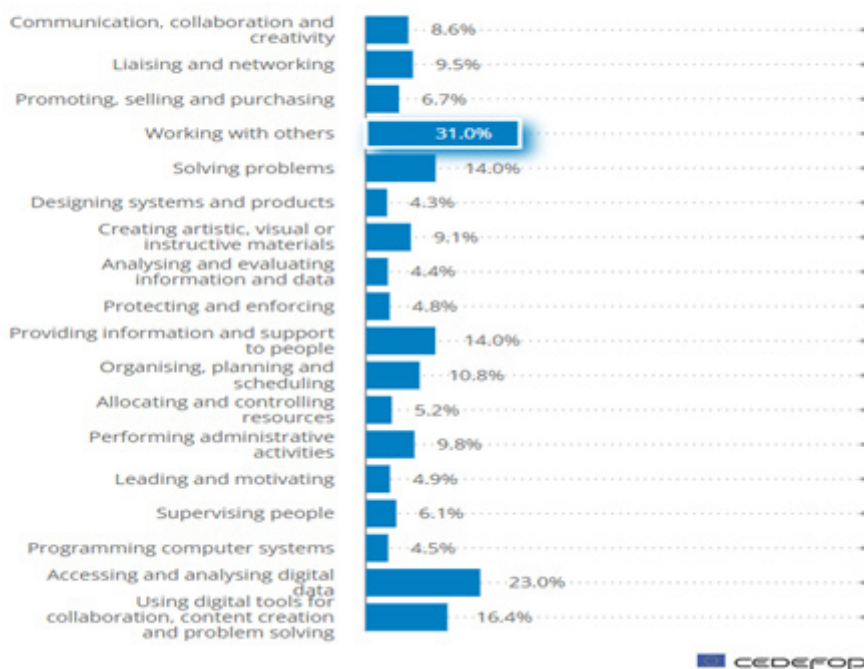
Megállapítható továbbá, hogy a felmérés adatai szerint a képzési kapacitás és hajlandóság növelése nem egyszerű ötlet vagy múltó szeszély, hanem a válaszadók több mint fele eltökélt volt abban, hogy a 2020-as évben többet költ a munkatársak képzésre, mint amennyit 2019-ben költöttek. Összefoglalva megállapítható, hogy az üzleti vállalkozások a lezárások hatására máshogy reagált, azaz újabb lehetőséget kíván adni a munkavállalók számára.

Vajon milyen készségek, kompetenciák hiányozhatnak a pandémiát átélt vállalkozások szerint? Erre a kérdésre kereste a választ az Európai Unió szakképzés fejlesztési központja, a CEDEFOP „Skills Panorama<sup>4</sup>” felmérése, a 2021-ben megjelent online álláshirdetések adatbázis-elemzése alapján:

4 [https://skillspanorama.cedefop.europa.eu/en/dashboard/skills-online-job-advertisements?country=EU27\\_2020](https://skillspanorama.cedefop.europa.eu/en/dashboard/skills-online-job-advertisements?country=EU27_2020)



### Most requested skills in online job ads in EU27 in Skills in 2021 (ESCO level 2)



3. ábra: Skills-OVATE felmérés eredménye, az online álláshirdetésekből elvárt készségek, EU-27, 2021

(Forrás: CEDEFOP: [https://www.cedefop.europa.eu/en/tools/skills-intelligence/skills-online-job-advertisements?country=EU27\\_2020&year=2021&occupation=#1](https://www.cedefop.europa.eu/en/tools/skills-intelligence/skills-online-job-advertisements?country=EU27_2020&year=2021&occupation=#1))

A csapattagok, munkavállalók, beosztottak és vezetők közötti együttműködés 31,0%-os elvárása érte el a legmagasabb szintet, azaz a munkaadók felismerték, hogy akár az otthoni konyhaasztal, akár a munkahelyi iroda, de az együttműködési készség nélkül nem lesz megoldás, nem lesz innováció, nem lesz bevétel.

#### IV. Képzettség, eszközök és lehetőségek-2022-2030

Az ENSZ tripartit foglalkoztatásügyi szervezete, az ILO 114 országra kiterjedő felméréssel<sup>5</sup> vizsgálta a COVID-19 járvány alatti munkaerőpiaci képzések helyzetét, és megállapította, hogy a felmérésben résztvevő szervezetek, vállalatok szerint, a kényszerű munkaszünet alatti képzések esetében 41,8% arányban kihívást jelentett a képzésben résztvevők digitális kompetencia jártassági szintje.

Az együttműködés új színtere a digitális felület, ezért a rövid összefoglaló végén érdemes kiemelni az egyéni és munkavállalói digitális kompetencia jártassági szintet feltáró DESI index<sup>6</sup> humántőke – tudásszint eredményeit, amelyek rámutatnak arra, hogy mekkora szakadék a kereslet-kínálat, a vállalati elvárások és az átlagos munkavállaló digitális kompetencia szintje között:

A 2019-es adatokat tekintve (DESI 2020 jelentés<sup>7</sup>) fejlődés mutatkozott mind az alapvető, mind a fejlett informatikai tudás (IKT szakon végzettség száma és IKT szakemberek száma) tekintetében. 2015 és 2019 között az alapvető digitális tudással rendelkező állampolgárok

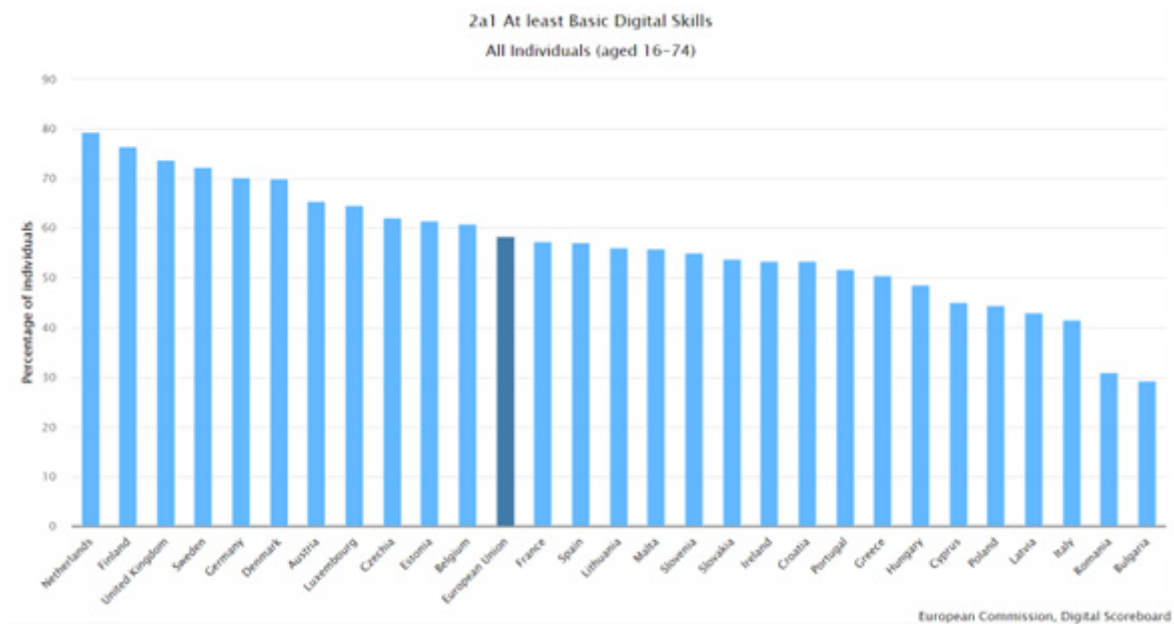
5 [https://www.ilo.org/skills/Whatsnew/WCMS\\_743243/lang--en/index.htm](https://www.ilo.org/skills/Whatsnew/WCMS_743243/lang--en/index.htm)

6 <https://digital-strategy.ec.europa.eu/en/policies/desi>

7 <https://digital-strategy.ec.europa.eu/en/policies/desi>

száma három százalékkal nőtt, 55%-ról 58%-ra, de így is igen nagy azok száma, akik nem lesznek képesek megfelelni egy egyre növekvő digitális versenynek.

Az IKT szakemberek számát tekintve 9,1 millió ember dolgozott informatikusként az Európai Unióban, amely 1,6 millió fővel több, mint négy évvel korábban. Az informatikai szakemberek utáni felfokozott kereslet azonban továbbra is magas szintű, a nagyvállalatok 64%, a KKV-ék 56% vélekedett úgy, hogy továbbra is nehéz betölteni az üres informatikai álláshelyeket, 2018-ban. A humán tőke fejlettségi szint a DESI számai alapján Svédországban, Finnországban és Észtországban a legmagasabb.



4. ábra Alapvető digitális készségek, 16-74 év, EU 27

(Forrás: DESI: <https://digital-strategy.ec.europa.eu/en/policies/desi-hungary>)

Az alapvető digitális kompetencia elemei tehát -a fenti adatok alapján- az európai átlagot tekintve 41,7% arányban hiányoznak az átlagos állampolgár tudásából, és ez a tényező hatott a COVID-19 járvány alatt, sokszor akár az alapvető szükségletek ellátása, pl. a védőoltásra való jelentkezés, vagy egy recept kiváltása esetében is. Ezt a hiányt kell felszámolni, a szükséges képzéseket, támogatásokat meg kell szervezni, a digitálisan magabiztos állampolgárok számának növelése érdekében.

Az európai szinten elfogadott „Digitális Évtized” szakpolitikai program<sup>8</sup> a 41,7% kompetenciahiány mérséklése érdekében 2030-as dátummal a következő célokat fogadta el: az IKT szakemberek számát 20 millióra kell növelni, továbbá a nemek közötti kiegyensúlyozottság megteremtése érdekében a női IKT szakemberek számát növelni kell, a digitális alapkészségek terén pedig az európai lakosok legalább 80%-át megfelelő szintre kell juttatni.

8 [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030\\_hu](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_hu)

Összefoglalva elmondhatjuk, hogy a járvány okozta világválság megrengette a képzési piacot, elvárásokat, trendeket, új lehetőségeket nyitott fel, és végérvényesen lezárt folyamatokat. A munkaerőpiaci képzések fontossága soha nem volt ilyen magas szintű, az együttműködési készség, képesség iránti munkaerőpiaci elvárás kimagasló, és mindehhez- az Európai Készségfejlesztési Programnak megfelelően – a felnőttkori tanulás új eszközeit, beleértve a digitális kompetencia fejlesztésre vonatkozó eszközöket kell kialakítani, és a megfelelő hálózati együttműködés keretében hatékonyan kell működtetni ezeket.

## Felhasznált irodalom

1. **EURÓPAI KÉSZSÉGFEJLESZTÉSI PROGRAM:** Európai Bizottság, COM (2020) 274: A fenntartható versenyképességre, a társadalmi méltányosságra és a rezilienciára vonatkozó európai készségfejlesztési program <https://ec.europa.eu/social/main.jsp?langId=en&catId=1223>
2. **THE FUTURE OF JOBS REPORT 2020:** World Economic Forum, 20 October 2020: <https://www.weforum.org/reports/the-future-of-jobs-report-2020#report-nav>
3. **BUILDING WORKFORCE SKILLS AT SCALE TO THRIVE DURING—AND AFTER—THE COVID-19 CRISIS** April 30, 2021 | McKinsey Survey <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis>
4. **ONLINE JOB VACANCIES AND SKILLS ANALYSIS: CEDEFOP, SKILLS-OVATE EU-27:** [https://www.cedefop.europa.eu/en/tools/skills-intelligence/skills-online-job-advertisements?country=EU27\\_2020&occupation=&year=2021#1](https://www.cedefop.europa.eu/en/tools/skills-intelligence/skills-online-job-advertisements?country=EU27_2020&occupation=&year=2021#1)
5. **THE DIGITAL ECONOMY AND SOCIETY INDEX (DESI):** <https://digital-strategy.ec.europa.eu/en/policies/desi>

## Megújuló könyvtári statisztika: az egységes adatstruktúra és a korszerű megjelenítés kialakításának útján

Békésiné Bognár Noémi Erika

*Országos Széchényi Könyvtár, Könyvtári Intézet, Kutatási és Elemző Osztály*  
[bognar.noemi@oszk.hu](mailto:bognar.noemi@oszk.hu)

Nagy Andor

*Országos Széchényi Könyvtár, Könyvtári Intézet, Kutatási és Elemző Osztály*  
[nagy.andor@oszk.hu](mailto:nagy.andor@oszk.hu)

### Absztrakt

Az Országos Statisztikai Adatfelvételi Program keretében az ország területén működő könyvtárak és könyvtári szolgáltató helyek évről évre értékes adatokkal járulnak hozzá az országos könyvtárstatisztikai adatbázishoz. A „Jelentés a könyvtárak tevékenységéről” című OSAP 1442 adatlap változásainak, valamint az adatrögzítésre szolgáló informatikai rendszer fejlesztéseinek következtében az összeállított adattáblák összehasonlíthatósága sokáig rendkívül korlátozott volt, és azok elemzése, értelmezése nehézségbe ütközhetett. Ennek okán a Könyvtári Intézet Kutatási és Elemző Osztály kollégái a 2021-es évben 2015-ig visszamenőlegesen átfogó adattisztítást végeztek, és kialakítottak egy olyan egységes adatstruktúrát, amelynek révén a 2015 és 2020 közötti évek adatai elemi szinten is könnyen összehasonlíthatóvá váltak. A munka eredményeképp, a részletes szakmai, tudományos elemzésekhez elérhetővé váltak Excelben és SPSS-ben is jól kezelhető adatállományok. Az adatok könnyű és gyors áttekinthetősége érdekében korszerű, adatvizualizációs eszközök segítségével, webes felületen válik megtekinthetővé a Magyarországon működő könyvtári szolgáltatási pontok számának, működésének, szolgáltatásainak és használatának alakulása. Ezenfelül a Kutatási és Elemző Osztály munkatársai egy új elektronikus kiadványsorozatot indítottak, a Könyvtári trendjelentéseket, amely jelenleg tizenkét tematikus riportból áll, bemutatva az elmúlt években, a hazai könyvtári rendszerben bekövetkezett főbb változásokat.

**Kulcsszavak:** könyvtári statisztika, adatvizualizáció, könyvtári trend

**Title:** Renewable library statistics: on the way to creating a unified data structure and modern visualization

### Abstract

Every year, Hungarian libraries contribute valuable data to the National Library Statistics database through the National Statistical Data Collection Programme. Due to ongoing changes of the OSAP 1442 data sheet „Report on the activities of libraries” the comparability of the compiled datatables was limited and hard for a long time, so the analysis and interpretation was difficult too. For this reason, the colleagues of the Research and Analysis Department of the National Széchényi Library, Hungarian Library Institute carried out a comprehensive data cleaning project in 2021, where they drilled down until 2015, and created a uniformed data structure that made the data from 2015 to 2020 easily comparable at a basic level. As a result of the work, the new data sets can be easily managed in Excel and SPSS applications and have been made available for detailed professional and scientific analysis.

In order to make the data easily and quickly accessible, the most important performance indicators like trends of visitors, services and usage of the libraries in Hungary will be made available on a web interface using Microsoft Power BI data visualization tool. In addition, the Research and Analysis Department have launched new electronic publication series called Library Trends Reports, which currently consists of twelve thematic reports presenting the main changes of the Hungarian library system during recent years.

**Keywords:** library statistics, data visualization, library trend

## 1. Bevezetés

A Könyvtári Intézet elődjével, a Könyvtártudományi és Módszertani Központtal egyetemben 1995 óta a könyvtárstatisztikai adatszolgáltatás menedzselője, tehát az intézet szervezi az adatgyűjtést és teszi közzé az eredményeket. Az évek során nemcsak a törvényileg meghatározott adatszolgáltatók köre változott, de maga az adatgyűjtő adatlap is számos alkalommal módosult. Kezdetben az adatszolgáltatók papír alapon juttatták el az Intézethez az éves statisztikájukat, 2008-tól viszont már az Emberi Erőforrások Minisztériuma által fejlesztett számítógépes rendszerben.

Az adatszolgáltató felület teljes megújítására 2020-ban került sor. Ez egy nagyon jelentős mérföldkő a statisztikai adatszolgáltatás történetében, ugyanis az új Kulturális Statisztikai Adatgyűjtő Rendszer indulása révén először vált lehetségessé olyan automatikusan működő adatlapszintű ellenőrzések megvalósítása, amelyek nemcsak a kitöltőket segítik a munkájukban, de általuk sokkal pontosabb adatokat is kapunk, hiszen már a kitöltés során kapnak azonnali visszajelzést a beírt adatok vonatkozásában. Ezen felül, az adatszolgáltatók köréről is realisabb képet kapunk az új rendszer indulása óta azáltal, hogy minden adatszolgáltató önálló regisztrációval rendelkezik a felületen, és definiáltak az egyes intézmények közötti kapcsolatok, valamint a többszereplős ellenőrzési folyamatnak köszönhetően az egyes intézmények alapadatairól is realisabb képünk lett.

## 2. A könyvtári statisztika megújításának folyamata

A könyvtári statisztikai adatlap, valamint az egyes adatelemekhez tartozó kitöltési útmutató rendszeres változása miatt sok adat idősorosan nem értelmezhető, és egyes adatelemek, adatcsoportok pontosításra szorulnak. 2021-ben a Könyvtári Intézet Kutatási és Elemző Osztályán elkezdük a könyvtári statisztika teljes körű megújítását, annak érdekében, hogy az adatokat a kor elvárásainak megfelelő módon egy átlátható, webes felületen tudjuk megjeleníteni és szűrhetővé tenni.

Első lépésként a korábbi évek önálló, struktúrájában nem egységes Excel-táblázataiból készítettünk el egy adatbázist, jórészt az SPSS nevű statisztikai szoftver segítségével. Mivel korábban az egyes adatelemeknek nem volt fix azonosítójuk, és nem is voltak egyértelműen megfeleltethetőek egymással (jórészt a statisztikai adatlap többszöri változása miatt), létre kellett hoznunk mindazon fix változóneveket, amelyek alá évenként be tudtuk sorolni az azokhoz tartozó adatokat. Ez egy rendkívül időigényes, sok esetben manuális munka volt. Évenként közel 700 változóval kell dolgoznunk, változónként megközelítőleg 4000 egyéni könyvtár adatával, tehát évenként 2,8 millió cellával. A könyvtárak adatai idősorosan érdemben 2015-től hasonlíthatók össze, így a munkánk során egy közel 16 millió cellára épülő adatbázist hoztunk létre, amely a 2015 és 2020 közötti időszakot fedi le.



Az egységes idősoros adatbázis létrehozását követően a következő lépés ennek a 16 millió adatelemnek az ellenőrzése és javítása volt, annak érdekében, hogy a statisztika a korábbinál sokkal pontosabb és egységesebb legyen. Mivel az adatlapszintű, automatikusan végrehajtott ellenőrzést csak a 2020-ban indult új Rendszer tette lehetővé, így az általunk vizsgált adathalmazban számtalan olyan hibát kellett javítanunk, amely elsősorban abból ered, hogy a korszerű technikai feltételek híján az adatszolgáltatók kontroll és valódi támpontok nélkül szolgáltatott adatot.

Az adattisztítást számos részfeladatra bontottuk, a következőkben ezek közül említünk meg néhányat. Egyrészt eltávolítottuk azokat az adatelemeket, amelyeket azon adatszolgáltatók küldtek be, akiknek nem is kellett volna kitölteniük az adott adatmezőket, mivel vagy nem rájuk vonatkoztak, vagy bizonyos feltételek nem teljesültek. Ezek a tévesen megadott adatok torzították az összesített könyvtári statisztikát.

Megvizsgáltuk továbbá az egyes adatelemek közötti összefüggéseket és korrigáltuk a következetlenségeket. Ha pl. valamely adatszolgáltató több nyilvánosan használható számítógépet jelentett le, mint ahány számítógépe összesen volt, akkor korrigáltuk a helytelen értéket. Ezeken felül megvizsgáltuk a kiugró, extrém értékeket, amelyek vagy félreértésből vagy elírásból adódtak. Jellemző tévesztés volt pl. több esetben, amikor egy nullával több került a megadott szám mögé vagy az „ezer Ft-ban” megadandó értéket teljes formájában adták meg. Felülvizsgáltuk még az egyes intézmények besorolását is, így megszüntettük pl. az „egyéb” intézményi kategóriát és manuális munkával besoroltuk a korábban ide tartozó intézményeket újonnan létrehozott, részletesebb típusokba. Az itt felsoroltakon felül is számos olyan javítást végeztünk, amelyek célja, hogy az eddigieknél is pontosabb képet kapjunk a könyvtári rendszer szereplőinek tevékenységeiről, használati adatairól.

Célunk volt továbbá az, hogy a jövőben jóval kevesebb utólagos javítást kelljen végeznünk a beszolgáltatók által szolgáltatott adatokon, és ezt többféle módon is megpróbáltuk elérni. Egyrészt úgy kíséreltük meg a 2022 januárjában induló statisztikai adatgyűjtést gördülékenyebbé és pontosabbá tenni, hogy leszűrve a 2020-ban indult statisztikai megújítási folyamat tanulságait, számos, immár több, mint 420 ellenőrzőképletet építettünk be közvetlenül az elektronikus adatszolgáltató felületen elérhető statisztikai adatlapba. Ezek egy része olyan ellenőrzéseket takar, amely az egyértelműen téves információk beküldését gátolja meg, egy részük pedig a potenciálisan téves információkra hívja fel a beküldők figyelmét. Mindkét esetben figyelemfelhívó üzenetek tájékoztatnak a valós vagy vélt hibákról, és pontosan megmutatják, hogy az adatlapon belül mely sor tartalmazza azt. Az egyértelmű hibákat piros színnel jelöli a rendszer, és javításuk nélkül nem is engedi beküldeni az adatlapot, a lehetséges hibákat pedig sárgával, és inkább támpontul szolgálnak a kitöltőknek abban, hogy mire érdemes figyelni az adatok megadása során.

Ezenfelül azzal is próbáltuk segíteni az egységesebb országos könyvtárstatisztikai adatszolgáltatás létrejöttét, hogy az adatlaphoz készítettünk egy, a korábbinál sokkal részletesebb és pontosabb kitöltési útmutatót. Erre azért volt szükség, mert egyes adatok gyűjtése vagy kiszámítása nem minden esetben egyértelmű, ezért nagyon fontos, hogy az egyes adatszolgáltatók biztosan ugyanazt értsék a kérdések alatt, és azonos módszertant használva állapítsák meg a saját intézményükre vonatkozó adatokat. Az új kitöltési útmutatót nemcsak egy önálló, letölthető fájlban tettük közzé, hanem az elektronikus adatszolgáltatási felületen is beágyaztuk az egyes mezőkhöz, az azokhoz tartozó útmutatásokat.

Az elvégzett fejlesztések eredményeképp sokat egyszerűsödött az adatszolgáltatók munkája, és egységesebbé is váltak az összesített adatok.

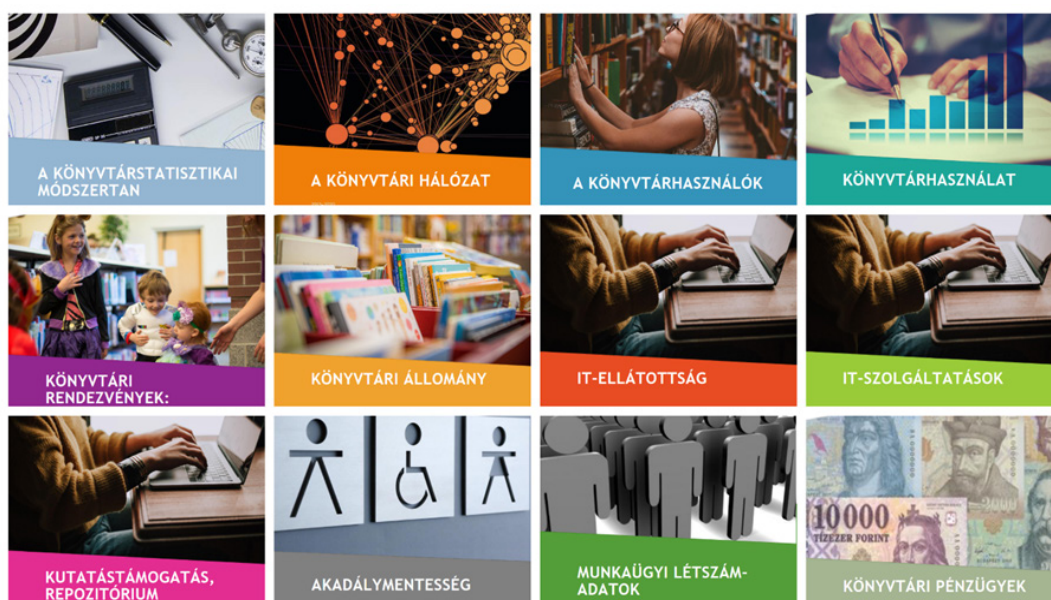


### 3. A könyvtárstatisztikai adatok bemutatása

Céljaink között nemcsak az új adatbázis összeállítása, az adatok tisztítása és a jövőbeni adatszolgáltatás adatminőségének javítása szerepelt, hanem a létrejött, harmonizált adatok különböző módon történő felhasználása és bemutatása is az érdeklődők számára. A könyvtárakról szóló adatok szemléltetését két új szolgáltatással kívántuk megvalósítani. Az egyik a Könyvtári trendjelentések című kiadványsorozat, a másik pedig az országos könyvtárstatisztikai adatok bemutatása adatvizualizációs eszközökkel. Amellett, hogy kutatók, szakdolgozók számára továbbra is elérhetők közvetlenül az adatok táblázatos formában is, szeretnénk volna ezzel a két új szolgáltatással jobbra tenni az adatok áttekinthetőségét.

#### 3.1 Könyvtári trendjelentések<sup>1</sup>

2021 végén az idősoros adatállomány felhasználásával a 2015 és 2020 közötti időszakról egy kiadványsorozatot indítottunk. Ennek keretében 2021-ben tizenkét tematikus összefoglaló készült, amelynek címe Könyvtári trendjelentések.



1. ábra Könyvtári trendjelentések című kiadványsorozat első tizenkét száma

Célja, hogy témakörönként mutassa be és foglalja össze a hazai könyvtári rendszerben bekövetkezett főbb változásokat, trendeket. A trendjelentések tartalmazzák a könyvtári rendszer jellemzőit, információkat a munkatársakról, a költségvetésről, az állományról, az IT-infrastruktúráról és -szolgáltatásokról, valamint az akadálymentességre vonatkozó helyzetről is. Külön trendjelentés foglalkozik a könyvtárakban szervezett programokkal, a kutatástámogatási tevékenységgel, a könyvtárhasználókkal és a könyvtárhasználati adatokkal.

<sup>1</sup> Könyvtári Intézet: Könyvtári trendjelentések. Hozzáférés: 2022. 06. 20. <https://ki.oszk.hu/informacioszolgalatas/konyvtari-trendjelentések>

### 3.2 Az országos könyvtári statisztikai adatok bemutatása adatvizualizációs eszközökkel<sup>2</sup>

A másik új szolgáltatásunk szintén az országos könyvtári statisztikai adatokat kívánja bemutatni, azonban egy dinamikusabb, személyre szabhatóbb módon, adatvizualizációs eszközök segítségével. E szolgáltatás kialakításával könnyen áttekinthetővé kívántuk tenni a Magyarországon működő könyvtári szolgáltatási pontok számának, működésének, szolgáltatásainak és használatának 2015 és 2020 közötti alakulását (összhangban a Könyvtári trendjelentésekkel témákban és arculatban) mind az ágazati irányítás, a könyvtárvezetők, a könyvtárosok és más érdeklődők számára. A rendszer kialakításánál törekedtünk arra, hogy a jövőbeni beérkező adatok könnyen beemelhetővé váljanak.

A lekérézéshez kattintson az alábbi ikonok valamelyikére:



2. ábra A vizualizált könyvtárstatisztikai adatok témakörei

A legfontosabb adatokat tizenegy témakörbe gyűjtve, idősorosan ismertetjük külön aloldalakon. Ahol nem áll rendelkezésre 2015-től adat vagy az adatlap időközi változása miatt idősorosan nem ábrázolható egységesen a teljes időszak, ott a diagramok szűkebb időtartamot fednek le. A látogatók a szűrők segítségével minden diagramot ugyanazon szempontok (év, funkció, fenntartó típusa, nyilvánosság, gazdálkodás és szervezet jellege, területi adatok: megye, település, településméret) szerint tudnak megtekinteni, leszűrni. Minden, az adott jelentésben szereplő vizualizáció, változik a szűrők beállítása után.

<sup>2</sup> Könyvtári Intézet: Az országos könyvtári statisztikai adatok bemutatása adatvizualizációs eszközökkel. Hozzáférés: 2020. 06. 20. <https://ki.oszk.hu/adatvizualizacio>

### Szűrési lehetőségek

Év

Az összes kijelölése

2015

2016

2017

2018

2019

2020

Funkció

Mind

Részletes funkció

Mind

Fenntartó típusa

Mind

Nyilvánosság

Mind

Gazdálkodás jellege

Mind

Szervezet jellege

Mind

Megye

Mind

Település

Mind

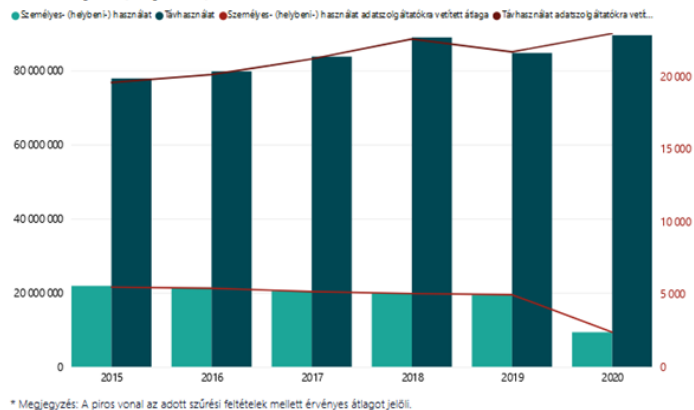
Településméret

Mind

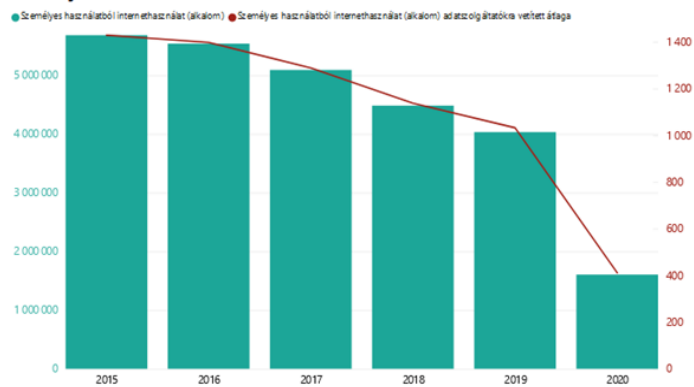
A könyvtári szolgáltatási pont neve

Mind

### A személyes- (helybeni-) és a távhasználat alakulása



### Személyes használatból internethasználat



3. ábra A könyvtárhasználatról szóló aloldal, a dinamikusan változó diagramokkal

A vizualizációkon megtekinthetők az adott változó egyedi és összesített adata, és ahol a szemléltetés úgy kívánta, a diagramok tartalmazznak az adott szűrési feltételek mellett érvényes, adatszolgáltatókra vetített átlagot is. Ezzel a funkcióval lehetővé válik az összehasonlítás az azonos intézményi körhöz tartozó könyvtári szolgáltatási pontok adatai között. Így például, ha egy városi könyvtár vezetője a saját intézményének beküldött adatait kívánja összevetni hasonló könyvtárak eredményeivel, akkor először könnyen visszanezheti a saját maga által beküldött adatokat idősorosan, majd lehetősége van megtekinteni bármely más könyvtár egyedi adatait vagy a részletes funkció szűrő használatával kizárólag a városi könyvtárakra vonatkozó adatokat leszűrni, valamint akár leszűkíteni a találatokat a hozzá hasonló méretű városok könyvtáira a településméret szerint. Ezután ennek a kiválasztott körnek az átlagát a saját eredményeihez tudja hasonlítani.

Az adatvizualizációs rendszer és a téma aloldalak kialakításához a Microsoft üzleti intelligencia szolgáltatását, a Power BI-t használtuk. Ez a rendszer széles körben elterjedt, könnyen használható és jó kompatibilitást biztosít számos adatbázis-típussal (pl. Excel, SPSS, SQL). Részletes adatelemzésre képes, nagy adathalmazból lehet jelentéseket készíteni, különböző típusú, az adott adatsorhoz leginkább illő grafikonokat kiválasztani és alkalmazni. Mivel ezt a tevékenységet a Kutatási és Elemző Osztály több munkatársa is végzi, fontos volt, hogy lehetőséget biztosít a rendszer a közös munkára ugyanazonokon az adatokon és jelentéseken. A szolgáltatás kialakításakor a publikus megosztási lehetőség lehetővé tette, hogy az elkészített jelentéseinket a Könyvtári Intézet oldalára be tudtuk ágyazni.

A jövőbeni terveink között szerepel a szolgáltatás továbbfejlesztése egyrészt a jövőbeni adatok beemelésével, másrészt a használat segítségével (további információk, leírások elhelyezése a felületen).

Kéziratoss dallamlejegyzések feldolgozása MI-vel támogatott digitális környezetben

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

## AI-supported Processing of Handwritten Transcriptions for Hungarian Folk Songs in a Digital Environment

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

My research focuses on creating an AI-supported Digital Research Environment (DRE) that helps analyzing and systematizing folk music tunes with the help of the latest information theory and database management results. The study may be extended to the entire source material accumulated by researchers integrating Hungarian ethnomusicology results of the last hundred years. In this way, new dimensions of structural analysis open and a large amount of information can be processed that already exceeds the limits of human musical memory.

Previous computerized music analysis experiments in Hungary have inappropriately defined the role of artificial intelligence. In our case, the AI-supported digital environment that is the subject of the research does not work independently, because of the researcher's scientifically abstract thinking, preferences, and recognizing characteristic melodic elements cannot yet be replaced by computer data processing.

Crucial goal of the research is to precisely define the researcher's role in the musical data processing. Thus the attitude of the researcher who rejects software support may be moderated. For the first time in Hungarian folk music research history, a detailed and documented digital research environment can be created, integrating the useful, relevant software tools. We can map out data entry problems and define the standard format of the musical data suitable for mass input and analysis. If possible, we will replace the previously widely used optional data with scalable data to have a broader range of parametrization and search options, and their free combination allows us to study new scientific models. With DRE, the validity range of the previous scientific musical classification can be more precisely specified and the processing as well as classification of unreported melodies and the process of type creating can be significantly accelerated.

The most significant debate in the previous research has been the dataset specification of analyses. I am convinced that only similarly processed tune-data-elements can be compared, so one of the most critical tasks is determining the input data's standard format and information density. As a first step, the digital conversion of the musical manuscript needs to be solved. International research has mainly led to results in the recognition of printed music, some of which can be used in the project, but many new developments are also needed.

**Keywords:** AI-supported Digital Research Environment (DRE), Optical Music Recognition (OMR), Musical Manuscripts, Hungarian Folk Songs, scientific musical classification, ethnomusicology, digital archives, folklore database.

**Tárgyszavak:** optikai zenefelismerés, népzenei dallamrendezés, magyar népdal, népzene kutatás, digitális archívum, folklór adatbázis, MI-támogatás, digitális kutatási környezet

## 1. Bevezetés<sup>1</sup>

A hároméves kutatási program fókuszában a népzenei dallamok elemzését és rendszerezését segítő, mesterséges intelligenciával támogatott digitális környezet kialakítása áll a magyar népzene kutatás tapasztalatai, valamint az információelmélet és adatbáziskezelés legújabb eredményeinek segítségével. Ebben a cikkben a 2021 végén indult program legfontosabb kérdésvetéseit és első eredményeit mutatom be.<sup>2</sup> Ez a kutatástámogató szoftverkörnyezet eszközként digitális dallamadatbázist és MI-t használ, a koncepciót pedig a zenetudomány adja. Integrálva az elmúlt száz év kutatási eredményeit, a vizsgálat kiterjeszhető az elődök által felhalmozott teljes forrásanyagra. Ilyen módon – nagyságrendeket lépve – olyan nagy mennyiségű információt lehet feldolgozni, amely már meghaladja az emberi zenei memória határait.

Az első, már valóban modern módszertannal dolgozó népzene kutatók nagy érdeklődéssel fordultak a dallamok rendezésével kapcsolatos kérdések felé Európa szerte. Mindez jól illeszkedett a 19. és a 20. század fordulójának uralkodó szellemi áramlataiba: a magyar kutatók kezén összegyűlt ezres nagyságrendű dallamadat már elegendő volt az érvényes vizsgálatokhoz és egyúttal kielégítette a pozitívizmus ténytiszteletét és rendszerezési igényét is. Ezt egészítette ki az evolucionizmus elméletének hatása, amely az általános emberi kultúra szerves részét képező zene kutatásában is tetten érhető; egymást követő korszakokat, szakaszosan végbemenő változásokat rajzolt a magyar népzene történet üres lapjaira.

A rendezések elsődleges – és gyakorlati – célja a hatalmasra duzzadt papíralapú gyűjtemények nyilvántartása volt, hiszen azok csak így válhattak hatékony kutatási segédanyaggá. Első népzene kutatóink a kutatási tevékenység mellett magasan kvalifikált zenészek is voltak, így a szükséges nyilvántartási elvek kialakítása mellett hamar megjelent egy összefoglaló tudományos koncepció igénye is. Ezt követően a két tevékenység – vagyis a nyilvántartás és a rendszeralkotás – párhuzamosan futott; a kutatói habitus, a fókusz és a munkabírás határozta meg e tudományos konstrukciók készülségi fokát és utóéletét.

A dallamrendezés alapelveinek megtalálásához először el kellett szakadni a nyelvtudományok módszertanától. A századfordulón nyilvánvalónak tűnt, hogy a dallamok rendezését a szöveges analógiák alapján lehet elvégezni. Ez a gondolati korlát abból adódott, hogy – köszönhetően a több évszázados könyvtári gyakorlatnak – sokkal nagyobb tapasztalat állt rendelkezésre a szövegek és a szöveges információk rendezése terén. Az ábécé-sorrend analógiájára megalkotott dallamkezdő hangok magasság szerinti rendezése azonban nem

---

1 A cikk a Debrecenben megrendezett NETWORKSHOP 2022 országos informatikai konferencián 2022. április 22-én elhangzott előadás leírt és szerkesztett változata. Az előadáshoz tartozó prezentáció megtekinthető itt: <https://prezi.com/view/4WH6mWduD8DuUHwiwKOp/>

2 Bolya Mátyás, *Népzenei dallamrendezés MI-vel támogatott digitális környezetben*, „OTKA” posztdoktori kiválósági program (PD\_21), projekt azonosító: 137969.





volt sikeres: tapasztalati úton kiderült, hogy a dallamok eleje egyszerűen nem ad karakteres mintázatot, így nem alkalmazható csoportképző elemként.<sup>3</sup>

A megoldás röviden: a népdalok túlnyomó többsége sorokba rendeződik, a sorvégző hangok relatív elhelyezkedését leíró számsor pedig olyan egyedi azonosító, amely már alkalmas nagyobb tömbök kialakítására és rokonsági viszonyok megállapítására. Ehhez a dallamokat a könnyebb összehasonlíthatóság kedvéért közös alaphangra transzponálták és a módszertan fejlődése során további, itt nem részletezett vizsgálati szempontot és technikai rendezőelvet vezettek be a szótagszámtól a ritmikai mintázatokon keresztül egészen a dallamvonalak általánosításáig.

## 2. Előzmények

A kutatást közvetlenül megalapozó eredményeket önálló kötetben publikáltam *Információelmélet és népzene kutatás* címmel.<sup>4</sup> A könyvvel lezárt kutatási szakaszban megvizsgáltam a magyar népzene rendezésének legfontosabb állomásait az információelmélet segítségével, összegyűjtöttem a dallamelemzés szempontjait, csoportosítottam ezeket, majd az így kialakított egységes szempontrendszer szerint megvizsgáltam a korábbi rendezési elveket, végül felírtam a rendezési szekvenciákat. A különböző tudományos modellek így összehasonlíthatóvá váltak, világosan feltárultak az alkotói szándék hangsúlyai és világossá vált az is, hogy a vizsgált rendszerek a rendszeralkotás mely fázisában tartanak.

Kutatási programom előzményének tekinthető több adatbázis is,<sup>5</sup> de a legfontosabb a *Zenetudományi Intézet Hangarchívum*,<sup>6</sup> amelynek 2021-ben már a negyedik fejlesztési üteme valósult meg. A hangarchívum nem csupán előzmény, hanem a kutatás elsődleges forrásadatbázisa is. Az ingyenesen elérhető hatalmas adatbázis<sup>7</sup> zenei leíró adatokat is tartalmaz, sok dimenziós hivatkozási rendszere a metaadatokat a hangfelvételekkel, a lejegyzésekkel, a papíralapú dokumentációval, valamint geokódokkal és idő kódokkal kapcsolja össze.

3 A zenei szempontú rendezés igényét jól mutatja egy 1900-ban kiírt pályázat; a felhívás fordítása: „Mi a legjobb módszer a népdalok és népies dalok zenei (és nem szöveges) jellemzőik szerinti, szótárszerű rendezésére?” A kétféle szemlélet pedig tökéletesen tükröződik a két beérkezett pályázatban. A bécsi zenetörténész, Oswald Koller a szótári betűrend mintájára a dallamok kezdőhangjait tette a rendezés alapjává. A másik pályázó a finn Ilmari Krohn volt, akinek ötletét Bartók Béla és Kodály Zoltán fejlesztette tovább és alkalmazta a magyar népdalok rendezésére. Scheurleer, Daniel Francois, „Welches ist die beste Methode, um Volk- und volksmäßige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen?“, *Zeitschrift der Internationalen Musikgesellschaft* 1. sz. (1899–1900): 219–220.

4 Bolya Mátyás, *Információelmélet és népzene kutatás: Rendszeralkotás, nyilvántartás, digitális archívum*. (Budapest: L'Harmattan, MTA BTK Zenetudományi Intézet, 2019).

5 Bolya Mátyás, Magyar citerás antológia, Liszt Ferenc Zeneművészeti Egyetem, 2014, [www.mca.lfze.hu](http://www.mca.lfze.hu); Bolya Mátyás, Both Miklós és Kukár Manó. *Ethiofolk: Online etióp népzene- és néptáncarchívum*, MTA BTK Zenetudományi Intézet, Polyphony Project, 2019. [www.ethiofolk.com/](http://www.ethiofolk.com/); Both Miklós et al. Polyphony Project. Internetes népzenei archívum. 2018. Online archiválási rendszer ukrán népzenei gyűjtések feldolgozására és publikálására. [www.polyphonyproject.com](http://www.polyphonyproject.com); Bolya Mátyás, Both Miklós et al. *Folk\_ME. Folk Music Education for Future Generations. Creative folk music educational toolkit*. Polyphony, 2021. <https://www.folk-me.com/hu>. It is an innovative education project, which targets public education. The project's main output is a newly developed web-based digital toolkit, which can integrate traditions into a wider context creatively in high school education, by which the project intends to give a solution to save disappearing European folk heritage inspired by the Polyphony project's research and online archive. Partners: Sibelius Academy – University of Arts Helsinki, Lviv National Musical Academy named after Mykola Lysenko, and RCH Institute for Musicology (Budapest).

6 Bolya Mátyás, szerk., *Zenetudományi Intézet Hangarchívum*. ELKH BTK Zenetudományi Intézet, 2021. <https://zti.hungaricana.hu/hu/>

7 Körülbelül 200 000 dallam, amely 12 000 óra hangfelvételt jelent.



A hazaival szemben a nemzetközi szakirodalom sokkal gazdagabb. A külföldön már hagyományosnak tekinthető digitális bölcsészet részének tekintik a digitális muzikológiát.<sup>8</sup> Ennek ellenére kettősség figyelhető meg ezen a tudományterületen: miközben az elméleti kutatás számos eredményével találkozhatunk projektekben és szócikkekben,<sup>9</sup> addig ezek az eredmények ritkán épülnek be a nagy adatbázisok kereső és elemző moduljaiba.<sup>10</sup>

### 3. A kutatás célja

A dallamrendszerezés tudományos modellalkotása terén jobbra néhány kiemelkedő képességű kutató zenei memóriájára voltunk utalva és leginkább empirikus úton történt a zenei típusalkotás. Bartók, Kodály, Járdányi, Dobszay és Bereczky sok ezer dallam áttanulmányozása során „megérzett” mintázatok alapján építették fel rendszereiket és alapozták meg a méltán világhírű magyar analitikus népzenei rendszerezés tudományát. Kutatásaik archívumi háttéréül szolgáló, a Zenetudományi Intézetben található közel 250 000 dallamot számláló Központi Népzenei Gyűjtemény átfogó vizsgálata azonban meghaladja az emberi memória határait.

Ma a számítógépes adatfeldolgozás kapacitása és a folyamatosan fejlődő technológia együttesen nyújthat segítséget a módszertani továbblépéshez. Ezáltal lehetőségünk nyílik a dallamrendezés kiterjesztésére, valamint a korábbi kutatói közgondolkodást erősen meghatározó tudományos struktúráktól független, kizárólag a leíró adatok alapján történő halmazvizsgálatra, statisztikai elemzésre.

Úgy gondolom, hogy az ismertetett kutatási program a digitális kutatási környezet kialakítása révén képes integrálni az elmúlt száz év magyar népzenei kutatási eredményeit, és a vizsgálat így kiterjeszthető az elődök által felhalmozott teljes forrásanyagra. Ilyen módon a strukturális elemzés új dimenziói nyílnak meg és olyan nagy mennyiségű információt lehet feldolgozni, amely már meghaladja az ember zenei memóriájának határait. Az eddigi számítógépes dallamelemzési kísérletek nem megfelelően becsülték meg a mesterséges intelligencia szerepét. Esetünkben a kutatás tárgyát képező, MI-vel támogatott digitális segédeszköz nem önállóan dolgozik, hiszen a kutatók tudományosan absztrakt gondolkodását, súlyozását, illetve a meghatározó dallamalkotó-elemek azonosítását ma még nem válthatjuk ki számítógépes adatfeldolgozással.

---

8 IMS Study Group “Digital Musicology” <https://www.musicology.org/networks/sg/digital-musicology>; Collections of articles on Frontiers: <https://www.frontiersin.org/journals/digital-humanities/sections/digital-musicology>

9 Blot, Guillaume et al. *Melody and Rhythm Through Network Visualization Techniques*. 2017. [https://www.researchgate.net/publication/319863548\\_Melody\\_and\\_Rhythm\\_Through\\_Network\\_Visualization\\_Techniques](https://www.researchgate.net/publication/319863548_Melody_and_Rhythm_Through_Network_Visualization_Techniques); Paulus, Jouni et al. Audio-based music structure analysis. 2010. [https://www.researchgate.net/publication/289662964\\_Audio-based\\_music\\_structure\\_analysis](https://www.researchgate.net/publication/289662964_Audio-based_music_structure_analysis); West, Kris et al. *The Networked Environment for Music Analysis (NEMA)*. 2010. [https://www.researchgate.net/publication/220985364\\_The\\_Networked\\_Environment\\_for\\_Music\\_Analysis\\_NEMA](https://www.researchgate.net/publication/220985364_The_Networked_Environment_for_Music_Analysis_NEMA); De Roure, David et al. *SALAMI: Structural Analysis of Large Amounts of Music Information*. 2010. [https://www.researchgate.net/publication/44023837\\_SALAMI\\_Structural\\_Analysis\\_of\\_Large\\_Amounts\\_of\\_Music\\_Information](https://www.researchgate.net/publication/44023837_SALAMI_Structural_Analysis_of_Large_Amounts_of_Music_Information); AHRC Research Centre for the History and Analysis of Recorded Music (CHARM). 2009. <https://charm.rhul.ac.uk/about/about.html> etc.

10 Robert Glaubitz, *The Aria Database*, 1996–2013. <http://www.aria-database.com/>; Giles Bergel et al. *Broadside Ballads Online from the Bodleian Libraries*. <http://ballads.bodleian.ox.ac.uk/>; Eerola, Tuomas, and Petri Toiviainen. *Digital Archive of Finnish Folk Tunes*. Database, 2004. [http://esavelmat.jyu.fi/index\\_en.html](http://esavelmat.jyu.fi/index_en.html); *British Library Sounds: World & traditional music*. 2020. <https://sounds.bl.uk/World-and-traditional-music/>

#### 4. A feldolgozás lépései

Fontos célkitűzés pontosan meghatározni a kutató helyét a dallamadat-feldolgozás folyamatában, ezzel mérséklődhet a szoftveres támogatást elutasító kutatói attitűd is. A hazai kutatástörténetben elsőként készülhet egy részletesen dokumentált digitális kutatástámogató környezet, integrálva a népzene kutatás számára hasznosítható szoftveres eszközöket. Feltérképezzük az adatbevitel problémáit, meghatározzuk a dallamadat szabványos, tömeges bevitelre és elemzésre alkalmas formátumát. Az eddig használt opciós leíró adatokat lehetőség szerint skálázható adatokkal helyettesítjük, így szélesebb paraméterezési és keresési lehetőségek állnak rendelkezésünkre, szabad kombinációjuk pedig eddig nem vizsgált tudományos modellek kipróbálását teszi lehetővé. Az eszköz segítségével pontosabban behatárolható a korábbi tudományos dallamrendezések érvényességi tartománya is, illetve jelentősen felgyorsítható a beosztatlan dallamok feldolgozása és besorolása, a típusalkotás munkája.

Az eddigi kísérletek során a legnagyobb vitát kiváltó tényező az elemzések adatbázisa volt. Meggyőződésem, hogy csak hasonló feldolgozottságú dallamadatbázis-elemeket lehet összehasonlítani, ezért az egyik legfontosabb feladat meghatározni a bemeneti adatok szabványos formáját és információsűrűségét. A gépi intelligencia jelenleg nem tudja felismerni egy egyszerűsített dallamváz és egy más szótagszámú, gazdagon díszített dallamváltozat tudományos részletességű lejegyzése közötti összefüggést. Egy ilyen dallamrokonság felismerése még a szakmán belül is gyakran vitákat vált ki, amely azt jelzi, hogy a modellalkotásban fontos szerepe van a kutatói – jó értelemben vett és tudományosan indokolt – szubjektivitásnak. Ezzel párhuzamosan feltérképezzük az adatbevitel problémáit és meghatározzuk a dallamadat (vagyis idő-hangmagasság függvény) szabványos, tömeges bevitelre és elemzésre alkalmas formátumát.

Első lépésként a kéziratos dallamlejegyzések digitális konverzióját kell megoldani. A nemzetközi kutatások leginkább a nyomtatott kották felismerésében hoztak eredményeket, ezek egy része hasznosítható a projekt során, azonban számos új fejlesztésre is szükség van. Ezt követi majd a dallamváz megállapítása – ebben segíthet a mesterséges intelligencia – és a dallamadatbázis létrehozása, végül az összehasonlító elemzés digitális környezetének kialakítása. Ezekről a fázisokról később fogok beszámolni.

## 5. Az optikai kottafelismerés

A nyomtatott kották felismerése, vagyis digitális képformátum konvertálása digitális hanginformációvá (Optical Music Recognition)<sup>11</sup> egy dinamikus fejlődő fejlesztési terület ma már számos, piacon beszerezhető szoftvermegoldással. A zene leegyszerűsítve egy időhangmagasság függvény, ennek képi ábrázolása a kotta. Fontos kiemelni, hogy a hangértéket kifejező szimbólumok grafikai elhelyezése nem feleltethető meg a hangértéknek, nem arányos azzal. Ennek oka az olvashatóság, ugyanis túlságosan széteső grafikai képet kapnánk, ha betartanánk az időegységeket is. Tehát a függvényként értelmezhető zene gyakorlati ábrázolása nem felel meg a matematikai szabályoknak, ez fontos elem lesz a kódolásnál.



1. ábra. A kottafejek időérték szerinti elosztása (Time Signature Spacing)



2. ábra. A kottafejek optikai elosztása (Note Spacing)

A kották felismerésének folyamatát önkéntelenül a szövegfelismeréshez hasonlítjuk, azonban sok lényeges különbség van. Bár az általános kottairás jelrendszere – nem számolva a hangszerspecifikus jeleket – viszonylag kevés szimbólumból áll, ezek elhelyezése és kombinációja meghatározza az értelmezést. Tehát a zeneírásnak speciális nyelvtana van, amely nélkül lehetetlen a dekódolás. Ezzel szemben a karakterfelismerés nem lép túl a betűk és a szavak azonosításán, vagyis az alapvonal referenciájához igazodó egydimenziós információolvasásról van szó.

De mi a helyzet a kéziratos kottákkal? A zenei információ dekódolása összetett feladat nem csak az algoritmusok, hanem egy zenész számára is. Ebből következik, hogy a kéziratos kották szükségszerűen pontosabbak, mint egy kézírásos szöveg. Egy kiolvashatatlan szó nem akadályozza a szöveg további értelmezését, a nyelvismeret és a kontextus automatikus

11 Bainbridge, David és Bell, Tim, „The Challenge of Optical Music Recognition.” *Computers and the Humanities* 35, 95–121 (2001). <https://doi.org/10.1023/A:1002485918032>; Rebelo, Ana, Fujinaga, Ichiro, Paszkiewicz, Filipe et al. „Optical music recognition: state-of-the-art and open issues.” *International Journal of Multimedia Information Retrieval* 1, (2012): 173–190. <https://doi.org/10.1007/s13735-012-0004-6>



javítási folyamatot indít el az olvasóban (tehát nem feltétlenül a dekódoló program végzi). A kotta viszont szemantika nélkül nem értelmes. Ha a zenész például nem tudja eldönteni egy hangról, hogy vonalra vagy vonalközbe esik, akkor megszakad a kódolás folyamata. Kétségtelen azonban, hogy a kéziratos kották esetén gyakran találkozhatunk az írást gyorsító egyszerűsítésekkel és a nyelvtani szabályok megsértésével is.

## 6. A kéziratok feldolgozásánál jelentkező problémák

A magyar népzene kutatás elmúlt 120 évében közel 200 000, különböző részletességű kéziratos lejegyzés készült. Miért fontos ezek digitalizálása akkor, amikor a szoftveres hangfelismerés és -feldolgozás szinte követhetetlen tempóban fejlődik? A kérdés jogos és biztos, hogy a kutatás egy későbbi fázisában e két adatgyűjtési módszer találkozni fog. Ezek a lejegyzések olyan időkapuszulák, amelyekben megtalálhatjuk az egykori kutatók tudományosan absztrakt gondolkodását. E kéziratok feldolgozásával kutatói generációk pótolhatatlan tapasztalatát csatornázzhatjuk be az adatelemzésbe.

Ezzel el is jutottunk az első gyakorlati feladathoz, vagyis az MI és az emberi erőforrás megfelelő arányának megállapításához a feldolgozás során. Mivel a kéziratok szerkezete és jelölései nem egységesek, ezért bizonyos feladatok elvégzése nem bízható gépi algoritmusokra. Pontosabban fogalmazva a fejlesztési ráfordítás nem lenne arányos a várható eredménnyel. Ilyen, az elemzések szempontjából kulcsfontosságú elem például a dallam tagolásának felismerése, a sorok azonosítása, a rövidítések dekódolása, illetve az elemzési terület meghatározása. Több ezer kéziratos lejegyzés tanulmányozása után a következő tipikus, a dekódolást nehezítő tényezőt sikerült feltérképezni: 1) az egyszerűsítések különböző fajtái, amelyek egyrészt a takarékos papírhasználat, másrészt a gyorsabb munkatempó miatt terjedtek el; 2) több dallamadat egy képen; 3) a hasznos információt tartalmazó terület azonosítása.<sup>12</sup>

## 7. Tervek

Megállapíthatjuk, hogy az eddigi számítógépes dallamelemzési kísérletek nem jól határozták meg a kutatók és a mesterséges intelligencia szerepét és ez bizalmatlanságot szült a népzene kutatók részéről az efféle kezdeményezésekkel szemben. Reményeink szerint kutatásunk feloldja ezt az ellentmondást, hiszen megmarad a kutatói kontroll, és mindenki szabadon kipróbálhatja az eszközt már a fejlesztés fázisában is.

A kutatásnak köszönhetően határozottabb vonal rajzolható az MI és a kutatói szerepvállalás átmeneti területére, illetve lehetőség nyílik az ember zenei memóriáját messze túlszárnyaló elemzésekre. Az eredmények kiterjesztik a magyar népzene kutatás hatósugarát a nem strofikus anyag és a hangszeres zene területére is. A kialakított rendszer bármilyen népzenei adatbázis feldolgozására alkalmas lesz, nemzetközi együttműködések keretében pedig különböző zenekultúrák is összehasonlíthatóvá válnak. A számítógéppel a strukturális elemzés új dimenziói nyílnak meg és nagy mennyiségű információt lehet feldolgozni.

A kutatók képzésében is jól hasznosulhatnak az eredmények, hiszen a felület használatával elsajátítható a magyar népzene kutatás méltán elismert analitikus gondolkodásmódja. A tapasztalatok általánosításával a kutatás később kiterjeszthető a más műfajokban fellelhető nagyobb zenei egységek vizsgálatára, illetve interdiszciplináris együttműködések felé is. Hosszabb távon az eszköz nemzetközi érdeklődésre is számíthat, hiszen ehhez hasonló, komplex szoftveres dallamelemzési támogatás jelenleg nincs a piacon.

<sup>12</sup> Példák a *Függelék*ben találhatók.

## 8. Függelék

A függelékben közölt kéziratok azonosításához szükséges rövidítések feloldása:

AP – A Zenetudományi Intézet Népzenei Archívumában található egyedi gyártású hanglemezek azonosítója, az *Akadémiai Pyral* rövidítése. A rövidítés a gyűjtemény jogtulajdonosára (MTA), valamint az alapanyagot és technológiát biztosító francia *Pyral* cégre utal. Az archívumi gyakorlat szerint a technikusok a népzenei vagy néptáncgyűjtésekkor az eredetileg magnetofonszalagra vett zenét speciális lakklemezre másolták. A kutatók e hanglemezeket – úgynevezett nyúzópéldányokat – használták a felvételek lejegyzéséhez, publikációikban pedig rendszerint az AP-számot adták meg forrásként.

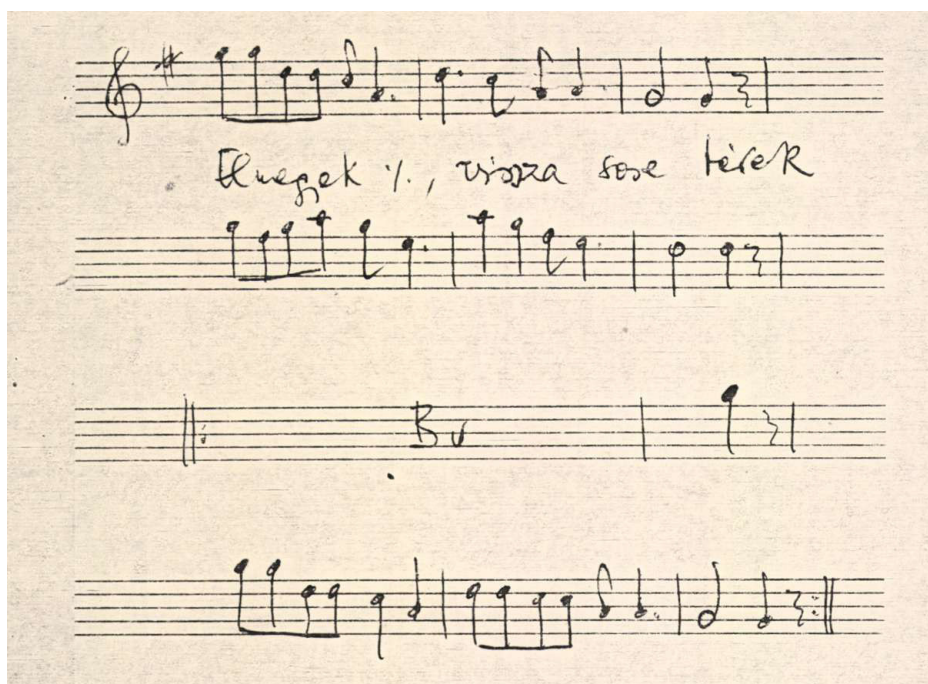
BR – prefix a Bartók-rend lejegyzéseinek leltári számához

D – Webster-drót (mágneses hanghordozó)

Mg – magnószalag (mágneses hanghordozó)

MH – a Néprajzi Múzeum (Budapest) tulajdonában levő fonográf-henger („múzeumi henger”)

ZTI – Zenetudományi Intézet (Budapest)

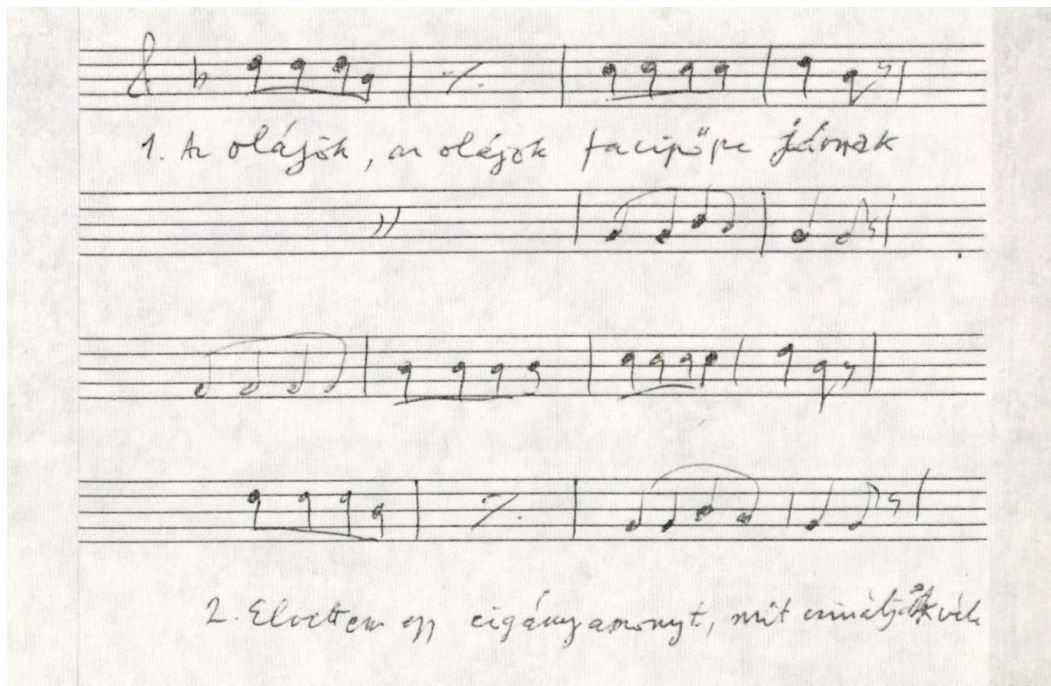


3. ábra. Egyszerűsítések 1. Variáns esetén csak az eltérés jelölése

Archívumi ID: ZTI\_AP\_01011a; hangfelvétel: ZTI\_D\_215-016

[https://library.hungaricana.hu/hu/view/ZTI\\_AP\\_01011-01036/?pg=0&layout=s](https://library.hungaricana.hu/hu/view/ZTI_AP_01011-01036/?pg=0&layout=s)

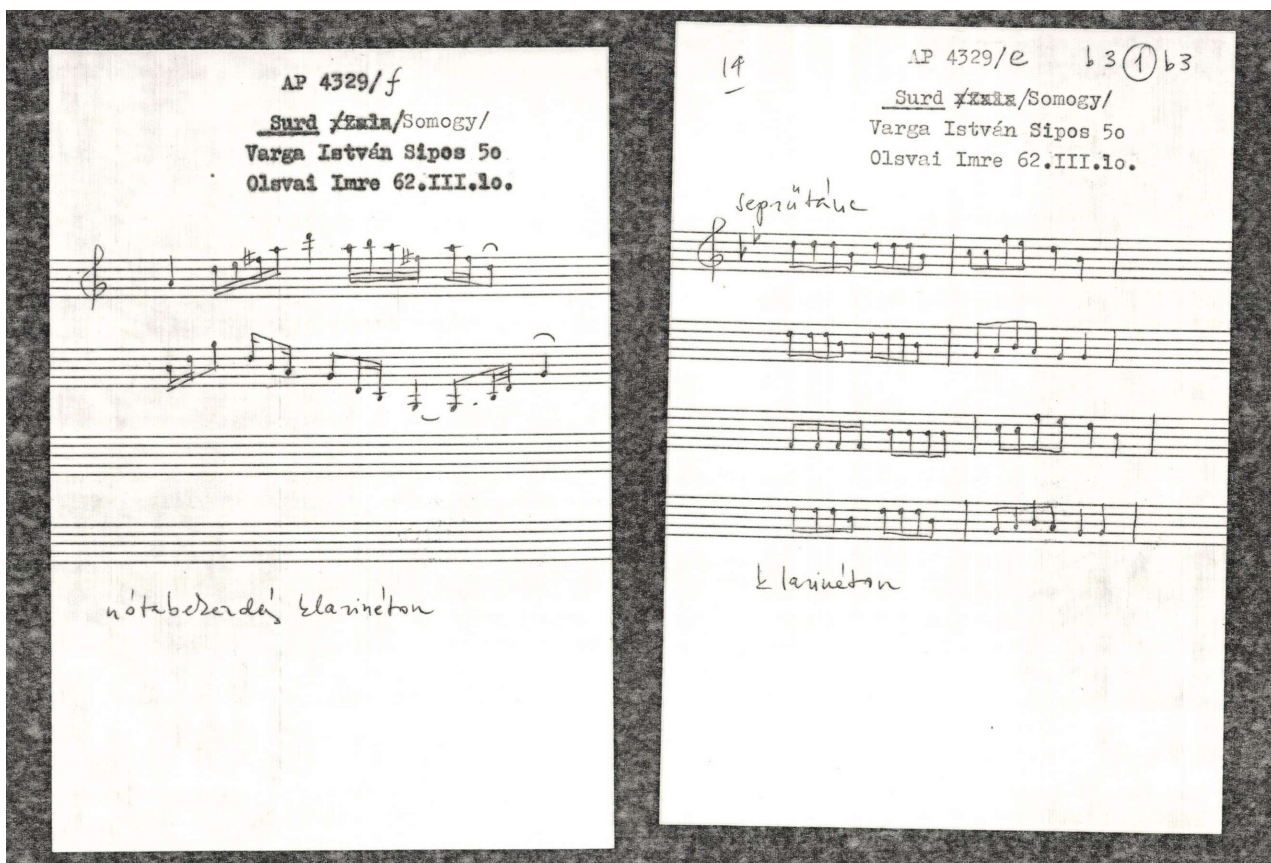




4. ábra. Egyszerűsítések 2. Ismétlődő részek jelölése

Archívumi ID: ZTI\_AP\_08663f; hangfelvétel: ZTI\_Mg\_02806

[https://library.hungaricana.hu/hu/view/ZTI\\_AP\\_08649-08680/?pg=206&layout=s](https://library.hungaricana.hu/hu/view/ZTI_AP_08649-08680/?pg=206&layout=s)



5. ábra. Több adat egy képen

Archívumi ID: ZTI\_AP\_04329e-f; hangfelvétel: ZTI\_Mg\_01276A

[https://library.hungaricana.hu/hu/view/ZTI\\_AP\\_04301-04330/?pg=400&layout=s](https://library.hungaricana.hu/hu/view/ZTI_AP_04301-04330/?pg=400&layout=s)



$\frac{8}{1-7}$ 
1274.
D(1)E

C II (1)<sup>13</sup>

M. H. 2418 a)

Hiskundororsma (Szongrád) 1930.  
 Bálint Gyámsóné sz. Lajkó Mária 44.  
 gy: Lajtha László.

3. üstös  
 1=116.

1=120.

1. Mit keres mi-nalunk a csőr? 2. Megkínálom bir én apjuk,  
 Anyjok ennek többlet ne főzz, Mert mi egymást ritkán látjuk;  
 Még ne kinsáld jó boromból, Ha kend kimen a tanjára,  
 Sem a lovát abrakomból, Elgyjn hozzám éca-kára.

6. ábra. Egy eredeti kézirat feldolgozás előtt  
 Archivumi ID: BR\_03943; hangfelvétel: MH\_2418a  
<http://systems.zti.hu/br/hu/search/3969?cod=2418a>



8  
 1-7

1274.

C II (1)  
13

D(1) E

M. H. 2418 a)

Kiskunbörzsoma (Kongrad) 1930.  
Bálint Jánosné sz. Lajkó Mária 44.  
gy: Lajtha László.

3/8-as  
1=116.

1. Mit keres mi-malunk a csőr? 2. Megkínálom bir én apjuk,

1=120

Anyjok ennek többet ne főzz, Mert mi egymást ritkán látjuk;

1=132.

1=126.

Meg ne kinald jó boromból, Ha kend kimen a tanyára,

1=120

Sem a lovát abrakomból. Elgyjn. horra. éca-kita.

7. ábra. Előkészítő elemzés a gépi feldolgozáshoz

6. 8. 8, 2060 b *Vagy Károly* 5. 1. 1  
 1-8 (2.)  
 M. F.: 1133 &  
 Pócsy (Seren) Bozka Pál Garay

a) 1. Ej, haj, áubri pajtás, mére lánmár ve u-da-ás? Ej, haj,  
 a-la-fe-li, a le-re-és er-dő fe-li.  
 a) 2. vr.  
 a) 3. vr.  
 b) 1. vr.

2. Ej, haj, áubri pajtás,  
 Micsoda jár af fehér har.  
 Ej, haj, dengyel város,  
 Ott lakik a Komisaros.

3. Ej, haj, áubri pajtás,  
 Elloptak a pey-aripát.  
 Ej, haj, csak hádd vigyék,  
 Van még ammén-gabén eleg.

4. Ej, haj, áubri pajtás,  
 Elloptak a cifra sulát.  
 Ej, haj, csak hádd vigyék,  
 Van még a rücsöknél eleg.

b) 1. Rivkomáromi erdöbe  
 Áubri Józka fegy ricsibe.  
 Harmincháron nemöt mént Ri,  
 Áubri hármát mind megfogá.

2. Rivkomáromi kert alatt,  
 Áubri Józka majd ott maradt,  
 Cifra saurbe gyöcsgályába,  
 Erüst sárkányus cipismába.

(csak a tota Barabás-é) 599/a var. u. ot (K. d. gy.) 60a

8. ábra. Egy elrettentő példa Bartók legendás zöld tintás javításaival  
 Archivumi ID: BR\_01194; hangfelvétel: MH\_1502a  
<http://systems.zti.hu/br/hu/search/1194?cod=1502a>





## Forrásjegyzék

- AHRC Research Centre for the History and Analysis of Recorded Music (CHARM). 2009. Hozzáférés: 2022.06.20. <https://charm.rhul.ac.uk/about/about.html>
- Bainbridge, David, és Bell, Tim. „The Challenge of Optical Music Recognition.” *Computers and the Humanities* 35, (2001): 95–121. Hozzáférés: 2022.06.20. <https://doi.org/10.1023/A:1002485918032>
- Blot, Guillaume et al. *Melody and Rhythm Through Network Visualization Techniques*. 2017. Hozzáférés: 2022.06.20. [https://www.researchgate.net/publication/319863548\\_Melody\\_and\\_Rhythm\\_Through\\_Network\\_Visualization\\_Techniques](https://www.researchgate.net/publication/319863548_Melody_and_Rhythm_Through_Network_Visualization_Techniques)
- Bolya Mátyás, Both Miklós és Kukár Manó. *Ethiofolk*. Online etióp népzene- és néptáncarchívum. MTA BTK Zenetudományi Intézet, Polyphony Project, 2019. [www.ethiofolk.com/](http://www.ethiofolk.com/)
- Bolya Mátyás, Both Miklós et al. *Folk\_ME*. *Folk Music Education for Future Generations*. Creative folk music educational toolkit. Polyphony, 2021. Hozzáférés: 2022.06.20. <https://www.folk-me.com/hu>
- Bolya Mátyás. *Információelmélet és népzene kutatás*. Rendszeralkotás, nyilvántartás, digitális archívum. Budapest: L'Harmattan, MTA BTK Zenetudományi Intézet, 2019.
- Bolya Mátyás. *Kéziratos dallamjegyzések feldolgozása MI-vel támogatott digitális környezetben*. 2022. Prezentáció. <https://prezi.com/view/4WH6mWduD8DuUHyywKOp/>
- Bolya Mátyás. *Magyar citerás antológia*. Liszt Ferenc Zeneművészeti Egyetem, 2014. [www.mca.lfze.hu](http://www.mca.lfze.hu)
- Bolya Mátyás. *Népzenei dallamrendezés MI-vel támogatott digitális környezetben*. „OTKA” posztdoktori kiválósági program (PD\_21), projekt azonosító: 137969. <https://nkfih.gov.hu/palyazoknak/nkfi-alap/tamogatott-projektek-pd21>
- Bolya Mátyás, szerk. *Zenetudományi Intézet Hangarchívum*. ELKH BTK Zenetudományi Intézet, 2021. Hozzáférés: 2022.06.20. <https://zti.hungaricana.hu/hu/>
- Both Miklós et al. *Polyphony Project*. *Internetes népzenei archívum*. 2018. Online archiválási rendszer ukrán népzenei gyűjtések feldolgozására és publikálására. [www.polyphonyproject.com](http://www.polyphonyproject.com).
- British Library Sounds*. *World & traditional music*. 2020. Hozzáférés: 2022.06.20. <https://sounds.bl.uk/World-and-traditional-music/>
- Collections of articles on Frontiers. Hozzáférés: 2022.06.20. <https://www.frontiersin.org/journals/digital-humanities/sections/digital-musicology>
- De Roure, David et al. *SALAMI: Structural Analysis of Large Amounts of Music Information*. 2010. Hozzáférés: 2022.06.20. [https://www.researchgate.net/publication/44023837\\_SALAMI\\_Structural\\_Analysis\\_of\\_Large\\_Amounts\\_of\\_Music\\_Information](https://www.researchgate.net/publication/44023837_SALAMI_Structural_Analysis_of_Large_Amounts_of_Music_Information)
- Eerola, Tuomas, and Petri Toiviainen. *Digital Archive of Finnish Folk Tunes*. Database, 2004. Hozzáférés: 2022.06.20. [http://esavelmat.jyu.fi/index\\_en.html](http://esavelmat.jyu.fi/index_en.html)
- Giles Bergel et al. *Broadside Ballads Online from the Bodleian Libraries*. Hozzáférés: 2022.06.20. <http://ballads.bodleian.ox.ac.uk/>
- IMS Study Group “Digital Musicology”. Hozzáférés: 2022.06.20. <https://www.musicology.org/networks/sg/digital-musicology>

- Paulus, Jouni et al. *Audio-based music structure analysis*. 2010. Hozzáférés: 2022.06.20.  
[https://www.researchgate.net/publication/289662964\\_Audio-based\\_music\\_structure\\_analysis](https://www.researchgate.net/publication/289662964_Audio-based_music_structure_analysis)
- Rebelo, Ana, Fujinaga, Ichiro, Paszkiewicz, Filipe, R. S. Marcal, Andre, Guedes, Carlos és S. Cardoso, Jaime. „Optical music recognition: state-of-the-art and open issues.” *International Journal of Multimedia Information Retrieval* 1, (2012): 173–190. Hozzáférés: 2022.06.20.  
<https://doi.org/10.1007/s13735-012-0004-6>
- Robert Glaubitz. *The Aria Database*. 1996–2013. Hozzáférés: 2022.06.20.  
<http://www.aria-database.com/>
- Scheurleer, Daniel Francois. „Welches ist die beste Methode, um Volk- und volksmäßige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen?“, *Zeitschrift der Internationalen Musikgesellschaft* 1. sz. (1899–1900): 219–220.  
<https://archive.org/details/ZeitschriftDerInternationalenMusikgesellschaft011899-1900/page/n221/mode/2up>
- West, Kris et al. *The Networked Environment for Music Analysis (NEMA)*. 2010. Hozzáférés: 2022.06.20.  
[https://www.researchgate.net/publication/220985364\\_The\\_Networked\\_Environment\\_for\\_Music\\_Analysis\\_NEMA](https://www.researchgate.net/publication/220985364_The_Networked_Environment_for_Music_Analysis_NEMA)

A cikk a 137969 számú projekt keretein belül (*Népzenei dallamrendezés MI-vel támogatott digitális környezetben*), az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, az „OTKA” posztdoktori kiválósági pályázati program (PD\_21) finanszírozásában valósult meg.

## Régi magyar verskorpusz összeállítása stilometriai és számítógépes metrikai kutatásokhoz<sup>1</sup>

Maróthy Szilvia

ELTE BTK Magyar Irodalom- és Kultúratudományi Intézet OTKA 135631 kutatócsoport  
[mthy.szilvi@gmail.com](mailto:mthy.szilvi@gmail.com)

ORCID: [0000-0003-2558-9504](https://orcid.org/0000-0003-2558-9504)

Seláf Levente

ELTE BTK Magyar Irodalom- és Kultúratudományi Intézet  
[selaf.levente@btk.elte.hu](mailto:selaf.levente@btk.elte.hu)

ORCID: [0000-0002-6052-9841](https://orcid.org/0000-0002-6052-9841)

Vigyikán Villó

ELTE BTK Magyar Irodalom- és Kultúratudományi Intézet OTKA 135631 kutatócsoport

The project “Computerized metrical and stylometric study of early modern Hungarian poetry” (OTKA, 2020–23) aims to computationally process and analyse Hungarian historical songs (epic poetry from the 16–17<sup>th</sup> century). Our paper introduces the methodology of building a corpus of 174 poems (length between 50 and 1000 strophes). The corpus mostly contains OCRd and corrected texts from printed critical editions, but born digital scholarly editions (in PDF, HTML, XML, DOC formats) are also represented. We had to make further changes to the texts to make them suitable for NLP tools: create modernized transcription, encode hiatus or non-metrical paratexts (titles, arguments). The research is assisted by the *Répertoire de la poésie hongroise ancienne* (<https://f-book.com/rpha/>) database, which serves the literary historical and metrical metadata for the analysed texts. During the mentioned OTKA research project a new version of the database is being developed, and content update is in progress.

Keywords: stichometry, stylometry, literary corpus, epic song, digitization

### Bevezetés

A 2020-ban indult „A régi magyar költészet számítógépes metrikai és stilometriai vizsgálata” kutatási projekt három területre fókuszál: a *Répertoire de la poésie hongroise ancienne* adatbázis fejlesztésére; a 16. század második felében kibontakozó szerelmi költészetre; valamint a 16. század egyik legnépszerűbb műfajára, a históriás énekekre.

A *Répertoire de la poésie hongroise ancienne*,<sup>2</sup> azaz a régi magyarversek adatbázisa (továbbiakban RPHA) egy 1976 óta mind technológiájában (először lyukkártyás, majd sorban nagygépes, CDS/ISIS, webes verziók), mind tartalmában fejlődő kutatói adatbázis, mely a kezdetektől az 1600-ig tartó időszak magyar nyelvű verseinek és azok forrásainak jegyzékét, részletes irodalomtörténeti és poétikai leírását adja strukturált rendszerben. A jelenleg zajló OTKA projektben készül az adatbázis hetedik verziója, mely számos webtechnológiai újításon túl

- 1 Az 135631 számú projekt az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, az OTKA-K pályázati program finanszírozásában valósult meg.
- 2 Horváth Iván és mtsai., „Répertoire de la poésie hongroise ancienne, v. 7.0”, 1979. 2022, <https://f-book.com/rpha/v7/>.



jelentős tartalmi gazdagodáson is keresztülmegy, a versek metaadatai mellé nagy számban felkerülnek azok szövegei is. Az RPHA emellett a POSTDATA nemzetközi verstörténeti adatbázis projekthez is kapcsolódik, melynek keretében mind az RPHA metaadat állománya, mind az adatbázisban már elérhető szövegek lekérdezhetők lesznek a POSTDATA közös felületein is, és az ahhoz kialakított ontológia alapján kerülnek leírásra.<sup>3</sup>

A kutatásnak több része van. Az egyik a kora újkori magyar szerelmi, s különösen a felsőbb (ún. arisztokratikus) regiszterbe sorolt udvari szerelmi líra kibontakozása, mely elsősorban Balassi Bálint és követőinek költészetéhez fűződik. Kutatásunkban a líra és az epika nyelvi és metrikai kapcsolódási pontjait vizsgáljuk a szerelmi költészetben. Milyen művek hathattak Balassi költészetére, s milyen a Balassi-versek utóélete lírában és epikában (a szerelmi tárgyú históriákban)? Stilometriai és számítógépes metrikai eszközökkel finomíthatjuk-e a Balassi-követők és Balassi műveinek kapcsolatáról alkotott képünket?

A projekt másik fontos része a históriás ének műfajának vizsgálata. Ide olyan, általában hosszabb epikus költemények tartoznak, melyek megtörtént eseményeket vagy bibliai, ill. fiktív (szerelmi, mitológiai) történeteket beszélnek el. A műfaj a 15. század végén jelenik meg, a 16. században teljeseedik ki, s a 17. század elejére már csak kis számban van jelen. Kutatásunkban a műfaj metrikai és stilisztikai egységét, sokszínűségét, valamint az oralitáshoz való kapcsolatát vizsgáljuk számítógépes metrikai és stilometriai eszközökkel. Az alábbiakban az OTKA kutatás ezen alprojektjének első eredményeiről számolunk be részletesebben.

## 1. A históriás énekek alprojekt

A kora újkori magyar költészet egyik legfontosabb műfaja a históriás éneké, arányuk a 16. századi magyar versek között jelentős: számukat tekintve több mint 12%, a terjedelem szempontjából, a sorok számát nézve pedig több mint 50% tartozik ide. A kevés kortárs irodalomelméleti és esztétikai reflexió szerint is a magyar kora újkor reprezentatív műfaját képviseli. A kor legfontosabb alkotói közül Tinódi Sebestyén vagy Bogáti Fazakas Miklós is számos históriás éneket írt különféle tematikával, történelmi, bibliai témákban.

Kutatásunk arra irányul, hogy az énekeket jellemző, sokak által monotonnak, túl egyszerűnek tekintett verselés pontos jellemzőit meghatározzuk, és megvizsgáljuk, egy bő évszázad alatt hogyan alakult át a magyar verselés logikája, jellege. Mennyire volt elterjedt az önrím, a ragrím, mennyire tiszták a rímek vagy például a szótagszámok és a sormetszetek szabályosságára mennyire ügyeltek a költők, esetleg felfedhetők-e eddig nem ismert strófaszerkesztési szabályok (pl. sormetszetek változása<sup>4</sup>). Stilometriai és sztachometriai módszerek segítségével anonim vagy bizonytalan szerzőségű versek alkotóit is megpróbáljuk azonosítani a kiterjedt verskorpusz segítségével (például az *Eurialus* és *Lucretia* című szerelmi históriáét).

A históriás énekkorpusz a 15. század végéről 4 éneket, a 16. századból 180 éneket (ebből 10 szövegét nem ismerjük), a 17. századból kb. 23–25 éneket számlál (a korpusz jelenleg a 16. század végéig teljes). Műfaj történeti áttekintésre a jelen írás keretei nem adnak lehetőséget,<sup>5</sup>

3 L. ennek kifejtését: Horváth Andor, „Poetry Database Connector (PDC)”, *Digitális Bölcsészet* 6. sz., 5 (2022): megjelenés alatt. További információ a POSTDATA nemzetközi költészettörténeti projektről: Poetry Standardization and Linked Open Data, hozzáférés: 2022. június 18., <https://postdata.linhd.uned.es>.

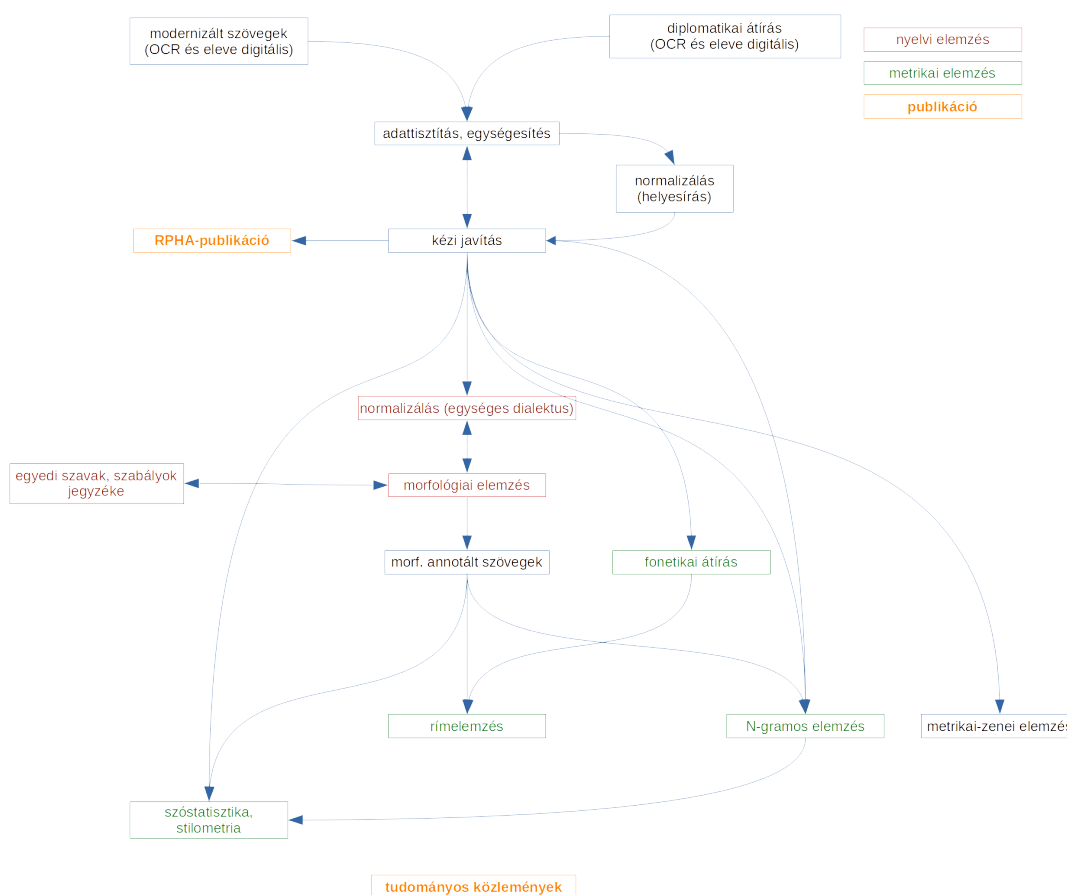
4 Erre az esetre szolgáltatott jó példát újabban Szatmári Áron előadása, mely Bogáti Fazakas (RPHA 2013) és Csanádi Demeter (RPHA 0230) egy-egy históriás énekének közös verselési szabályát fedte fel. Eszerint a 13 (7,6) sorok sormetszete a strófák harmadik sorában (az énekek második felére) nagy számban következetesen hiányzik. Szatmári Áron, „7+6, avagy Demeter király esete a kanásztáncossal”, *A históriás ének: poétikai és filológiai kérdések* műhelykonferencia, ELTE BTK Budapest, 2022. június 10.

5 Műfaj történeti áttekintéshez I. Pap Balázs, *Históriák és énekek*, Pannónia könyvek (Pécs: Pro Pannonia, 2014).

így csak röviden utalunk arra, hogy a művek kiválasztásakor a 15–16. századi anyag esetében az RPHA műfaji besorolása szerint gyűjtöttünk (*vallásos>historia* vagy *világi>historia* műfaji besorolású művek).<sup>6</sup> A 17. századi anyag összegyűjtése a releváns szakirodalom alapján történik és még folyamatban van. Gyűjtésünk az RMKT 17. századi sorozatában kiadott műveket tekinti magnak, azonban a barokk epikus költeményeket, melyek már távol állnak a históriás ének műfajától, nem tartalmazza. Ilyenek például Zrínyi Miklós *Obsidio Szigetiana* című eposza vagy Gyöngyösi István epikus költeményei. A projekt keretében a 17. századi énekek RPHA-szerkezetnek megfelelő bibliográfiai, irodalomtörténeti és poétikai leírása is elkészül, így válik majd a korpusz a szöveg és a metaadatok szintjén is együtt elemezhetővé. (Az RPHA eredeti gyűjtőköre az 1600-as évvel zárul, az OTKA projekt keretében azonban ilyen jellegű tartalmi bővítéseket is tervezünk.)

## 2. A korpuszépítés lépései

A históriás énekkorpusz összeállításának és elemzésének lépéseit az alábbi folyamatábrán foglaltuk össze. A jelen írás az előkészítő fázist mutatja be, mely során a nyomtatott, digitalizált vagy eleve digitális forrásokból előállt a nyelvi annotációkat tartalmazó, számítógépes elemzésre alkalmas, megbízható tudományos minőségű szöveganyag.



1. ábra: Folyamatábra a munka menetéről

Fontos kiemelni, hogy a régiség szöveganyagával való munka számos nehézséget rejt magában, hiszen különösen jellemző rá az egyedi és nem következetes ortográfia, a különféle nyelvjárások megléte, az archaikus grammatikai paradigmák és lexikális elemek nagy száma.

<sup>6</sup> A lekérdezés linkje: <https://f-book.com/rpha/v7/results.php?boole1=EMPTY&field1=M%C5%B1faj&value1=049&method1=LIKE&boole2=OR&field2=M%C5%B1faj&value2=002&method2=LIKE&boole3=AND&field3=A+szereztet%C3%A9s+ideje&value3=&method3=LIKE&boole4=AND&field4=R%C3%Admk%C3%A9plet&value4=&method4=LIKE>.

Ezekre az eddig fejlesztett nyelvi elemzők nincsenek felkészülve, így az előfeldolgozás nagy kihívások elé állított minket.

A korpusz alapját zömmel nyomtatott szövegforrások, a Régi Magyar Költők Tára sorozat már bescannelt és OCR-ezett, online elérhető,<sup>7</sup> valamint az általunk újrascannelt és OCR-ezett kötetei adták. Ezen felül kerültek bele digitális nyomdai fájlok (PDF), tudományos igényű internetes szövegkiadások (HTML) és kutatók által modernizált szövegek digitális kéziratban<sup>8</sup> (DOC, DOCX), valamint előfordultak egyéb, például folyóiratbeli szövegközlések is.

A szövegek kinyerését Abbyy OCR szoftverrel végeztük. Mivel az interneten elérhető különféle források scan és OCR minősége igen heterogén volt, a szövegfelismertetést sok esetben újra el kellett végezni. Az így kapott nyers szöveget több lépcsőben kézi és automatizált javításoknak vetettük alá. Az automatizált javításokat regexszel végeztük, ez elsősorban a nem releváns alfanumerikus elemek kiszűrését (pl. a sorszámozás és a fejlécek törlése), valamint a szöveg szegmentálását jelentette (sor és strófahatárok egységes jelölése). A kézi javítás során az így előfeldolgozott szöveget olvastuk össze a digitalizált forrással, az OCR (betűtévesztési és esetleges értelmi) hibáit szűrve ki. A kimenet egy egyszerű szöveg lett minimális, Markdown-szerű szintaxissal (sor- és strófahatárok, hiányok, fejezethatárok, paratextusok jelölésével/elkülönítésével). Ez a többszörös ellenőrzési folyamat jóval időigényesebb, azonban ettől lesz megbízhatóbb a nyelvi korpusz, ellentétben olyan heterogén forrású, szövegkritikailag nem ellenőrzött digitális bölcsészeti korpuszokkal, melyek a webes szövegforrásokat kontroll nélkül inkorporálják.

A kézi és automatizált javítások következő fázisa a modernizálás vagy normalizálás. Modernizálásra esetünkben két okból volt szükség. Egyrészt a korpusz jelentős része eleve modernizált kiadásban jelent meg az RMKT köteteiben, így a homogén korpusz előállításához minden szövegnél modernizálni kellett. Maga a több mint egy évszázadra visszatekintő RMKT sorozat is tükrözi a magyar helyesírás változásait, ezekre is tekintettel kellett lennünk (pl. c hangérték cz betűvel való jelölése a régebbi kötetekben). A modernizálás másik oka, hogy a nyelvi elemző programok a különféle ortográfiájú, nyelvjárású szövegeket jelentős hibaszázalékkal ismerik csak fel, s ezért eredménytelenek vagy fals eredményeket hoznak. A projektben elsősorban nem olyan jellegű kutatásokat végzünk, ahol az ortográfiai jellegzetességeknek jelentősége lenne (vö. némely stilometriai, szerzői attribúciót célzó kutatással), így az eredeti ortográfia az elemzés szempontjából nem releváns. Fontos azonban hangsúlyozni, hogy a munkafolyamat során létrejött különféle szövegátiratokat megtartjuk, azokra a munkafolyamat különböző állomásain szükség van. Például az eredeti, kritikai kiadáshoz közel álló szöveg az RPHA adatbázis felületén jelenik meg, a már csak verses szövegelemeket tartalmazó, de nem modernizált szöveg fonetikai annotációt kap, a számítógépes elemzésre szánt modernizált szöveg pedig részint nyelvi, részint metrikai annotációt stb. (Lásd a folyamatábrán fentebb.)

A kézi modernizálás számára közös szabályrendszert alkottunk, az egyes eseteket példákkal is illusztrálva. Noha törekedtünk az egységességre, a gyakorlat sokszor felülírta a szabályokat, következetlenségeket idézett elő, hiszen egyszerre kellett figyelembe venni a nyelvjárás vagy verstani szempontból fontos alakok megtartását, s ugyanakkor a (gépi) érthetőséget is szem előtt kellett tartanunk. A modernizáláshoz szükséges idő nagyon eltérő volt, attól függően, milyen a 16. századi nyomtatványok, kéziratok helyesírása. A 20. század második felében elterjedt, a régi helyesírást, sőt néha betűképet is pontosan tükröző kritikai kiadások esetében

7 Nemzeti klasszikusok kritikai kiadásai, BTK Irodalomtudományi Intézet, hozzáférés: 2022. 06. 18., <https://szovegtar.iti.mta.hu/hu/sorozatok/rmkt-xvii-szazad/>.

8 Ezúton mondunk köszönetet Pap Balázsnak és az elhunyt Vadai István örököseinek a rendelkezésünkre bocsátott szövegekért.



ez meglehetősen időigényes munka. Soronként gyakran 8–10 változtatásra is szükség volt, ha például az eredeti kiadás nem használt egyáltalán ékezeteket (vagy csak az a/á, e/é betűpároknál), vagy ha fonetikus és nem szóelemző helyesírással éltek a nyomdászok és másolók. Egy kb. 100 strófás vers modernizálásához korábbi filológiai tapasztalat birtokában is minimum 4–5 munkaóra van szükség.

Az automatizált előnormalizálás vagy modernizálás<sup>9</sup> során a nyelvjárásában, ortográfiájában homogénebb szövegcsoportokra a tipikus esetek figyelembe vételével szabályokat alkottunk, ezek mentén cserélte ki az előnormalizáló program a szövegek betűkészletének adott részeit. Önmagában ez a módszer sem nyújt megoldást, hiszen a kézi modernizálással szemben „túlságosan” is következetes, így nem minden esetet fed le és használatával a javításkor néha sajnos rontunk is. Például egy ö-ző szövegben az ö-k cseréje e-re az öröm (/N+Nom) szót az erem (/N+Poss.1Sg+Nom) szóra cseréli. Így mindkét esetben szükség van kézi ellenőrzésre is; a legcélravezetőbb a két módszer kombinálása és iteratív alkalmazása a megfelelő szövegállapot eléréséig.

A munkafolyamat következő lépése a szövegannotáció. A nyelvi elemzéshez az E-magyar, illetve az emMorph elemző ó- és középmagyarrá alakított változatát, az emMorphOMH-ot használtuk.<sup>10</sup> Utóbbi mellett azért döntöttünk, mert jóval alacsonyabb hibaszázalékkal ismeri fel a nyelvi elemeket, hiszen a régies nyelvi paradigmák, lexémák jó részét tartalmazza, melyeket az E-magyar nem.

A rímek elemzését a RhymeTagger<sup>11</sup> programmal végezzük, melynek implementálásában a készítő, Petr Plecháč segíti kutatócsoportunkat. A program bemenete a versszövegek IPA szabvány szerinti fonetikus átírása, melynek előállítását az Epitran és az Espeak programokkal végeztük tokenizálást követően.

Az így összeállt históriás énekkorpusz mérete számokban: 174 vers (a fennmaradó 11 vers töredék, melyet csak említésből ismerünk), 25 007 versszak, 98 141 verssor és 527 598 szövegszó. Összehasonlítás végett más történeti korpuszok mérete: az Ómagyar Korpusz<sup>12</sup> 3,2 millió szövegszót, a Történeti Magánéleti Korpusz<sup>13</sup> 1 millió 112 ezer elemzett szövegszót, a Magyar Történeti Szövegtár pedig 3 millió szövegszót tartalmaz. A kutatásaink során létrejött

9 Ezen munkafolyamatban kutatócsoportunk tagja, Simon Eszter volt segítségünkre.

10 Novák Attila, „Milyen a jó Humor? [What is good Humor like?]", in *I. Magyar Számítógépes Nyelvészeti Konferencia* (Szeged: SZTE, 2003), 138–44; Attila Novák, „A New Form of Humor – Mapping Constraint-Based Computational Morphologies to a Finite-State Representation”, in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, szerk. Nicoletta Calzolari és mtsai. (Reykjavik, Iceland: European Language Resources Association (ELRA), 2014), 26–31; Attila Novák, Borbála Siklósi, és Csaba Oravecz, „A New Integrated Open-Source Morphological Analyzer for Hungarian”, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, szerk. Nicoletta Calzolari és mtsai. (Paris, France: European Language Resources Association (ELRA), 2016), 23–28; Váradi Tamás és mtsai., „Az e-magyar digitális nyelvfeldolgozó rendszer”, in *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*, szerk. Berend Gábor, Gosztolya Gábor, és Vincze Veronika (Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, 2017), 49–60; Tamás Váradi és mtsai., „E-Magyar – A Digital Language Processing System”, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, szerk. Nicoletta Calzolari (Conference chair) és mtsai. (Miyazaki, Japan: European Language Resources Association (ELRA), 2018). Ezúton köszönet Sass Bálint és Novák Attilának, hogy rendelkezésünkre bocsátotta az elemzőt.

11 RhymeTagger, hozzáférés 2022. június 16., <https://github.com/versotym/rhymetagger>. Vö. Petr Plecháč, „A Collocation-Driven Method of Discovering Rhymes (in Czech, English, and French Poetry)”, in *Taming the Corpus: From Inflection and Lexis to Interpretation*, szerk. Masako Fidler és Václav Cvrček, *Quantitative Methods in the Humanities and Social Sciences* (Cham: Springer International Publishing, 2018), 79–95, [https://doi.org/10.1007/978-3-319-98017-1\\_5](https://doi.org/10.1007/978-3-319-98017-1_5).

12 <http://omagyarkorpusz.nytud.hu>

13 <http://tmk.nytud.hu/>

szövegtörzset nyílt licenc alatt tesszük közzé, így gazdagítva az elérhető magyar nyelvi korpuszok halmazát. (Részletek a kutatás folyamatosan frissülő GitHub repozitóriumban: <https://github.com/versotym/oldhun>.)

### 3. Első eredmények és a folytatás

Első kísérleti tanulmányunkban<sup>14</sup> 26 történelmi ének számítógépes metrikai vizsgálatát és annak eredményeit mutattuk be. A kutatás a történelmi énekkorpusz verstani homogenitását vizsgálta. A tanulmány elsősorban annak a közel-kortársi vélekedésnek járt utána kvantitatív eszközökkel, miszerint a történelmi énekek verselés monoton, kevés poétikai invencióval szolgál – ilyen a 17. század elején Szenci Molnár Albert „számtalan az sok vala vala vala” kijelentése is (*Psalterium Ungaricum*, 1607, Herborn). Ez a kis korpusz ugyan még korántsem volt reprezentatív, de elemzése tendenciákat így is sejtet. Eszerint az idő előrehaladtával határozottan változik a verselés technikája. Így például egyre kevesebb (a szövegekben kisebb arányú) a morfémarím, a szóismétlő rím és a rímtelen sorpár a vizsgált epikus énekekben.

Az elemzés arra is rámutatott, hogy egyes szerzők művei számítógépes metrikai eszközökkel is jól elkülöníthető csoportokat alkotnak. Az önrímek időbeli megoszlását figyelve például a történelmi énekköltészetben is kiemelkedő Tinódi Sebestyén és Bogáti Fazakas Miklós énekei alkotta csoportok láthatóan elkülönülnek, a korabeli átlagtól távol esnek. Tinódinál ugyan a három vizsgált énekből egy épp az átlag fölé kerül, azaz több benne az önrím – ezt az is indokolhatja, hogy az újabb kutatások szerint a szöveg valójában egy nagyobb történelmi ének töredékesen fennmaradt része, nem pedig különálló mű. Az eltérés pontos okára és további verstani és stilometriai mintázatokra a teljes Tinódi-korpusz vizsgálata deríthet majd fényt.

A történelmi énekköltészetet vizsgáló projektünk jelen célkitűzése, hogy

- a) a 16. századra összeállt teljes korpuszon elvégezzük a fenti vizsgálatokat;
- b) az énekkorpuszt kibővítsük a 17. századi anyaggal, mind a metaadatok, mind pedig a szövegek szintjén;
- c) a számítógépes metrikai elemzések mellett stilometriai vizsgálatokat végezzünk a nyelvileg annotált szövegtörzseten (pl. szerzői és műfaji csoportok, vagy az oralitás és írásbeliség vizsgálata);
- d) az RPHA szolgáltatotta metaadatok segítségével a statisztikai elemzéseken túl intertextuális és interperszonális jelenségeket kutassunk, a szöveg és a metaadatok elemzését összekapcsoljuk;
- e) az RPHA adatbázison további tartalmi és technológiai fejlesztéseket végezzünk.

A projektet megelőzően ilyen mélységű és ilyen változatos számítógépes elemzés még nem készült magyar verses korpuszon.

---

14 Szilvia Maróthy, Levente Seláf, és Petr Plecháč, „Rhyme in 16th-Century Hungarian Historical Songs: A Pilot Study”, in *Tackling the Toolkit: Plotting Poetry through Computational Literary Studies*, szerk. Petr Plecháč és mtsai. (Institute of Czech Literature of the Czech Academy of Sciences, 2022), 47–62, <https://doi.org/10.51305/ICL.CZ.9788076580336.04>.



## Kéziratok forrásainak transzformációinak lehetőségei a közgyűjteményekben Transformations of manuscripts in public collections

Szűcs Kata Ágnes  
Digitális Bölcsészeti Központ  
Országos Széchényi Könyvtár  
[szucs.kata@oszk.hu](mailto:szucs.kata@oszk.hu)

### Abstract

Providing digital accessibility for manuscript resources in public collections is feasible at various levels. Though what does digital processing of manuscripts mean? Filling in record fields of an online catalogue, uploading scanned images into Transkribus, and digital publishing a TEI XML file can all line up behind the concept. In my paper, I discuss the changes a manuscript undergoes and the multiple media formats it becomes interpretable and provided for the public. From a practical point of view, I demonstrate how the use of Transkribus is transforming a workflow of a public collection. The paper also investigates methods for building HTR models by comparing two basic theoretical strategies in practice. The first approach is to add the previously generated Ground Truth (GT) as a Base Model (BM). The other is to train an entirely new model by merging the old and new GTs without a base model.

**Keywords:** htr, handwritten text recognition, Transkribus, digital processing, digital edition, manuscript, public collection

### 1. Bevezetés

A Digitális Bölcsészeti Központ Kiss József-projektje egyrészt mint digitális forráskiadás, másrészt mint kéziratok feldolgozását és közzétételét kidolgozó közgyűjteményi feladat és végül mint automatikus kézírásfelismerést szorgalmazó vállalkozás is fontos célokat tűzött ki maga elé és valósított meg.

AdHUpla<sup>1</sup> – Digital Humanities Platform – digitális szövegkiadásokat és a szövegfeldolgozáson alapuló kreatív tartalmakat tesz közzé. Tehát egy olyan publikálási környezet, amely a közgyűjtemények szöveges forrásainak digitális megjelenítésére szolgál. Az elmúlt két évben, a Móricz-kutatócsoport<sup>2</sup> munkája mellett, ez a közgyűjteményi pilot projekt segítette feltérképezni és kialakítani a megjelenítéstől kezdve az xml publikáció tagkészletén át, a kép és szöveg összekapcsolásának lehetőségeit.

<https://dhupla.hu/collection/kiss-jozsef-levelazes>

A Kiss József-projektben továbbá kialakult egy olyan workflow, amely bármely kézirat esetében alkalmazható lépéseket ír elő a feldolgozáshoz és nem utolsósorban a publikálás elkészítéséhez, másrészt pedig a szoftverhasználatra is létrejött egy részletesen kidolgozott, mégis rugalmas ajánlás. Ez főképp a Transkribus platformot,<sup>3</sup> amely történelmi dokumentumok

1 Digital Humanities Platform, [www.dhupla.hu](http://www.dhupla.hu) hozzáférés: 2022. június 23.

2 A munkáról bővebben a Móricz-műhely jelenleg költöztetés alatt álló oldalán lehet olvasni, <https://pim.hu/hu/digitalis-bolcseszeti-kozpont/moricz>, hozzáférés: 2022. június 23.

3 READ-COOP, Transkribus, <https://readcoop.eu/transkribus> hozzáférés: 2022. június 23.



szövegfelismerésére, képelemzésére és szerkezetfelismerésére alkalmas, és az Oxygen tei-xml szerkesztő szoftvert<sup>4</sup> helyezi előtérbe.

A közgyűjteményekben fellelhető kéziratos források digitális szolgáltatása különböző szinteken lehetséges. A kéziratok mediális transzformációja ebben a projektben az alábbi, egymáshoz is szervesen kapcsolódó állomások köré szerveződött:

- Fizikai példány
- Digitalizált másolat(ok)
- Transkribus dokumentum
- Kétrétegű PDF
- XML-file
- Online publikáció (dHUplá)

A kéziratok minden formájának megvan a maga jelentősége és önmagában olyan információkat hordoz, amelyeket más változatok nem. A Transkribus dokumentum már egy átírással rendelkező, szegmentált képet, illetve xml dokumentumot tartalmazó objektum, amely ebben a formában lehetőséget teremtett nem csak a digitális forráskiadás létrehozására, hanem automatikus kézírásfelismerő modell építésére is. A továbbiakban esettanulmány jelleggel az e téren elért újabb eredményekről lesz szó.

## 2. Automatikus kézírásfelismerés (Handwritten Text Recognition – HTR)

A szoftver a gyakorlatban is alkalmasnak bizonyult a közgyűjteményekben fellelhető különböző kéziratok kezelésére és publikálására. A jelenlegi workflowban egyrészt az átírások elkészítésére és ezek különböző formátumokba történő exportálásra használjuk, illetve folytattuk az automatikus kézírásfelismertetést is a Kiss József-levelezésen. A Transkribus szervezet<sup>5</sup> és a körülötte kialakult felhasználói közösség<sup>6</sup> pedig több téren is segítette a kézírásfelismerésben végzett munkánkat.

Egy korábbi publikációmban<sup>7</sup> bemutattam az általánosságoktól a konkrét kimenetelig a Transkribus program használatát, a mesterséges intelligencián alapuló kézírásfelismerő-modell építésének módját, illetve a Kiss József kézírásán tanult első modellek eredményeit is. Az egy kéz által írt leveleken tanult eddigi legjobb modellünk 6,94-es hibaszázalékkal rendelkezik.

Hosszútávú terveink közé tartozik egy minél általánosabb, a 19–20. századi magyar írók kézírását könnyen felismerő modell létrehozása. Efelé tett következő lépés az volt, hogy a Kiss József-levelezés másik felét, azaz a 19. századi magyar költőnek, A Hét című hetilap szerkesztőjének címzett leveleket is bevontuk a modellépítésbe. Az így keletkezett vegyes kézírásos korpusz már elegendő mennyiségű adattal rendelkezett egy általánosabb magyar

4 Oxygen XML Editor, <https://www.oxygenxml.com/> hozzáférés: 2022. június 23.

5 A kezdeményezés 2016 óta létezik és az Európai Unió 2020-as Horizon programja által finanszírozott kutatási projektből nőtte ki magát. 2019 július 1-jétől Read COOP néven és immár vállalati alapokon folytatja megkezdett munkáját. European Commission CORDIS research results, *Recognition and Enrichment of Archival Documents*, hozzáférés: 2022. június 23., <https://cordis.europa.eu/project/id/674943>. Vö.: READ-COOP, *Our Story*, hozzáférés: 2022. június 23., <https://readcoop.eu/our-story/>.

6 „Transkribus users | Facebook”, hozzáférés: 2022. június 23., <https://www.facebook.com/groups/614090738935143>.

7 Szűcs Kata Ágnes, „Automatikus kézírás-felismertetés Kiss József levelezésén”, in *Online térben az online térért: Networkshop 30: országos online konferencia*. 2021. április 6-9. Eötvös Loránd Tudományegyetem (Networkshop, HUNGARNET Egyesület, 2021), 73–80, <https://doi.org/10.31915/NWS.2021.8>.



nyelvű kézírást felismerő modell létrehozásához, amely azóta nyilvánosan is hozzáférhető a Transkribus oldalán<sup>8</sup> és az asztali alkalmazáson keresztül is.

Neve: *Transkribus Hungarian Handwriting 19th–20th Century*

Epoch-ok száma: 250

Szavak száma: 74.862

Sorok száma: 16.630

## 2.1. Vegyes kézírásos tanult modellek

A jelenlegi modellhez felhasznált kéziratok a Petőfi Irodalmi Múzeumban találhatóak, melyek között szerepelnek borítékok, képeslapok, hagyományos és fejléces levelek, névjegykártyák. A levélírók Kiss József és családja, illetve a századforduló írói, újságírói és művészei. Ez nagyságrendileg 300 darab változó hosszúságú és minőségű levelet jelent. A levelezés további kéziratjai az Országos Széchényi Könyvtárban találhatóak, amelyekkel terveink szerint a modell tanítókorpusza a későbbiekben bővülni fog.

A vegyes kézírásmodell építéséhez alapvetően több adat szükséges, mivel nagyobb változatosság mutatkozik az egyes írásképek között, mint egyetlen ember esetében. Például, a Kiss József kézírásán tanult modellnél körülbelül 1/3 mennyiségű adat állt rendelkezésre, mégis 2,27%-kal jobb eredményt sikerült elérni.<sup>9</sup>

Az egyik legfontosabb mérőszám a HTR modell építésekor, a Character Error Rate (CER) on Validation Set,<sup>10</sup> azaz, az ellenőrző korpuszon ejtett karakterhibaérték aránya. A jelenlegi legjobb eredmény 9,19% – és bár akadnak olyan levelek a korpuszban, amelyeket szinte tökéletesen képes átírni, azt nem lehet elvárni ettől a modelltől, hogy ezt bármilyen magyar nyelvű kézirattal megtegye.

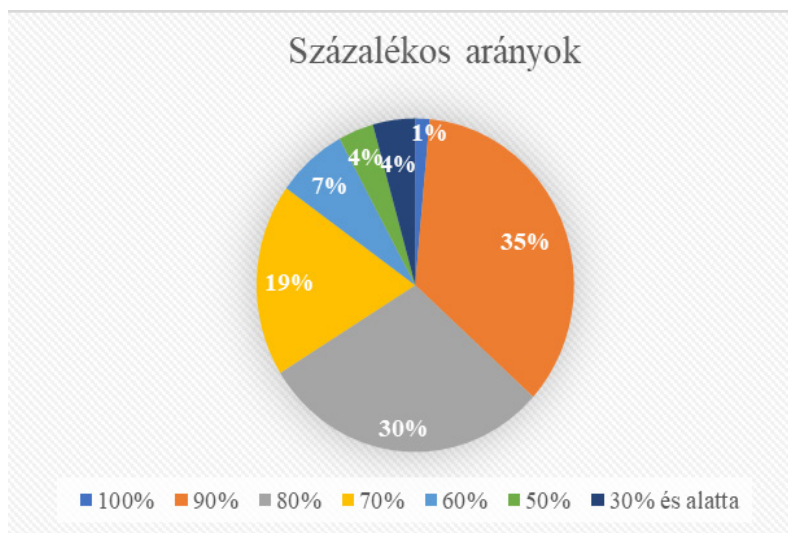
Egyrészt azért, mert a modell egy adott korpuszon, Kiss József-levelezésen tanult (1265 oldal), és így jelenleg ennek a felismerésére a legalkalmasabb. A 9,19-es érték pedig egy előjelzett átlag, amelyet a Validation Set nevű elkülönített ellenőrzőkorpuszon (138 oldal) számol ki az algoritmus.

Tehát a modell ezen a tanulási folyamatból kimaradó kézírathalmazon átlagosan közel 90%-os pontossággal meg tudja állapítani, hogy milyen karakterekből áll össze egy levél szövege, amely jó indikátora annak, hogy hogyan teljesítene egy új, ismeretlen kéziraton. Ez az eredmény egyes leveleknél valóban 90-, vagy akár 100%-os pontosságot jelent, de más, nehezebb kézírások/írásképek vagy oldaltükrök esetében 60-, 20- vagy akár 0%-os eredményt is produkálhat. Ezeket az olvasatokat átlagolja az algoritmus a CER meghatározásakor. Egy általános, a magyar nyelvű kéziratokat széleskörűen felismerő modellhez a jelenleginél sokkal több adatra van szükség.

8 READ-COOP, *Transkribus Hungarian Handwriting 19th–20th Cent.*, hozzáférés: 2022. június 23., <https://readcoop.eu/model/hungarian/>.

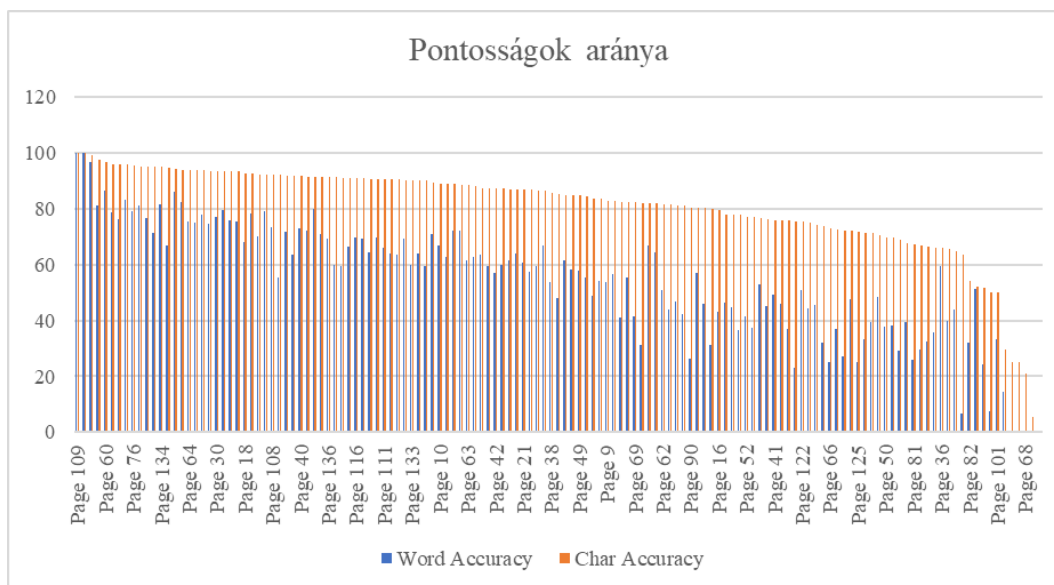
9 CER on VS: 6,94%  
Epoch-ok száma: 200  
Szavak száma: 74.862 25.348  
Sorok száma: 16.630 5.634

10 Vö.: Elisabeth Heigl, *CER? Don't Worry!*, hozzáférés: 2022. június 23., <https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/cer-dont-worry/>.



1. ábra A CER on Validation Set százalékos eloszlása: 100% tökéletes, átírás 30% és alatta, jelentős hibákat tartalmazó átírás.

Egy másik mérőszám, a Word Error Rate (WER) a szavak hibaarányát jelzi. A helyesen átírt szavak megoszlása a karakterekhez képest nagyobb kilengéseket mutat, mert egy karakter hibás felismerésekor az adott szót is, amelyben az szerepel automatikusan hibásnak fogja értékelni a program. Ez abban az esetben is igaz, ha csak egy nagybetű-kisbetű tévesztésről, esetleg a szó végén lévő írásjel elhagyásáról van szó.



2. ábra Szavak pontosságának az alakulása a karakterpontossághoz képest.

## 2.2. Modellek egymásba építése

A vegyes kéziratkorpusz folyamatos feldolgozása új kérdéseket vetett fel a modellek egymásba építésével kapcsolatban. Mivel a modellépítéshez szükséges átírt kéziratok folyamatosan készülnek, ezért egy-egy nagyobb mennyiségnél fontossá vált, hogy milyen módon tudjuk azokat beépíteni a már meglévő modellbe. Ezáltal be tudtunk kapcsolódni egy aktív nemzetközi diskurzusba, melyből kiderült, hogy ezek a kérdések nemcsak a Transkribus használói, hanem a fejlesztői számára sem egyértelműek.



Két elméleti lehetőség van, de az, hogy melyik a jó megoldás az a kéziratok korpusztól is függ. Eddigi tapasztalataink szerint befolyásolja az eredményeket, hogy mennyire homogén a kézírás, időben közel vagy távolabb vannak-e egymástól az egyes dokumentumok, műfajuk megegyezik-e stb.

Az első opció, hogy az új anyagba alapmodellként építjük be az eddigi kéziratokon tanult modellt és így fejlesztjük tovább.<sup>11</sup> Ebben az esetben a modell minden egyes tanulási ciklusa egy már meglévő modellre, azaz alapmodellre (Base Modellre – BM) épül. Az alapmodellek „emlékeznek” arra, amit már megtanultak, ezért elméletileg minden egyes új tanulási ciklus javítja a modell minőségét. Az új modell tanul az elődjéből, és így egyre jobbá válik.<sup>12</sup>

A második, hogy az elkészült új anyagot Ground Truth-ként (GT) adjuk hozzá a régihez és ebből hozunk létre új modellt. A GT vagy alapigazság egy statisztikából ismert alapfogalom, amely egy adott kérdéssel kapcsolatos igazság ismeretére vonatkozik, ez az ideális elvárt eredmény. Kéziratok esetében ez egy dokumentum 100 %-ban helyes gépelt példányát jelenti, amely a Mesterséges Intelligencia betanítására szolgál.<sup>13</sup>

Mindegyik modell HTR+ technológiával készült és a modellépítésnél be volt kapcsolva az *omit line by tags* funkció, melynek hatására az algoritmus automatikusan figyelmen kívül hagyja azokat a sorokat, ahol bizonytalan olvasatot vagy olvashatatlan szövegrészt jelöltünk. Ez minimum 2–3% javulást eredményezett a modelleknél, annak ellenére, hogy ezáltal kb. 1000 sorral kevesebb lett a GT. Az első lépésben elkülönítettünk egy kisebb Tesztkorpuszt (TK). Ez az anyag egyáltalán nem vett részt a modellépítésben, így befolyásoltság (bias) nélküli tesztelésre adott lehetőséget a modellépítés során.

### 2.3. A Base Modellel tanított modellek

Az első vegyeskézírásból készült modellünk 484 oldalnyi szövegen tanult (6.716 sor), alapmodell nélkül 10,38%-os lett. Ebbe építettük be a Kiss József kézírásán tanult modellt (CER on VS: 6,94%), amellyel 0,27%-ot javult a teljesítménye.

A második vegyeskézírásból készült modell 298 új oldalnyi szövegen tanult és azt a modellt (ID: 34757 Vegyes\_kézírás\_13) használtuk Base Modellként, amelybe Kiss József kézírása is be volt építve alapmodellként. Amikor az első vegyeskézírás korpuszon épült modellt, (ID: 34600 Vegyes\_kézírás\_11) használtuk alapmodellként, 0,1%-kal kaptunk rosszabb eredményt. Ebből az látszik, hogy egy alapmodell alapmodelljeként beépített – jelen esetben a Kiss József kézírásán tanult – modell jelenléte vagy hiánya alig okozott teljesítmény béli különbséget. Harmadik lépésként az összevont vegyes kézíráskorpuszhoz építettük be a Kiss József-kézírása alapmodellét, amely a legjobb eredményt hozta az első fázisban (9,92%). Tehát nagyobb hatása van az alapmodellnek, ha nincs annyira „eltemetve.”

11 Vö.: Dirk Alvermann, *Use Case: “Model Booster”*, hozzáférés: 2022. június 23., <https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/use-case-model-booster/>.

12 READ-COOP, *Base Models*, hozzáférés: 2022. június 23., <https://readcoop.eu/glossary/base-models/>.

13 READ-COOP, *Ground Truth*, hozzáférés: 2022. június 23., <https://readcoop.eu/glossary/ground-truth/>.

	50 epoch	100 epoch	150 epoch	200 epoch
BM 34757 Vegyes kézírás_13	10,05%	10,28%	10,35%	0%
BM: 34600 Vegyes kézírás_11	10,68%	10,15%	10,42%	10,20%
BM: 31617 Kiss József kézírása + Összevont GT	12,27%	10,28%	9,92%	10,30%

3. ábra. Második vegyes kézírásmodellek alapmodellel.

Az alapmodellel történő modellépítés egyik vitathatatlan előnye, hogy viszonylag kevés epoch alatt éri el a maximumát, ami idő- és energiatakarékos megoldás lehet bizonyos esetekben. Az epoch a Transzkibus terminológiájában azt az iterációs ciklust jelenti, ahányszor az algoritmus a tanulási folyamatot elvégzi az adott korpuszon.

## 2.4. Összevont Ground Truth alapú modellek

A második szakaszban az összevont GT-s modellekkel folytattuk a tesztelést.

Először kombináltuk az első és második vegyes kézírásmodell anyagát, és BM nélkül az algoritmus 10,34%-os CER értéket produkált a korpuszon. Ehhez hozzátettük a Kiss József kézírásával készült levelek korpuszát, ezen már 9,31%-os eredményt értünk el.

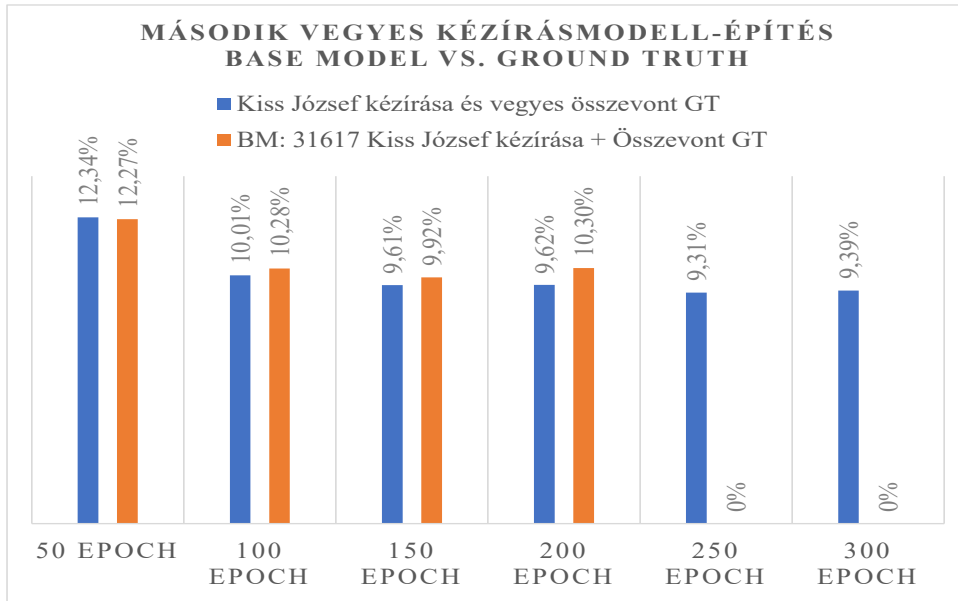
A különböző tesztek lefuttatása után a TK-t is beépítettük a modellbe és így értük a vegyes korpusz eddigi legjobb, 9,19%-os eredményét.

	50 epoch	100 epoch	150 epoch	200 epoch	250 epoch	300 epoch
Vegyes összevont GT	12,71%	10,96%	10,72%	10,55%	10,36%	10,34%
Kiss József kézírása és vegyes összevont GT	12,34%	10,01%	9,61%	9,62%	9,31%	9,39%
Kiss József kézírása és vegyes összevont GT + Teszt Korpusz	12,16%	10,14%	9,89%	9,19%	9,19%	0%

4. ábra. Második vegyes kézírásmodellek összevont alapigazsággal.

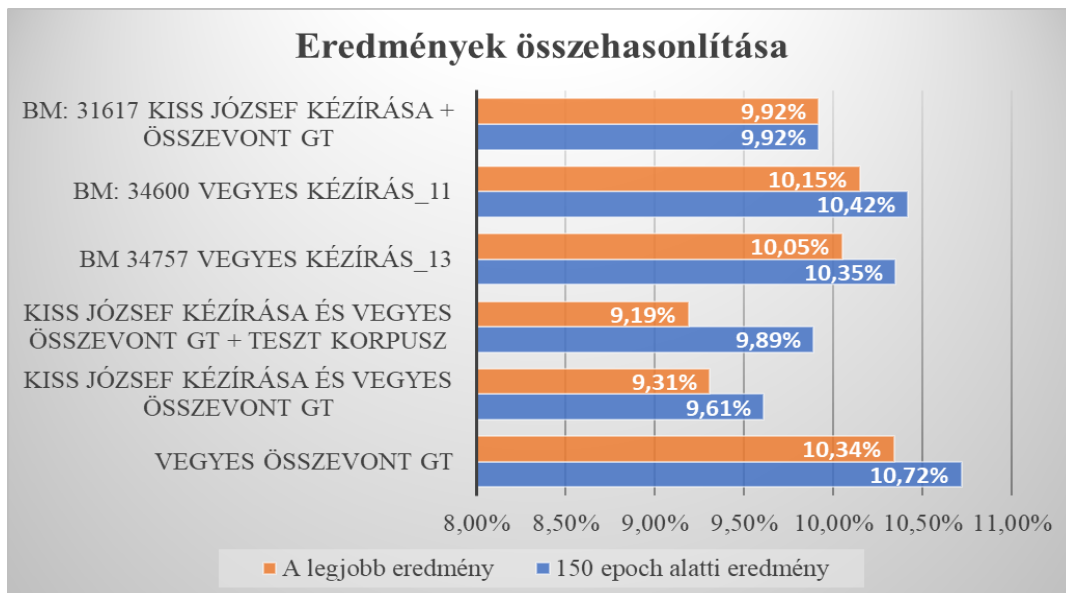
## 2.5. Az eljárások összehasonlítása

A Base Modelles és Ground Truth-os eljárásokat nehéz összehasonlítani, mert mechanizmusukból fakadóan más-más esetben hoznak jó eredményt. A diagrammon is látszik, hogy az alapmodellel rendelkező modell kevesebb epoch alatt hozott jó eredményt és egy bizonyos mennyiségű ismétlés (150 epoch) felett már túltanulta magát, így nőni kezdett a hibaszázalék aránya. A GT-s modellek épp ellenkezőleg, eleinte sokkal rosszabb eredményeket hoztak és csak 150 epoch után kezdett az érték 10% alá csökkenni.



5. ábra. Az eljárások összehasonlítása.<sup>14</sup>

150 epoch-nál még mindegyik modell értékelhető eredményeket hozott és összehasonlítva ezeket már körvonalazódnak a végső eredmények is. A Base Modellel rendelkező modell itt érte el a maximumát, a Ground Truth-os modellek közül a végén a második és az első helyet cserél, ugyanis 250 epochon 0,12%-kal jobban teljesít a TK beépítése utáni adattöbblet miatt.



6. ábra. Az hibaszázalékok összehasonlítása 150 epoch-nál.

14 250 epoch-nál 0%-os érték szerepel, mert az egyvel korábbi szakaszon is romló tendenciát mutatott a modell, így utána nem futtattuk le a betanítást.



### 3. Kitekintés

Terveink közé tartozik, hogy folytatjuk az automatikus kézírásfelismerés kutatását és tesztelését. A következő lépés az, hogy az OSZK-ban található Kiss József-levelezés elemeit is beépítsük a modellbe, hogy minél inkább rátanítsuk azt a vegyes kézírás felismerésére.

A legjobb eredmény érdekében a különböző magyar nyelvű projekteknek, amelyek automatikus kézírásfelismertetést használnak, a jövőben össze kell fogniuk, hogy a saját korpuszokon betanított modelleket egymásba építve egyre általánosabb érvényű automatikus kézírásfelismertető eszköz jöjjön létre a magyar nyelvű források kutatható közzétételéhez. Meg kell találni azokat az eszközöket, amelyek segítségével a közgyűjteményekben rejtőző kéziratos kincsek a digitális térben hozzáférhetővé, feldolgozhatóvá, kutathatóvá válnak. A most nyilvánosságra hozott első magyar kézírásfelismerő modell ennek a folyamatnak volt fontos mérföldköve.

### 4. Felhasznált irodalmak

Alvermann, Dirk. *Use Case: "Model Booster."* Hozzáférés: 2022. június 23.

<https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/use-case-model-booster/>.

European Commission CORDIS Research Results. *Recognition and Enrichment of Archival Documents | READ Project | Fact Sheet | H2020.* Hozzáférés: 2022. június 23.

<https://cordis.europa.eu/project/id/674943>.

Heigl, Elisabeth. *CER? Don't Worry!*, Hozzáférés: 2022. június 23.

<https://rechtsprechung-im-ostseeraum.archiv.uni-greifswald.de/cer-dont-worry/>.

READ-COOP. *Base Models.* Hozzáférés: 2022. június 23.

<https://readcoop.eu/glossary/base-models/>.

READ-COOP. *Ground Truth.* Hozzáférés: 2022. június 23.

<https://readcoop.eu/glossary/ground-truth/>.

READ-COOP. *Our Story.* Hozzáférés: 2022. június 23. <https://readcoop.eu/our-story/>.

READ-COOP. *Transkribus Hungarian Handwriting 19th–20th Cent.* Hozzáférés: 2022. június 23. <https://readcoop.eu/model/hungarian/>.

Szűcs, Kata Ágnes. „Automatikus kézírás-felismertetés Kiss József levelezésén”. In *Online térben az online térért: Networkshop 30: országos online konferencia.* 2021. április 6-9. Eötvös Loránd Tudományegyetem, 73–80. HUNGARNET Egyesület, 2021.

<https://doi.org/10.31915/NWS.2021.8>.

*Transkribus users | Facebook.* Hozzáférés: 2022. június 23.

<https://www.facebook.com/groups/614090738935143>.

## A digitális filológia infrastruktúrái. A DigiPhil megújulásáról.

Fellegi Zsófia

Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet

[fellegi.zsofia@abtk.hu](mailto:fellegi.zsofia@abtk.hu)

### 1. Bevezetés

A Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézete 2014-ben indította el a Petőfi Irodalmi Múzeummal együttműködésben digitális filológiai projektjét, a DigiPhil (A magyar irodalomtudomány filológiai portálja) szolgáltatást. 2021 májusától a szolgáltatás az Irodalomtudományi Intézet szakmai felügyeletével és keretei között működik tovább. Ezzel egy időben a Bölcsészettudományi Kutatóközpont konzorciumi partnerként csatlakozott a Digitális Örökség Nemzeti Laboratóriumhoz, és az együttműködés keretein belül megkezdődött a projekt szakmai és technikai átalakítása. A folyamat részeként a DigiPhil teljes technikai infrastruktúrája megújul, a szolgáltatás alapját az egyre szélesebb körben elterjedt szemantikus web technológia adja. A DigiPhil korábban kritikai kiadásokat informatizált,<sup>1</sup> ún. born digital kiadások számára nyújtott szakmai segítséget és publikációs platformot, valamint bibliográfiai adatbázist épített. A megújulás részeként, a korábbi tapasztalatokat továbbfejlesztve olyan virtuális, irodai szoftvert imitáló kutatókörnyezetet hozott létre, ahol egy kritikai kiadást készítő kutatócsoport online publikálásra kész, a standardoknak megfelelő jelölőnyelvi átíratot állíthat elő, bonyolult kódok írása nélkül. Így az ITI-ben és a partnerek által készülő kritikai kiadások esetén a nyomtatott kiadás státusza megváltozik, csupán egy adott verziója lesz a digitális kiadásnak, abból automatikusan előállítva. A DigiPhil nem csak a szolgáltatás mögött álló teljes infrastruktúrát alakítja át, hanem a TEI XML jelölőnyelvi átírat vizualizációjára szolgáló megjelenítő eszközt is lecseréli, amely így képes kiszolgálni a különböző filológiai elvek alapján készített kritikai kiadások egyedi igényeit.

A következőkben az átalakítás folyamata, a migráció nehézségei, az új szolgáltatások és rendszerek kapcsolata, valamint a legújabb fejlesztések kerülnek bemutatásra.

### 2. DigiPhil 1.0

Már a DigiPhil szolgáltatás kialakításakor, 2013-ban is számos olyan kritérium fogalmazódott meg, amelyek a mai napig hatással vannak a szolgáltatás működésére és fejlesztésére. Az egyik legfontosabb kritérium, hogy a DigiPhil infrastruktúrája megfeleljen a digitális kulturális örökség elvárásainak, amelyeket olyan nemzetközi intézmények, mint például az Europeana fogalmaztak meg. Ezek az elvárások a következők: platformfüggetlenség, nyílt forráskódú eszközök használata, metaadatszabványok használata és a nyílt hozzáférés biztosítása a kulturális örökséghez.

Aszolgáltatás kialakítása többlépcsős folyamat volt. Első lépésként ki kellett alakítani a kiadások alapját képező XML fájlokra egy olyan specifikációt, amely a Text Encoding Initiative ajánlását követi, annál azonban szűkebb. Ennek oka, hogy a TEI ajánlása rendkívül szerteágazó, egyes filológiai jelenségekre különböző jelölési lehetőséget biztosít, egy szolgáltatás fejlesztésekor azonban választani kell a lehetőségek közül. A DigiPhil specifikációja folyamatosan fejlődik,

1 A digitalizálás kifejezés nem elég kifejező a nyomtatásban megjelent kritikai kiadások adatgazdagított jelölőnyelvi átíratai, illetve az azokból készített kétrétegű PDF-ek közötti különbségek leírására, ezért, Palkó Gábor nyomán az *informatizálás* kifejezést használom a folyamat leírására. A fogalmak közötti különbségekről bővebben lásd: Fellegi Zsófia, *Átmenet az analógból a digitális filológiába: médiumváltás?: Madách Imre: Az ember tragédiája. Szinoptikus kritikai kiadás*, Helikon, 2021/I., 91.

annak érdekében, hogy ki tudja szolgálni a szolgáltatásban elérhető kritikai kiadások eltérő igényeit. A specifikációról készült dokumentáció hamarosan elérhető lesz a DigiPhil oldalán. Egy levelezés kiadás elkészítéséhez mind az XML fejlécében, mind a szövegtestben eltérő elemekre van szükség, mint egy genetikus kritikai kiadás, a szövegváltozatokat egy fájlban leíró jelölőnyelvi átíratában.

A kezdeti kísérleteket és teszteket<sup>2</sup> követően választottuk ki a szolgáltatás alapját képző, a repozitálást is lehetővé tevő infrastruktúrát. A Grazi Egyetemhez tartozó Zentrum für Informationsmodellierung, az Austrian Center for Digital Humanities (ACDH) tagja 2014. augusztus 27-én, az azóta a digitális bölcsészet és filológia nélkülözhetetlen eszközévé vált Github kód repositóriumban tette közzé,<sup>3</sup> az akkor már évek óta saját, digitális filológiai céljaira fejlesztett eszközt, a GAMS (Geisteswissenschaftliches Asset Management System) alapját adó Cirilo-t szoftvert.

A Cirilo egyszerre alkalmas TEI XML átíratok, az arra vonatkozó és azokból kiolvasott metaadatok kezelésére, a vizualizációs stíluslapok implementálására.<sup>4</sup> Az eszköz kiválasztása befolyásolta az infrastruktúra kiépítését, valamint maguknak a TEI XML fájloknek a felépítését is. A Cirilo rendszer alapját egy Fedora 3.6 repositórium adja, amelyben a digitális objektumok úgynevezett „context-object” viszonyban állnak egymással, az egyes „context”-ek, amelyek az egyes „object”-eket összekapcsolják pedig egymásba ágyazhatók. Az egyes kiadások egészét, vagy egy-egy részét reprezentáló TEI XML fájlok egy „context” alá sorolva alkottak egységet. A DigiPhil 1.0 szolgáltatásban ezért az egyes digitális kritikai kiadások úgy épültek fel, hogy a digitális objektumok, ami lehet egy teljes regény, vagy csupán egy vers jelölőnyelvi átírata, összessége rendelődött a kiadást jelentő kontextus alá. A TEI XML fájlok vizualizációjára a, szintén digitális bölcsészek által fejlesztett, Versioning Machine, bonyolult, sok helyen hiányosan dokumentált XSLT stíluslapokból álló eszközt használta a szolgáltatás. A DigiPhil hosszú éveken át fejlesztette és szabta testre a megjelenítő eszközt, azonban a Versioning Machine egyik legnagyobb hibáját nem sikerült orvosolni. Az eszköz, bár rendkívül stabilnak bizonyult, egy nagy méretű TEI XML fájlt rendkívül lassan volt képes csak megjeleníteni, ha egyáltalán meg tudott.

A munkafájlok és a publikált kiadások TEI XML fájljainak verziókövetése és archiválása kódrepositóriumokban történt. A kiadás készítésében részt vevő kollégák a munkafájlokat egy saját SVN szerverre mentették, onnan, az automatikus feldolgozást (például azonosítók kiosztása) követően, a DigiPhil zárt, Bitbucket git repositóriumába töltötték át. Erre azért volt szükség, mert a Bitbucket rendszerében az ingyenes hozzáférés korlátozott volt, egy-egy gyűjteményt mindössze öt személlyel lehetett megosztani. Ez a gyakorlat azonban rendkívül nagy odafigyelést és szakértelmet igényelt, magas volt az adatvesztés lehetősége.

A TEI XML struktúrája, mivel rendkívül széles kutatóréteg igényeire alakították ki, nem alkalmas az igazán strukturált adatkezelésre, így a személy- és helynevek, valamint bibliográfiai adatok leírására. A szolgáltatást ezért egy integrált könyvtári rendszer, a Koha<sup>5</sup> egészítette ki, amely tartalmazta a Petőfi Irodalmi Múzeum névterét, valamint a Geotauruszt. Emellett a DigiPhil külső adatbázisokból (pl. OSZK, Humanus) automatikusan töltött be bibliográfiai adatokat, amelyeket a munkatársak a rendszerben bővítettek és javítottak (pl. Kosztolányi

---

2 A DigiPhil elsőként a SADE (Scalable Architecture for Digital Editions) akkoriban fejlesztett eszközét tesztelte, azonban akkor még nem állt a készültség megfelelő fokán.

3 A Cirilo dokumentációja az alábbi linken érhető el: <http://gams.uni-graz.at/archive/objects/o:gams.doku/methods/sdef:TEI/get?locale=de> (hozzáférés: 2022.09.20.)

4 A TEI XML fájlok vizualizációjára gyakran XSLT (Extensible Stylesheet) stíluslapokat alkalmaznak, amely automatikusan generálja az XML fájlból a megjelenítésre alkalmas HTML-t.

5 A DigiPhil könyvtári adatbázisa a teljes állásig az alábbi linken érhető el: <http://biblio.digiphil.hu> (hozzáférés: 2022. 09. 20.)



forrásjegyzék). A TEI fájlokban a felismert entitások a Koha-ban található besorolási rekordokkal lettek összekötve. Kiderült azonban, hogy egy könyvtári rendszer nem alkalmas a digitális filológiai igények kiszolgálására, a MARC21 szabvány túlzottan merev, ezért bizonyos filológiai adatok leírása, mint például a kéziratok közötti bonyolult viszonyrendszer, nem lehetséges.

A bonyolult és eltérő felépítésű TEI XML fájlokban történő keresés nem triviális feladat. A DigiPhil szolgáltatásban eleinte egy informatikus kolléga által fejlesztett Solr alapú, facettás keresőfelület tesztelése zajlott, azonban miután az együttműködés megszűnt, lehetetlenné vált megfelelő és megfizethető, a Solr-hoz értő szakembert találni. Ezt követően egy jóval egyszerűbb, és pont ezért kevésbé kifinomult rendszert épített a DigiPhil, amelynek az alapját egy ExistDB adta. Az ExistDB azonban nem kezeli jól a nagy mennyiségű XML fájlokat, a DigiPhil fejlődésével pedig a kezelendő XML fájlok száma lassan elérte a tízezres nagyságrendet.

1. ábra A DigiPhil Arany János kéziratkeresője

## 2. DigiPhil 2.0

A DigiPhil infrastruktúrájának megújítása égetővé vált, hiszen a digitális kritikai kiadások száma folyamatosan emelkedik, a filológusok közül egyre többen szeretnének elektronikus kiadást készíteni. A Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézetében készülő minden új kritikai kiadásnál elvárás, hogy eleve digitálisan, TEI XML alapon szülessen, a nagy múltú Bibliotheca Scriptorum Medii Recentisque Aevorum (BSMRAe) sorozat minden új kiadása szintén a DigiPhil szakmai segítségével készül. Bár a digitális kritikai kiadások publikálása egyre elterjedtebb, továbbra is elvárás, hogy a digitális mellett nyomtatott formában is megjelenjenek a kiadások. A DigiPhil erre az igényre válaszol olyan algoritmusok fejlesztésébe kezdett, amelyek a kritikai kiadásokat reprezentáló TEI XML átíratokból automatikusan előállítja a nyomdai előkészítésre alkalmas formátumot. Az első ilyen kiadás Oláh Miklós levelezésének második kötete, itt az XML fájlokból LaTeX formátum jön létre.

A hosszú távú tervek között szerepel olyan algoritmus fejlesztése, amely a sokkal elterjedtebb, InDesign tördelőprogram által használt formátumot tudja előállítani.

A DigiPhil infrastruktúra megújulásának része az adatgazdagítás támogatására korábban használt könyvtári rendszer leváltása. A digitális filológia területén szinte adja magát, hogy a szemantikus web technológiáit integráljuk szolgáltatásainkba. Az elmúlt években egyre elterjedtebb Wikibase alapú rendszerek építése a kutatások támogatására.<sup>6</sup> A DigiPhil a Digitális Örökség Nemzeti Laboratóriummal együttműködésben, a korábban az ELTE Digitális Bölcsészeti Tanszéke által épített ELTEdata-t vette mintául. Az adatszerkezet kialakításakor a csoport figyelembe vette a különböző adattípusok jellemzőit, valamint a külső rendszerekhez való kapcsolódási lehetőségeket. A rendszerbe első lépésként betöltjük a PIM névterét, majd az ITI Bibliográfiai Osztálya által, részben retrokonverzióval,<sup>7</sup> részben új gyűjtéssel létrehozott Magyar Irodalomtörténet Bibliográfiáját, ami mintegy 130.000 rekordból áll. A MIB névtérre fektetése, a szemantikus kapcsolatok kialakítása automatikusan zajlik a Qulto Kft.-vel együttműködésben. A földrajzi neveket a Wikidata rendszeréből azonosítjuk. Emellett a DigiPhil korábbi könyvtári rendszeréből áttöltjük a Kosztolányi forrásjegyzék, a Sajtótörténeti Bibliográfia, valamint az Arany János Bibliográfia, valamint az Arany kéziratkeresőben található adatokat. Ezen felül a rendszerben kéziratos hagyatékok metaadatai és a szövegek közötti szemantikus kapcsolatok is leképezhetők, míg egy szigorú, a MARC21 szabványt követő könyvtári rendszerben ez elképzelhetetlen lenne. A DigiPhil középtávú tervei között szerepel a TEI XML fájlok metaadatokat tartalmazó fejléceiből kiolvasni az adatokat és azokat is leképezni a rendszerben. Ezzel a lépéssel a kritikai kiadások adatai olyan szemantikus hálózatba rendeződnek, amely nemzetközi szinten is egyedülálló. Az ITIdata egyik legnagyobb előnye, hogy alkalmas heterogén források kezelésére, így olyan tudástár hozható létre, amely hiánypótló az irodalomtudományi kutatások területén.

Az infrastruktúra egy másik pillére a repozitórium, a DigiPhil a jövőben a Cirilo szolgáltatást a CERN által fejlesztett Invenio RDM repozitóriumszoftver váltja. Az Invenio RDM felel majd a TEI XML fájlok tárolásáért és vizualizációjáért. Az új keresőszolgáltatás szintén a repozitóriumban található fájlokat indexeli majd, szemben például az ExistDB-vel, ahol a fájlokat két helyen kellett tárolni. A nyers munkafájlok esetében szintén nagy előrelépés történt a korábbiakhoz képest. Miután a GitHub ingyenessé tette a felhasználók számára a korlátlan privát repozitórium létrehozását, a DigiPhil minden gyűjteményét migrálta a rendszerbe, így a munkafájlok és a publikált kiadások egy helyen érhetők el. Az Invenio RDM és a GitHub összekapcsolása lehetővé teszi a kritikai kiadást készítő munkatársak számára, hogy a munkafájlok vizualizációját egy zárt rendszerben lássák, ami jelentősen megkönnyíti a lektorálási folyamatot.

A DigiPhil új vizualizációs eszköze számos olyan problémára jelent megoldást, amelyeket a korábbi eszközökkel nem lehetett megoldani. A Versioning Machine, amelyet elsősorban az egyes szövegváltozatok egymás melletti megjelenítésére fejlesztettek, a TEI kódolás természetéből fakadóan olyan szövegeket generált, amelyek valójában abban a formájukban sosem léteztek. Az új megjelenítő eszköz egyik legfontosabb funkciója az úgynevezett „Minden verzió nézet”, amely nem bontja szét a szövegváltozatokat, hanem azokat egymás mellé helyezve, az adott szöveghelyen mutatja.

---

6 Az egyik legszemléletesebb példa a FactGrid ([https://database.factgrid.de/wiki/Main\\_Page](https://database.factgrid.de/wiki/Main_Page)), amelyet az Erfurt-i Egyetemen működő Gotha Research Center szolgáltatók történések számára.

7 A Magyar Irodalomtörténet Bibliográfiája tíz kötetes sorozatából két kötet feldolgozása történt meg eddig, a további kötetek egyelőre PDF formátumban érhetők el a honlapon: <http://www.iti.mta.hu/mib/> (hozzáférés: 2022. 09. 20.)



Verzió: Minden szövegváltozat

Összesen 5 verzió

Az ember trag&lt;&lt;ö&gt;&gt;édiája.\* / AZEMBER TRAGÉDIÁJA\*

&lt;&lt;d&gt;&gt; Drámai költemény.\* / Drámai költemény.\* / \*

\* / ÍRTA\*

\* / MADÁCH IMRE\*

\* / KIADTA A KISFALUDY-TÁRSASÁG. C \* / \*

\* / (MÁSODIK TETEMESÉN JAVÍTOTT KIADÁS)\*

\* / PEST.EMICH GUSZTÁV, MAGYAR AKADÉMIAI NYOMDÁSZ\* / PEST. EMICH GUSZTÁV SAJÁTJA.\*

\* / Pest, 1863. Nyomatott Emich Gusztáv magyar akad. nyomdásznál. C \*

\* / AZ EMBER TRAGÉDIÁJA. C \* / AZ EMBER TRAGÉDIÁJA. C \*

1

&lt;&lt;ö&gt;&gt; első szín.\* / ELSŐ SZÍN.\*

A mennyekben. Az úr\* / AZ ÚR\* dicstől környezetben  
 trónján. Angyalok serege\* / ANGYA- SEREGE\* / ANGYALOK SE- REGE\* térden.  
 A négy főangyal\* / NÉGY FŐANGYAL\* a trón mellett áll. Nagy  
 fényesség.\* / Nagy fényesség.\* / C

2. ábra Minta a DigiPhil új vizualizációjából

Ez különösen fontos az informatizált kiadások esetén, ahol a nyomtatott kiadás természete miatt számos eltérést nem jelöltek. Egy másik fontos újítás szintén a szövegváltozatokat tartalmazó kiadások vizualizációjára szolgál. A „Minden verzió nézet (egyszerűsített)” vizualizáció segítségével elrejtethők a kiadást készítő kutatók által nem relevánként megjelölt szövegváltozatok,<sup>8</sup> ezzel könnyítve a befogadást. Szemben a DigiPhil korábbi gyakorlatával, ahol korlátlanul lehetett egymás mellé helyezni a szövegdozokat, a szövegváltozatokat maximum két ablakban lehet együtt olvasni. Ez egyrészt jelentősen felgyorsítja a szolgáltatás sebességét, másrészt egyszerűsíti az olvasást. A rendszerben ezen felül lehetőségünk van olyan annotációk készítésére és megjelenítésére, amelyek a jelenkor szakértőinek a régi, elavult kritikai kiadásokra vonatkozó megjegyzéseit tartalmazza. Új kritikai kiadásokat készíteni a régiak helyett rendkívül sokáig tartana, azonban ezzel a megoldással a tudomány legfrissebb megállapításai is feltüntethetők, anélkül, hogy a kritikai

8 Egyes kritikai kiadásokban előfordulhatnak olyan eltérések, amelyek például a korabeli nyomdai eszközkészlet hiányából fakadnak. Ezeknek a jelölésére, szemben a nyomtatott kiadásokkal, természetesen lehetőség van a digitális kiadásban, azonban az átlagos felhasználó számára ez jelentősen megnehezítené a befogadást.



kiadást reprezentáló TEI XML-t megváltoztatnánk. A GitHub-bal való összekötésen túl, egy egyszerűsített nyomtatási nézet segíti a lektorálást, illetve az összes információ egy lapon történő megjelenítését.

### Az eljátszott öregség

{Ahogy}<sup>1</sup> {A hogy}<sup>2</sup>  
 nőnek az árnyak,  
 Ahogy fogynak az  
 {esték},<sup>1</sup> {esték}<sup>2</sup>  
 {Úgy}<sup>1</sup> {Ugy}<sup>2</sup>  
 fáj jobban és jobban  
 Az eljátszott öregség.

<sup>1</sup> A szöveg státusza: Kritikai kiadás  
<sup>2</sup> A szöveg státusza: Nyugat

Ez az én két vén  
 {szemem}<sup>3</sup> {szemem},<sup>4</sup>  
 Habár sok szépet látott,  
 Mosolygós öreg  
 {úrként}<sup>3</sup> {urként}<sup>4</sup>  
 Nem látja a világot.

<sup>3</sup> A szöveg státusza: Kritikai kiadás  
<sup>4</sup> A szöveg státusza: Nyugat

3. ábra A DigiPhil nyomtatási nézete

A DigiPhil új, rendkívül kifinomult keresőszolgáltatást fejleszt, amelynek több olyan funkciója is lesz, amelyek Magyarországon egyedülállók. A keresőfelületen a metaadatokban való facettás keresésen túl lehetőséget biztosítunk a teljes szövegű keresésre, akár az egyes szövegváltozatokra szűrve, illetve a találati listában megjelenítjük, hogy mely szövegváltozatban olvasható a keresett szövegrész.

luci\*

Összetett keresés

2 találat [↔](#) lista rendezése Alapértelmezett rendezés 10

1. Rózsa és Ibolya [☰](#)

kiemelés [^](#)

Szöveges tartalom:

- o.ajom3-ek.tei.xml(A kritikai kiadás szövege): Lucifer, a hajnalcsillag, lesz. A . J.
- o.ajom3-ek.tei.xml(Kézirat): Lucifer, a hajnalcsillag, lesz. A . J.
- o.ajom3-ek.tei.xml(Pesti Divatlap szövege): Lucifer, a hajnalcsillag, lesz. A . J.

2. Az ember tragédiája, szinoptikus hálózati kritikai kiadás – Második szín. [☰](#)

Szerző: Madách Imre

kiemelés [^](#)

Szöveges tartalom:

- mi-aet.genetikus.ll.tei.xml(Madách kézirata: Madách Imre és Arany János javításaival módosított szöveg): Luci Lucifer..., Luci Lucifer..., Luci Lucifer. A szél gallyakat ráz..., Luci Lucifer..., Luci Lucifer.
- mi-aet.genetikus.ll.tei.xml(Első kiadás): LUCIFER..., LUCIFER..., LUCIFER..., LUCIFER..., LUCIFER.
- mi-aet.genetikus.ll.tei.xml(Második, javított kiadás): LUCIFER..., LUCIFER..., LUCIFER..., LUCIFER..., LUCIFER.

4. ábra A DigiPhil keresőszolgáltatása



Ennek a sok komponensű, a digitális filológia és a szemantikus web elvárásainak megfelelő infrastruktúrát egy könnyed Wordpress alapú weboldal felől érhetjük el, ahol az egyes digitális kritikai kiadások a digitális objektumokból épülnek fel, szerkezetük valójában egy keresés eredménye. Így lehetőségünk van különböző megközelítésmódok biztosítására a felhasználó számára, líra esetében például kronologikus, vagy ciklusokba rendezett tartalomjegyzéket előállítva, mindezt automatikusan. A DigiPhil átállása az új infrastruktúrára folyamatos, a folyamat előreláthatólag 2022 végéig lezárul.

## Mi az a dHUpla? A Digitális Bölcsészeti Platform bemutatása

Mihály Eszter  
Országos Széchényi Könyvtár, Digitális Bölcsészeti Központ  
[mihaly.eszter@oszk.hu](mailto:mihaly.eszter@oszk.hu)

### Abstract

One of the main tasks of the National Széchényi Library's Digital Humanities Centre is to develop an online platform for textual resources and manuscripts held in public collections, providing a common research environment for literary studies, linguistics and other humanities disciplines. It enables the digital publication of a text corpus, the content analysis using natural language processing and other computer tools, the creation of data visualisations, and the processing and publication of born digital text content.

The service also makes available the wide knowledge of other institutions through the databases linked to the platform (namespaces, bibliographic databases) and facilitates the international integration of the field (Europeana).

The dHUpla platform ([dhupla.hu](http://dhupla.hu)) makes a part of the Hungarian cultural heritage accessible by digitising and computerising the national data assets and archiving them in a most advanced way for long-term preservation. The contemporary research and user interface also includes digital philological recommendations for publishing the content of scientific projects, responds to the challenges of the 21st century transformation of public collection functions.

This paper presents the development of the platform so far and its ongoing projects.

**Keywords:** digital humanities, digital philology, digital text edition, searching, metadata, text-image linking, manuscript, text corpus, data visualisation, corpus search, annotation.

### Absztrakt

Az Országos Széchényi Könyvtár Digitális Bölcsészeti Központ egyik elsődleges feladata egy korszerű online platform fejlesztése a közgyűjteményekben őrzött szöveges források számára, amely egységes kutatói környezetet jelent az irodalomtudomány, a nyelvtudomány, és más humán tudományok számára. Lehetőséget teremt digitális szövegtörzsek publikálására, a platformon elérhető tartalmakon nyelvtechnológiai és egyéb számítógépes eszközökkel történő elemzések elvégzésére, adatvizualizációk létrehozására, illetve ún. born digital (digitálisan született) szöveges tartalmak feldolgozására és közzétételére is.

A szolgáltatás biztosítja a más intézményekben fellelhető tudás széles körű újrahasznosítását is a platformmal kapcsolatban álló adatbázisok révén (névterek, bibliográfiai adatbázisok), illetve elősegíti a szakterület nemzetközi integrációját (Europeana).

A dHUpla platform ([dhupla.hu](http://dhupla.hu)) hozzáférhetővé teszi a magyar kulturális örökség eddig elérhetetlen rétegét a nemzeti adatvagyon digitalizálásával és gépi feldolgozásra alkalmassá tételével, továbbá a legkorszerűbb módon

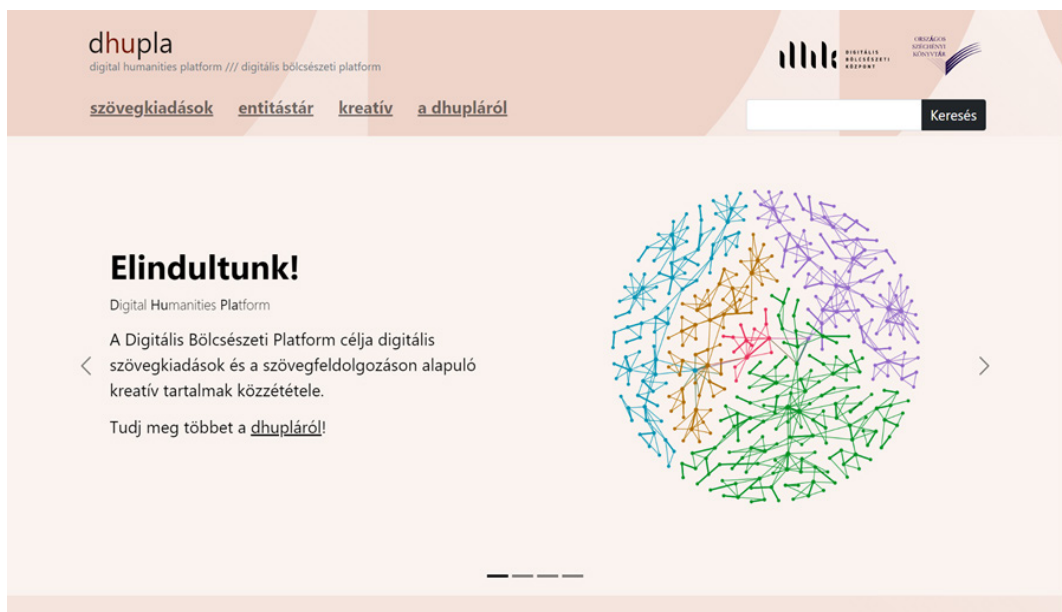


archiválja azt a hosszútávú megőrzés céljából. A digitális filológiai ajánlásokat is tartalmazó, tudományos projektek tartalmainak publikálására is lehetőséget nyújtó korszerű kutatói és felhasználói felület a közgyűjteményi funkciók átalakulásának 21. századi kihívásaira reagál.

Jelen tanulmány a platform fejlesztésének eddigi eredményeit és folyamatban lévő projektjeit mutatja be.

**Kulcsszavak:** digitális bölcsészet, digitális filológia, digitális szövegkiadás, keresés, metaadatok, text-image linking, kézirat, szövegtörzs, adatvizualizáció, korpuszlekérdezés, annotálás.

## Elsődleges profil



1. ábra: A dhUpla nyitóoldala

A szolgáltatás kialakításának alapvető célja eddig kiadatlan, nehezen hozzáférhető vagy valamilyen szempontból kiemelt jelentőségű, a tudományos-kulturális érdeklődés középpontjában álló, illetve éppen méltatlanul elhanyagolt *kézírtos* források digitális feldolgozása, közzé- és kutathatóvá tétele. A kézírtatok újszerű, pontos definícióját mellőzve<sup>1</sup> azt mondhatjuk, jelenleg ide soroljuk többek között a művek szerzői kéziratait, fogalmazványokat, naplókat, levelezéseket, feljegyzéseket stb., amelyek ma már természetesen szülehetnek eleve digitális formában is. A platform alkalmas nyomtatásban már megjelent szövegek digitális kiadására is, de nem ez a szöveges források elsődleges feldolgozandó köre. Vannak természetesen kivételek, mint például Kassák Lajos folyóiratainak publikálása: <https://dhupla.hu/collection/kassak-lajos-a-tett>.

## Publikáció

Már tavalyi előadásomban is felvettem, hogy a források digitális másolata létrehozásának általános gyakorlatával az is kérdésessé vált, hogy a közzététel mely szintjét nevezhetjük

1 Meglepő módon sem a magyar nyelvű, sem a nemzetközi szakirodalomban nem találtam közmegegyezésen alapuló meghatározást. E téma kibontása természetesen külön tanulmány keretein belül lehetséges.

publikációnak. Annyi bizonyos, hogy a szövegkiadás egyes szintjei mára a közgyűjtemények hatáskörébe (is) kerültek. A GLAM szektor és a felhasználók részéről is – teljes joggal – egyre gyakoribb az igény, hogy a szöveges tartalmak ne képként (digitális faksimile), hanem digitális szöveggként legyenek elérhetőek az online térben. A Digitális Bölcsészeti Központ (DBK) éppen ezért elsősorban digitális forráskiadásokat készít, illetve kiemelt projektek esetében digitális kritikai kiadásokat.

### Infrastruktúra főbb elemei

A kiépített infrastruktúra középpontjában egy verziókövető rendszer áll (GitLab),<sup>2</sup> amelyben a tárolt anyagok minden változása visszakövethető, s innen indul ki minden automatizált folyamat, ezzel áll kapcsolatban közvetlenül vagy közvetve az összes használt eszköz.

Az eredetileg tervezett komponensek közül az XML-adatbázist végül egyelőre mellőztük, ugyanis a gyakorlati tapasztalatok azt mutatták, hogy a jelenleg elvárt funkciók enélkül is mind megoldhatóak, s így tovább egyszerűsödik az informatikai háttér felépítése, ami fejlesztői és felhasználói szempontból is öröndetes.

Nagy előnye a git alapú rendszernek, hogy gyakorlatilag teljesen szoftverfüggetlen, így a közgyűjteményi szféra különböző infrastruktúráival könnyen kompatibilissé tehető. A szövegek elvárt formátuma TEI XML, amely nemzetközileg is a legáltalánosabban használt szövegjelölő nyelv, így a legtöbb projekt befogadhatóvá válik. A moduláris kialakítás továbbá lehetővé teszi, hogy a munkafolyamat különböző pontjain is be lehessen csatlakozni: ha például egy projekt nem a dHUpla szerkesztőségi környezetében, hanem más módon vagy más szoftverekkel szeretné előállítani a publikálandó tartalmakat, akkor megteheti, hogy a kész XML fájlokat adja át a Git-be való betöltésre.

### Szerkesztő program 1. Transkribus - Digitális átírat

A kéziratok kézi és automatikus gépi átírásához, valamint az ún. text-image linking (kép és szöveg összekötése) elvégzéséhez a Transkribus<sup>3</sup> programot használjuk, amely TEI XML formátumban is exportálható. A kézi átírás ingyenes, a mesterséges intelligencián alapuló automatikus kézírásfelismerő (Handwritten Text Recognition) modell építése is, a létrehozott modellel pedig 4-500 kézírásos oldal ingyenesen felismertethető, utána fizetnünk kell érte. Ennek fejében viszont készen kapunk egy eszközt az automatikus gépi 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. Az első publikus magyar nyelvű kézírásfelismerő modellt a Digitális Bölcsészeti Központ 2022. március 4-én tette nyilvánossá.<sup>4</sup>

Ahhoz, hogy a jövőben ezen a szakterületen tömeges feldolgozás vehesse kezdetét, a különböző magyar nyelvű projekteknek, amelyek automatikus kézírásfelismertést használnak vagy használnának, mindenképp össze kell fogniuk. Egyrészt azért, hogy a saját szövegtörzsekön betanított *modelleket egymásba építve* egyre általánosabb érvényű eszköz jöjjön létre a magyar kéziratok digitalizációjához, másrészt azért, hogy a tömeges adatfeldolgozás lebonyolításához megfelelő szakmai, személyi és anyagi feltételek

---

2 GitLab, hozzáférés: 2022. 09. 19., <https://about.gitlab.com/>.

3 Transkribus, ReadCoop, hozzáférés: 2022. 09. 19., <https://readcoop.eu/transkribus>.

4 *Hungarian handwriting 19th–20th cent.*, ReadCoop, hozzáférés: 2022. 09. 19., <https://readcoop.eu/model/hungarian/>. A Transkribus, illetve a HTR használatáról és eredményeiről dr. Szűcs Kata kötetben szereplő tanulmánya számol be részletesen.



mellett megvalósítható *intézményközi vagy akár nemzetközi megállapodásokat* tudjanak kötni az olyan fejlesztői műhelyekkel, mint a Transkribus.

### 1.1. Biztonság

A Transkribus programmal lehetséges a saját gépen való szerkesztés is, de az összes funkció kihasználásához a szerverre is fel lehet tölteni a dokumentumokat, amely szintén verziókezelővel biztosítja a tárolásukat. A számos felhasználói funkció teljes igénybevétele miatt a projektek során a szerveren dolgozunk, ezért saját fejlesztés keretében megoldottuk a Transkribus-gyűjtemények automatikus napi mentését saját szerverre is.

### Szerkesztő program 2. Oxygen XML Editor - TEI XML

A dHUpla oldalára szánt szövegeket a DBK által készített TEI XML specifikációk szabályai szerint kell létrehozni. Mivel a TEI igencsak rugalmas XML formátum, az általános elvekhez képest meglehetősen szigorú szabályrendszer létrehozására volt szükség ahhoz, hogy az annotációk alapján a platformon való megjelenítés, kereshetőség, lekérdezhetőség stb. jól működjön. Az útmutató külön foglalkozik a metaadatok, illetve a szövegtest tagelési módjával. A magyar nyelvű digitális szövegkiadások, azon belül a különböző kategóriákba sorolt szövegtípusok jelölési útmutatója hamarosan bárki számára elérhető lesz a DBK oldalán.

ATEI XML-ek további szerkesztésére az Oxygen XML Editor<sup>5</sup> programot használjuk. Ezen belül a szerkesztőségi környezet optimalizálása céljából kialakítunk egy ún. frameworköt, amely a dHUpla igényeire van szabva, annak minden szempontját figyelembe véve. A framework egyrészt kiterjedt textológiai-filológiai jelölésrendszer alkalmazására nyújt lehetőséget anélkül, hogy a szerkesztőnek a technológiai hátteret és az XML világot ismernie kéne, másrészt az adatgazdagítás, metaadatok részletes kitöltése, külső adatbázisok bekapcsolása, szerkezeti és szemantikus annotációk alkalmazása, alapszinten a text-image linking mind felhasználóbarát módon megoldható ebben a környezetben. A létrejövő TEI XML fájl végül tulajdonképpen önmagában egy strukturált adathalmazt képez, amely a legkülönfélébb módokon lesz feldolgozható (ld. kreatív tartalmak),<sup>6</sup> lekérdezhető.

A framework- és a git alapú funkciók teszik lehetővé a dHUpla rendszerének egyik legfontosabb elemét, a különböző automatizált műveletek beépítését (pl. automatikus transzformációk, azonosítókiosztás stb.), vagy éppen a szerkesztőségi folyamatok felügyeletét, irányítását.

### 2.1. Framework-funkciók

Az alább ismertetett funkciók jelenleg tesztelési fázisban vannak.

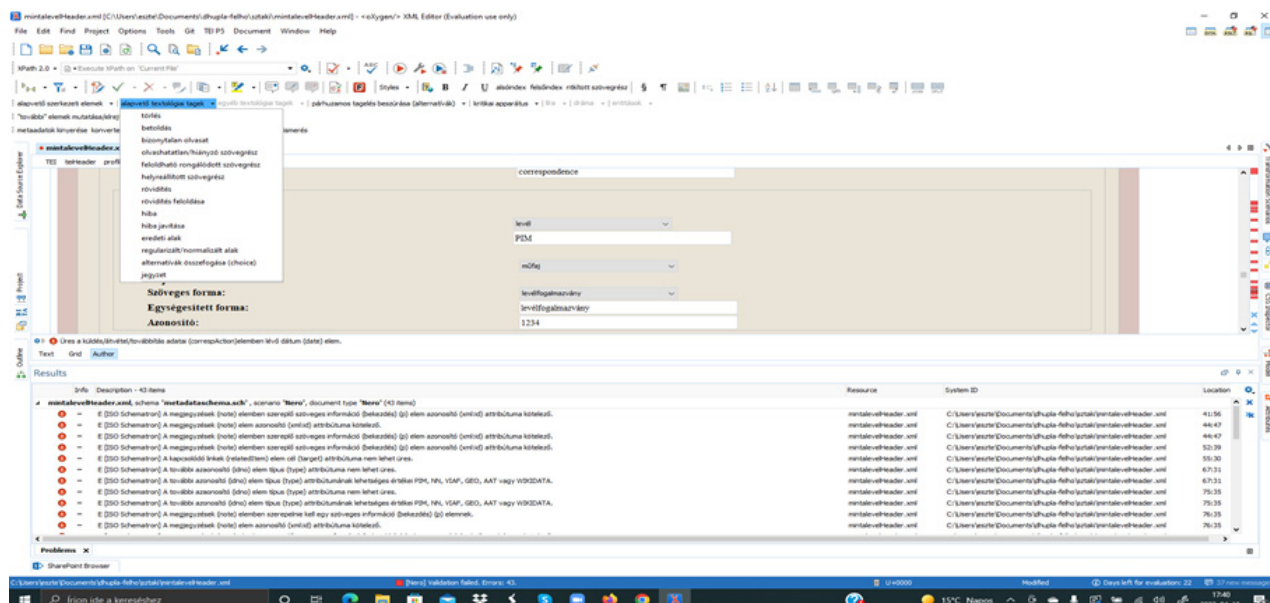
1. A szerkesztőfelületet igyekszünk úgy kialakítani, hogy a folyószöveg minél jobban elkülönüljön a TEI tagektől és azok attribútumaitól, hogy jól olvasható és áttekinthető legyen, ugyanakkor az annotáció hozzáadása vagy szerkesztése is egyszerű maradjon. Ezért a tagek csak akkor látszódnak, ha a felhasználó szeretné, továbbá a szerkesztő magyar terminusokkal dolgozhat, legördülő listákból választhat, gombnyomással annotálhat, felajánlott opciókból válogathat. Az ún. TEI fejléctet, tehát a metaadatokat jól strukturált űrlap kitöltésével lehet létrehozni.

<sup>5</sup> XML Editor, Oxygenxml, hozzáférés: 2022. 09. 19., [https://www.oxygenxml.com/xml\\_editor.html](https://www.oxygenxml.com/xml_editor.html).

<sup>6</sup> Kreatív tartalmak, dHUpla, hozzáférés: 2022. 09. 19., <https://dhupla.hu/page/kreativ/>.



2. A validáláshoz, illetve a felhasználó számára felajánlott segítő funkciók meghatározásához ún. sémafájlokat (RELAX NG, Schematron) alkalmazunk, amelyek az aktuális nemzetközi TEI szabvány mellett saját szabályrendszerünk szerint is definiálják és ellenőrzik a dokumentumokat. A Schematron sémával való ellenőrzéskor a felhasználó általunk megfogalmazott szöveges hibaüzenetet kap, ha nem megfelelő a tagelés (pl. „Hiányzik a bekezdés azonosítója”), vagy figyelmeztetheti olyan gyanús esetre is, amely adott szituációban lehet helytálló is (pl. „A levél mérete nem mm-ben van megadva. Biztos jó így?”). Mindez nagyban megkönnyíti a helyes formátumú TEI XML fájl kialakítását, hiszen kiszűri az emberi szemmel rendkívül nehezen, nagyobb mennyiségű szövegfeldolgozás esetén pedig egyáltalán nem felfedezhető tévedéseket.



2. ábra: A szerkesztőfelület

3. A különböző szövegtípusokhoz létrehozunk ún. template-fájlokat, amelyek egy új dokumentum készítéséhez szükséges alapvető elemeket tartalmazzák, ebből indulhat ki a szerkesztő minden új szöveges objektum létrehozásakor, így soha nem kell nulláról kezdenie a munkát, a keret már eleve adott lesz hozzá.

#### 4. Automatizált műveletek

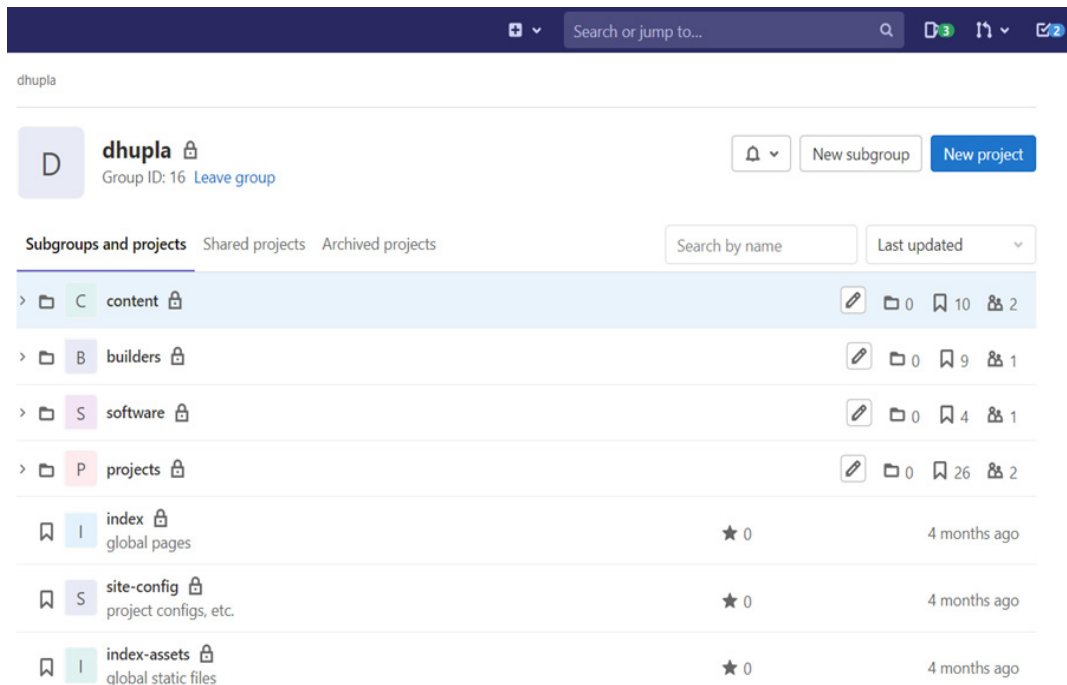
- a) Automatikus névelemfelismerés: egy adott szövegben vagy szövegrészben egy gombnyomásra lefuttathatjuk a mesterséges intelligencián alapuló ún. Named Entity Recognition<sup>7</sup> funkciót, amely felismeri az abban szereplő tulajdonneveket (személynév, helynév, testületi név, műcím), és el is látja azokat a megfelelő annotációkkal<sup>8</sup>.
- b) Entitásazonosítás: az ún. entitásokat (ld. előző pont, azonkívül ún. terminusok: pl. szakkifejezések, köznevek), valamint akár bibliográfiai tételeket is beazonosíthatjuk külső névterekben, adatbázisokban való közvetlen kereséssel, s egy kattintással elláthatjuk a megfelelő egyedi azonosítóval, valamint megadhatjuk az egységesített névformát is.

7 Named-entity recognition, Wikipedia, hozzáférés: 2022. 09. 19., [https://en.wikipedia.org/wiki/Named-entity\\_recognition](https://en.wikipedia.org/wiki/Named-entity_recognition).

8 A jelenlegi funkció a huSpacy (<https://github.com/huspace/huspace>) NER moduljára épül, mivel a magyar nyelvű írott normához közeli szövegtörzsek esetén jelenleg ez bizonyult a leghatékonyabbnak, ugyanakkor folynak párhuzamos fejlesztések e téren, tehát a modul felépítése még változhat.



- c) Automatikus transzformációk: jelenleg a Transkribusban létrehozott anyagokhoz készült ilyen funkció, amely az onnan exportált szövegeket automatikusan átalakítja a dHUpla számára valid TEI XML fájlkká.
  - d) PDF előállítás: a TEI XML-ekből egyedileg és tömegesen is létrehozhatunk PDF formátumú fájlokat, amelyek a born digital szövegkiadásból nyomtatott kiadás előkészítésére alkalmasak. A PDF elrendezését, formázását a DBK határozza meg.
  - e) Automatikus metaadatkinyerés adatvizualizációkhoz: jelenleg a digitális levelezéskiadásokhoz készült funkció, amellyel egy gombnyomással kinyerhető egy teljes szövegtörzs összes releváns metaadata, s máris felhasználható a legkülönbözőbb adatvizualizációk, statisztikák készítéséhez.
5. 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.
- 6.



3. ábra: A GitLab-repozitórium

A Git verziókezelő rendszer lát el gyakorlatilag minden olyan feladatot, amelyre a publikáció menedzseléséhez szükség van.

1. Repozitóriumként (repository) funkcionál, tehát nincs szükség külön adatbázis használatára és karbantartására, ami rendkívüli módon leegyszerűsíti az üzemeltetési és supportálási munkát és a munkafolyamatot is.
2. A verziókezeléssel megoldható, hogy egy szöveg életét a kezdetekig visszakövessük, megtaláljunk egy folyamatban elkövetett hibát, vagy akár visszaállítsunk egy korábbi verziót. Hosszútávú terveink közé tartozik, hogy a felhasználók számára is elérhetővé tesszük a korábbi ún. release<sup>9</sup>-ek verzióit.

<sup>9</sup> Egy teljes szövegtörzs újabb kiadásnak tekintett változata.

3. A szerkesztők párhuzamosan is végezhetik a munkájukat, akár ugyanazon a dokumentumon is, ilyenkor a változtatások merge-ölésével (összegyűrésével), illetve az ún. konfliktusok feloldásával lehet elmenteni a közös változtatásokat.
4. Mivel az Oxygen programnak van Git kliense, a felhasználók közvetlenül dolgozhatnak a Git-be, minden változást azonnal látnak magában az XML-szerkesztő programban.
5. A különböző szerepkörökhöz különböző szintű jogosultságokat osztunk ki, így meghatározhatunk pl. szerkesztői, főszerkesztői, lektori tevékenységekhez tartozó engedélyeket.
6. Az általunk használt Gitlab ún. issue tracker (feladatkövető) használatával írjuk össze és követjük nyomon az egyes anyagokkal kapcsolatos teendőket, hibajavításokat, tesztelési feladatokat, megfelelő személyhez rendeljük, határidőt szabunk a feladathoz.
7. Ún. pipeline-ok<sup>10</sup> alkalmazásával megoldhatóak az olyan automatizált eljárások, amelyeknek mindig a munkafolyamat egy adott pontján kell lezajlaniuk. Ilyen például az XML well-formedness (jól formáltság) ellenőrzése betöltéskor, vagy az állandó azonosítók kiosztása objektum szinten. Ha az automatikus validáció hibát jelez, fejlesztésünk segítségével a projekt-admin email-ben értesítést kap, s a hibás tartalmat nem lehet betölteni, amíg ki nem javítják.
8. A Git repositoryban egy helyen tárolhatjuk a tartalmakat és a programkódokat, mindent ugyanott kezelhetünk, átlátható struktúrába rendezve.
9. Az ún. build-művelet, amelynek során a TEI XML-ből előáll a honlapon publikált HTML tartalom, teljes egészében automatizált, semmilyen manuális munkára nincs hozzá szükség.
10. A docker-technológiával<sup>11</sup> való összehangolásnak köszönhetően lehetővé vált a projektszintű kezelhetőség. Ez azt jelenti, hogy egy adott projekt bármilyen szoftverrel, bármilyen módon előállíthatja a TEI XML fájlokból a megfelelő HTML-eket, s beépítheti az adott projekt build-folyamatába.
11. A különböző típusú szövegkiadásokat (gyűjteményeket) kategóriákba soroltuk, amelyek beállításai testreszabhatóak egyszerű szöveges fájlok (YAML) kitöltésével. Itt adhatjuk meg például, hogy az adott kiadásban milyen szűrési lehetőségek jelenjenek meg, vagy milyen szempontok alapján lehessen rendezni az objektumokat. A konfigurációs fájlokban bármikor új kategóriákat is meghatározhatunk.

---

<sup>10</sup> CI/CD pipelines, GitLab, hozzáférés: 2022. 09. 19., <https://docs.gitlab.com/ee/ci/pipelines/>.

<sup>11</sup> Docker (software), Wikipedia, hozzáférés: 2022. 09. 19., [https://en.wikipedia.org/wiki/Docker\\_\(software\)](https://en.wikipedia.org/wiki/Docker_(software)).

## A feldolgozás státuszai



4. ábra: A staging server

A szerkesztés alatt álló szövegkiadások megjelenésének és működésének ellenőrzésére létrehoztunk egy belső felhasználású ún. staging-szervert, amely az éles szolgáltatás másolata azzal a különbséggel, hogy itt nem csak a publikusra beállított tartalmak,<sup>12</sup> hanem minden a repozitóriumban található tartalom megtekinthető. Így a szövegkiadásban résztvevő munkatársak már a tartalomelőállítás folyamata közben is láthatják eredményeiket, a javítandó hibákat, illetve jelezhetnek informatikai teendőket is. Ezen a felületen tesztelhetünk újabb fejlesztéseket is akár olyan szövegkiadásban, amely célzottan csakis erre lett létrehozva (lásd: Demo gyűjtemény).

<sup>12</sup> A gyűjtemények, illetve a bennük szereplő objektumok státusza lehet draft (szerkesztés alatt), withdrawn (visszahívott), finished (kész) és published (publikus).

## A publikáció alapegységei

Digitális kritikai kiadás

# Móricz Zsigmond levelezése (1892–1913)

2.0 verzió

Az alap kutatás célja Móricz Zsigmond levelezésének (1892–1913) forrásfeltárása és közzététele digitális kritikai kiadás formájában. A projekt az NKFIH támogatásával valósult meg. [Tovább olvasom.///](#)

elejére vissza [tovább](#) [végére](#) 1-30/1526

[régebbiek elől](#) [újabbak elől](#) [ábécé szerint](#)

keresés

keresés a gyűjteményben

szűrési feltételek törlése

feladó-címzett	típus	keletkezés
<a href="#">Móricz Zsigmond – A Debreceni Református Kollégium tanári kara (1892-06-20)</a>	levél	1892-06-20
<a href="#">Pallagi Erzsébet – Móricz Zsigmond (1894-03-23)</a>	levél	1894-03-23
<a href="#">Móricz Zsigmond – Hajdu István (1894-09-10)</a>	levél	1894-09-10
<a href="#">Móricz Zsigmond – Pallagi Gyula (1896-09-03)</a>	verses levél	1896-09-03
<a href="#">Heeger János – Móricz Zsigmond (1898-06-16)</a>	levél	1898-06-16
<a href="#">Heeger János – Móricz Zsigmond (1898-06-19)</a>	levél	1898-06-19
<a href="#">Heeger János – Móricz Zsigmond (1898-07-16)</a>	levél	1898-07-16
<a href="#">Móricz Bálint, Móricz István, Pallagi Erzsébet – Móricz Zsigmond, Móricz Dezső (1899 után)</a>	levél	1899 után
<a href="#">Heeger János – Móricz Zsigmond (1899-03-18)</a>	levél	1899-03-18
<a href="#">Balázs József – Móricz Zsigmond (1899-08-25)</a>	levelezőlap	1899-08-25

**TÍPUS** ^

- levél (903)
- képeslap (249)
- levelezőlap (227)
- levélfogalmazvány (40)
- boríték (24)
- névjegy (16)

**SZERZŐ / LEVÉLÍRÓ** ^

- Móricz Zsigmond (495)
- Holics Janka (121)
- Móricz Dezső (77)
- Szklanár Teréz (69)
- Móricz István (58)
- Pallagi Gyula (43)

**FELADÁS HELYE** ^

5. ábra: A Móricz Zsigmond-levelezése gyűjtemény nyitóoldala

A dHUpla kiadásai alapvetően háromféle szerkezeti egységből épülnek fel. A legnagyobb egység az ún. gyűjtemény, amely egy teljes szövegkiadást foglal magában. Magához a gyűjteményhez is tartozik egy TEI XML fájl, amely tartalmazza a gyűjtemény metaadatait, leírását, állandó azonosítóját stb. Egy gyűjtemény lehet egyetlen regény, de akár egy teljes levelezés kiadása is. Ebben találhatóak az ún. digitális objektumok, azaz a kiadáshoz tartozó szöveges egységek TEI XML formátumban, valamint az ún. információs oldalak, amelyek Markdown vagy HTML formátumúak lehetnek.<sup>13</sup>

### Megjelenítés, szűrés, keresés, rendezés, hálózatoság

Az általunk alkalmazott text-image linking szövegfeldolgozó módszer rendkívül jól használható minden felhasználói réteg számára. Ennek lényege, hogy a digitális faksimile zónái össze vannak kapcsolva az átírt szöveggel, tehát a felhasználó együttesen láthatja a kettőt. A dHUplán többféle nézet állítható be, attól függően, hogy soronként vagy kisebb kontextusban szeretnénk látni az éppen olvasott szövegrész fotón megfeleltethető részét, de arra is lehetőség nyílik, hogy a képernyőt kettéosztva párhuzamosan olvassuk az eredeti kéziratot és az átíratát.

<sup>13</sup> Ilyenek például a kreatív tartalmak oldalai is, de létre lehet hozni bármilyen kiegészítő jellegű, nagyobb terjedelmű járulékos tartalomhoz.





6. ábra: Kép és szöveg

A szövegben szereplő annotációk jól láthatóvá válnak azáltal, hogy kiemeltük az annotált szövegrészeket, emellett igyekeztünk jól olvashatóvá tenni magát az annotáció tartalmát is (pl. textológiai-filológiai annotációk, tárgyi és magyarázó jegyzetek, entitások beazonosítása).



7. ábra: Az annotációk megjelenítése

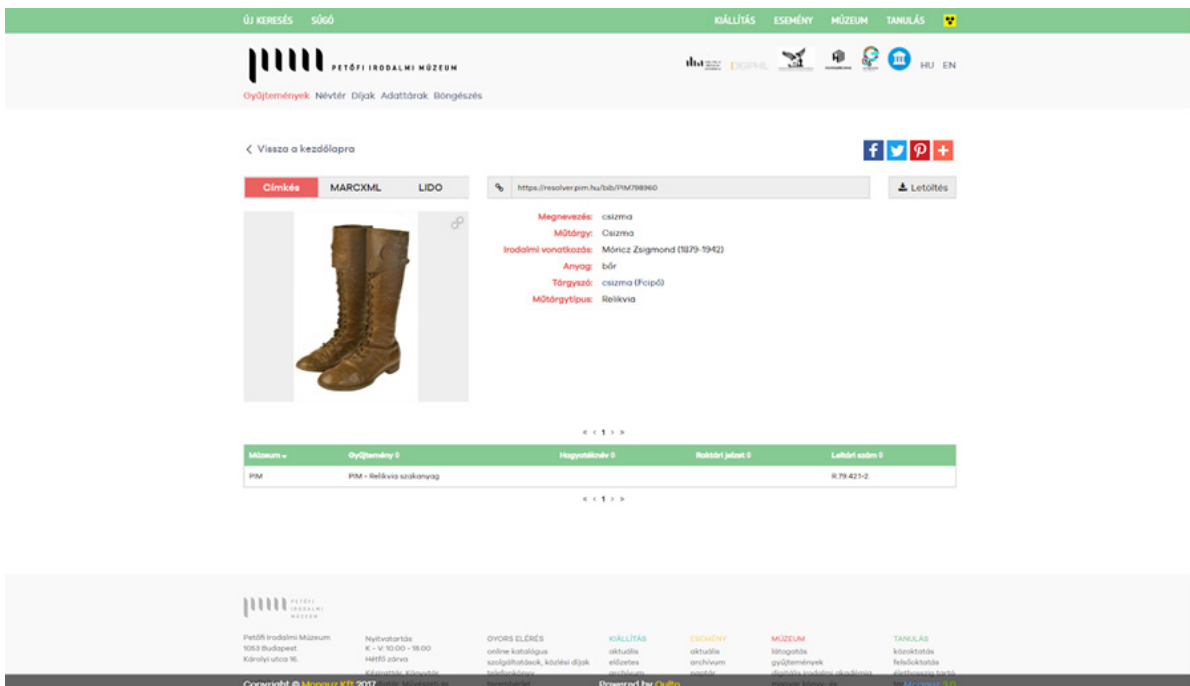
A szöveg a digitális térben menthetetlenül az adatok hálójába kerül, ami ez esetben nem csapdát vagy megkötöttséget, hanem éppen a határtalan szabadságot teremti meg számára. Jó példa erre Móricz Zsigmond levele, amelyben csizmájáról mesél:





8. ábra: Móricz Zsigmond levele

Afelhasználó egy kattintással tovább léphet a Petőfi Irodalmi Múzeum Relikviagyűjteményébe, és megtekintheti a szóban forgó csizmát:



9. ábra: Móricz Zsigmond csizmája a PIM Gyűjteményi katalógusból

Ehhez hasonló és még jóval komplexebb adatkapcsolatok építhetők ki a különböző adattárakkal, online tartalmakkal. A fent említett entitásazonosítások, bibliográfiai hivatkozások révén szintén szerteágazó adathalmazban találja magát az érdeklődő felhasználó.



Mindemellett a dHUpa oldalán belül is létrehoztunk egy ún. Entitástárat (<https://dhupla.hu/collection/nevter>), amely kifejezetten a platformon elérhető szövegekben szereplő nevek a kiadás szempontjából releváns adatait tartalmazza. Az egyes entitások itt is kapnak egyedi azonosítót, de természetesen összeköthetőek külső névterekkel, adatbázisokkal.



10. ábra: a dHUpa Entitástára

A keresési, szűrési műveleteket ún. facetek használatával tettük egyszerűvé, támogatóvá. A facetek a dokumentum metaadatai szerint is működhetnek, de a szövegben szereplő annotációk szerint is megadhatunk szűkítési lehetőségeket (pl. íróeszköz, említett műcímek szerint). Szabadszavas keresést egy gyűjteményen belül, de a teljes dHUpa tartalmaira is indíthatunk. A Solr<sup>14</sup> alapú keresőbe szótövelő eszközt is beépítettünk, hogy a keresett kifejezés összes toldalékolt változata is megjelenjen a találatok között.

14 Solr, hozzáférés: 2022. 09. 19., <https://solr.apache.org/>.

előjére vissza tovább végére 1-6/6 relevancia szerint régebbiek elől újabbak elől abcé szerint csizma

keresés a gyűjteményben

szűrési feltételek törlése

**TÍPUS** ^

levél (5)  
képeslap (1)

**SZERZŐ / LEVÉLÍRŐ** ^

Móricz Zsigmond (2)  
Ady Lajos (1)  
Holics Janka (1)  
Móricz István (1)  
Szilágyi Julianna (1)

**FELADÁS HELYE** ^

n.a. (5)  
Budapest (1)

**CÍMZETT** ^

Móricz Zsigmond (4)  
Holics Janka (2)

**ÁTVÉTEL HELYE** ^

n.a. (5)  
Budapest (1)

**KELETKEZÉS HELYE** ^

n.a. (1)  
Barszentdéd (1)  
Budapest (1)  
Kisújszállás (1)  
Kölköc (1)

képeslap  
**Ady Lajos – Móricz Zsigmond (1913-06-18)**  
Kézsók – az Asszonytól csók – meleg köszönet, bő szüret (majd őszre) és hozzá illő nagy **csizma!** Igaz szeretettel: Ady Lajosok

levél  
**Szilágyi Julianna – Móricz Zsigmond (1909-09-22)**  
Higye meg anyi bajom van – az a kis pénz a mit kapok ha kedves Apám a lakást és a háztartást nagy részbe nem fizetné nem tudom mit tennék, most megint kura alá kell jönnöm, a kislíknak ruha **csizma** kell, nekem sincs igazán meleg ruhám most azt beszerezni és miből, igazán nagyon szegénylenem de magát akarom megkérni ha kölcsönözne nekem januárig 100 vagy 150 frntot forintot a kamatját is fizetném decemberbe a mit kapok aból mindjárt le fog hatna lefoghathna 50 forintot.

levél  
**Holics Janka – Móricz Zsigmond (1904-12-24)**  
– Pali nem tud egyebet hirtelen kitalálni, hát Jenő : Egy csizmatalpat Pali : Ha még csak azt, egy egész **csizmat** huzzunk ki. – Hát ez nem furcsa?

levél  
**Móricz István – Móricz Zsigmond (1902-08-25)**  
De csak nekik rohan Dühös Varju Gábor Lazsnakolja őket Dühös Varju Gábor. Mikor oszt elfárad az útlegelesbe **Csizma** szára mellől Kést kap a kezébe. Dühös Varju Gábor Meg szurkálja őket Véres mán a kése De szurkálja őket.

levél  
**Móricz Zsigmond – Holics Janka (1904-08-29)**  
Édes Egyem, ne haragudj én rám, Soha, soha, Ha a sorsom, a mely ily bajt mért rám, – Oly mostoha: Irót, költőt, írka-fírka gyártót, Én belélem ilyen mestert gyártott: Hisz tudod, Jól tudod: A csizmadiának rongyos a **csizmája**. A kertésznek gyomos az ó háza tája, A kovácsnak nincsen patkó a **csizmáján**. A doktornak pedig baj lakozik máján, A szegény írónak, Rongyos firkálónak Soha sincs ideje sirni hogy ha fáj ám!

levél  
**Móricz Zsigmond – Holics Janka (1910-04-26)**  
Ma egész nap ezen ábrándoztam, hogy te kis gazdasszony volnál, én meg totyakos, terebélyes gazda. Vadász **csizmat** huznék, lovagolnék s zöld kalapom lenne gazdászosi! A tokámat három rófosé ereszteném s ha megunnánk az idillt, hip, hop – egy ugrás Pest , rögtön ott volnánk kultur embernek.

11. ábra: Keresés és szűrés

Egy gyűjtemény objektumai több szempontból rendezhetőek, s ahogy fentebb már szó volt róla, az aspektusok az igények szerint konfigurálhatóak:

Digitális kritikai kiadás

Móricz Zsigmond levelezése (1892–1913)

2.0 verzió

Az alapkutatás célja Móricz Zsigmond levelezésének (1892–1913) forrásfeltárása és közzététele digitális kritikai kiadás formájában. A projekt az NKFIH támogatásával valósult meg. [Tovább olvasom. ///](#)

előjére vissza tovább végére 1-30/1526 régebbiek elől újabbak elől abcé szerint keresés

keresés a gyűjteményben

szűrési feltételek törlése

**TÍPUS** ^

levél (903)  
képeslap (249)  
levelezőlap (227)  
levélfogalmazvány (40)  
boríték (24)  
névjegy (16)

**SZERZŐ / LEVÉLÍRŐ** ^

Móricz Zsigmond (495)  
Holics Janka (121)  
Móricz Dezső (77)  
Szklenár Teréz (69)  
Móricz István (58)  
Pallagi Gyula (43)

**FELADÁS HELYE** ^

feladó-címzett	típus	keletkezés
<a href="#">A Jövendő – Móricz Zsigmond (1910-04-22)</a>	levél	1910-04-22
<a href="#">Ács Emma – Móricz Zsigmond (1902-12-31)</a>	képeslap	
<a href="#">Ács Emma – Móricz Zsigmond (1903-05-11)</a>	levél	1903-05-11
<a href="#">Ács Vilma – Móricz Zsigmond (1900-09-27)</a>	képeslap	1900-09-27
<a href="#">Ady Endre – Móricz Zsigmond (1913-07)</a>	képeslap	
<a href="#">Ady Lajos – Móricz Zsigmond (1913-06-18)</a>	képeslap	
<a href="#">Ady Lajos – Móricz Zsigmond (1913-07-02)</a>	képeslap	1913-07-02
<a href="#">Ady Lajos – Móricz Zsigmond (1913-08-15)</a>	képeslap	1913-08-15
<a href="#">Ady Lajos – Móricz Zsigmond (1913-12-15)</a>	levél	1913-12-15
<a href="#">Ágai Béla – Móricz Zsigmond (1906)-08-18</a>	névjegy	[1906]-08-18
<a href="#">Árkosi-Ferenczi Kálmán – Móricz Zsigmond (1911-03-25)</a>	levél	1911-03-25

12. ábra: A digitális objektumok rendezése



## Adatvizualizációs lehetőségek

A metaadatok, illetve a szövegekben elhelyezett annotációk lekérdezésével és strukturált feldolgozásával statisztikák, vizualizációk végtelen sora állítható elő az egyes szövegkorpuszokról. Ilyenek többek között a dHUpla oldalán (<https://dhupla.hu/page/kreativ/>) található kapcsolati hálók ábrázolása, családfa-ábra, interaktív tartalmak (pl. virtuális séta), nyelvtechnológiai elemzések vizualizációja (pl. szófelhő), stilometriai vizsgálatok.

Legtöbb esetben a vizualizációk a dHUpla tartalmaira épülnek, de akadnak kivételek is, mint például az ukrán háborúról szóló online magyar sajtó szóhasználatának vizsgálata különböző módokon (<https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet><sup>15</sup>).

Mindezek a kreatív tartalmak arra hivatottak, hogy a közzétett magyar kulturális kincseket a felhasználók, kutatók egészen újszerű perspektívából tudják megközelíteni a források digitális közegbe áthelyezett új élete során.

---

15 Lásd még: *Magyar hírportálok orosz-ukrán háborús tartalmainak interaktív vizualizációja*, dHUpla, hozzáférés: 2022. 09. 19., <https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet-interaktiv>; Általános statisztikák, dHUpla, <https://dhupla.hu/page/kreativ/ukrajna-hirek-szokeszlet-diagram>.

## Szemantikus névelem-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)

Nemeskey Dávid Márk – Palkó Gábor  
 Eötvös Loránd Tudományegyetem Bölcsészettudományi Kar Történeti Intézet,  
 Digitális Bölcsészet Tanszék, Digitális Örökség Nemzeti Laboratórium  
[nemeskey.david@btk.elte.hu](mailto:nemeskey.david@btk.elte.hu), [palko.gabor@btk.elte.hu](mailto:palko.gabor@btk.elte.hu)

## Bevezetés

A közgyűjteményi és webes forrású anyagok szemantikus címkézését céljaul kitűző fejlesztés a Digitális Örökség Nemzeti Laboratórium keretei között valósul meg. A Nemzeti Laboratórium program az ITM kezdeményezésére 2020-ban indult, jelenleg összesen 18 Laboratórium működik párhuzamosan. Érdemes kiemelni, hogy bölcsész- és társadalomtudományi területen csak két laboratórium tevékenykedik. A társadalmi innovációt kutató és előmozdító TinLab szintén az ELTE vezetésével kezdte meg működését, akár a konzorciumi formában létrejött Digitális Örökség Nemzeti Laboratórium. Az összes többi laboratórium a sokkal könnyebben piacosiható „Hard Science” területén jött létre. A laboratóriumok ötéves futamidővel működnek, többnyire konzorciumi formában. A DH-LAB együttműködő partnere a Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézete, melynek keretei között a DigiPhil projekt működik, és amelynek technikai megújítása<sup>1</sup> a DH-LAB támogatásával valósul meg. A DH-LAB további partnere a Magyar Nemzeti Levéltár, amely elsősorban a gépi kézírásfelismeréssel kapcsolatos tevékenységekben, illetve a digitálisan született anyagok kezelése (born digital curation) területén aktív, ahogyan további konzorciumi partnerünk, a Miskolci Egyetem is.

A digitális örökség Nemzeti Laboratórium létrehozásának fő célja az volt, hogy kidolgozza a nemzeti kulturális örökség mesterséges intelligencia alapú feldolgozásának, kutatásának és oktatásának, valamint a lehető legszélesebb körű közzétételének a módszertanát, és mindezt úgy, hogy a kifejlesztett módszertanok és eszközök, valamint a szakértői kompetencia piaci hasznosítására is gyakorlatokat alakítson ki.

Ez a konferencián bemutatott mindkét eszközre, a szemantikus címkézőre és a gépi kézírásfelismerőre egyaránt érvényes.

A digitális kulturális örökség kutathatóvá tételével szélesebb felhasználói kör számára nyújthatunk célzottabb intelligens hozzáférést, még hozzá szemantikus mélységben. A digitalizálás folyamatának gyorsítása és javítása MI eszközök révén, valamint a tömeges digitalizálás eredményeinek gépi tárgyszavazása, a kéziratok automatikus felismertetése többszörösére emelheti a széles körben, intelligens módon felhasználható kulturális tartalmak mennyiségét.

De térjünk át az előadás voltaképpeni tárgyára, a szemantikus címkézőre. Adódik a kérdés: tulajdonképpen miért építünk tematikus címkéző eszközt? Amellett, hogy a digitális örökség minden területén hasznos, a DH-LAB három fejlesztési iránya is épít az ezen eszköz nyújtotta lehetőségekre.

Az első tevékenység, amit kiemelnék, a webaratás alprojekt keretei között folyik.<sup>2</sup> A fejlesztés középpontjában egy nyílt hozzáférésű szoftver, a Web Article Curator<sup>3</sup> áll. Ennek a szoftvernek

1 Lásd jelen kötetben

2 <https://keptar.oszk.hu/kereses/reszletes.phtml?id=78376>

3 The ELTE.DH Pilot Corpus – Creating a Handcrafted Gigaword Web Corpus with Metadata Balázs Indig, Árpád Knap, Zsófia Sárközi-Lindner, Mária Timári, Gábor Palkó In the Proceedings of the 12th Web as Corpus Workshop (WAC XII), pages 33-41 Marseille, France 2020, <https://doi.org/10.5281/zenodo.3755323>





a segítségével webes forrású cikkek aratása történik, a weboldalak, elsősorban hírportálok cikkei részletes és szabványos formátumú metaadatokkal együtt kerülnek be a DH-LAB fejlesztette keresőszolgáltatásokba. Ezek jelenleg mintegy 3 millió cikket tartalmaznak. Mivel ezek többnyire szerzői jogvédett tartalmak, ezért a szolgáltatásban a nagyközönség csak a metaadatok között kereshet. Kutatási együttműködés keretében ugyanakkor a webes forrású cikk-korpuszt bármely érdeklődő kutatóval megosztjuk.

A célunk az, hogy ezeket a cikkeket a szemantikus címkéző segítségével a teljes szövegű szabad szavas keresésénél mélyebben tegyük kereshetővé.

Egy másik projektünk, ahol a szemantikus indexelésnek nagy a jelentősége, DH-LAB készülő repozitóriumi metakeresője, amely a magyarországi repozitóriumok dokumentumainak metaadataiban és azok szövegében egyszerre képes kereséseket végrehajtani. A SZTAKI és az MTA KIK közös keresője hasonló céllal jött létre, ám jelenleg csak a metaadatokat gyűjti össze, a dokumentumok szövegét nem kezeli.

Célunk az, hogy a szolgáltatás a dokumentumok teljes szövegében is lehetővé tegye a keresést, még hozzá szemantikus módon is. Ennek egyik fő akadálya jelenleg, hogy a gyűjteményi anyagok a repozitóriumokban többnyire változó minőségű PDF formátumú fájlokban vannak jelen. Célunk, hogy ezekből a dokumentumokból gépi feldolgozásra alkalmas formátumban nyerjük ki az adatokat, lehetőleg szabványos XML formátumban. Ennek különféle módszerei lehetségesek, azon dolgozunk, hogy a szolgáltatás automatikusan a legmegfelelőbb munkamenetet válassza ki a lehetséges eszközláncok közül, hogy a gépi feldolgozás számára a legalkalmasabbat válasszuk ki.

De miért is van szükség arra, hogy dokumentumokat wikifikáljunk vagy más módon szemantikus címkével lássuk el a szövegek elemeit?

Idézzünk fel egy példát. A képernyőképen az Ontotext cég által létrehozott szolgáltatás egy oldala látható. Az Ontotex hírportálok szövegét gyűjti össze és azokat szemantikusan címkézi. A képen egy a digitális kulturális örökség szempontjából releváns, a mesterséges intelligencia és a múzeumok kapcsolatáról szóló egészen friss cikk látható. Az Ontotext rendszere nem pusztán a cikkek szövegében keres, de azok elemeit különféle kategóriákba sorolt címkékkel látja el (pl. személyek, intézmények, földrajzi helyek, témák), és az ezek alkotta szemantikus hálózatot is bejárhatjuk.

Amikor Tim Berners-Lee a szemantikus web fogalmát bevezette, éppen ilyen szolgáltatás lebeghetett a szeme előtt. Olyan eszközök létrehozását sürgette, amelyek kizárják a többértelműségből származó tévedéseket (példáiban a Carlsberg karaktersor szerepelt, ami egyszerre jelölhet egy sörmárkát, egy futballcsapatot és egy földrajzi entitást).

A szemantikus címkézés célja pontosan ez: entitások és entitástípusok azonosítása különféle szövegbeli megfogalmazásokat összekötve (pl. Budapest, Magyarország fővárosa), illetve azonos szövegbeli megfogalmazások értelmi differenciálása (pl. arany mint anyagnév és Arany [János] mint személy).

Az első olyan projekt, amely a Huwikifier szemantikus címkéző kutatási célú mintázatfelismerésre használja, helyneveket azonosít. A kísérlet során Kiss Tamás digitális bölcsész szakértővel együttműködve a COST Action *Distant reading for European Literary History* című projekt keretei között az ELTE Digitális Bölcsészeti tanszékén készült regénykorpusz regényeit elemeztük. A Huwikifier segítségével mintegy 14 ezer helyszínt azonosítottunk a 100 magyar nyelvű regényben, majd a Notegoat szolgáltatás segítségével térképes vizualizáció formájában jelenítettük meg az eredményeket. Egy ilyen elemzés képes lehet kulturális és/vagy irodalomtörténeti trendek felismerésére, illetve ilyen érvelések megerősítésére (vagy éppen cáfolatára).



## Implementáció

Az eddigiekben egy szélesebb áttekintést adtunk a feladatról és a környezetről, amiben a rendszert használjuk. Ebben a fejezetben a „színfalak mögé lépünk” és egy részletesebb képet nyújtunk a Huwikifiről és a wikifikáció folyamatáról.

## Wikifikáció

Definíciószerűleg a wikifikálás a szemantikus annotáció egy olyan változata, ahol a szöveg fontosabb szavait a Wikipédia<sup>4</sup> entitásaival (oldalai URL-jével) címkézzük fel. Ennek a hagyományos – egyedi ontológián alapuló – szemantikus annotációval szemben számos előnye van. Egyrészt lehetővé teszi az annotációt saját ontológia készítése nélkül, hiszen adott egy nagy, közösségileg épített, több nyelven is elérhető entitás-adatbázis. A Wikipédia-entitások ráadásul bárki számára ismerős és érthető fogalmakat takarnak. Másodrészt az entitásokhoz tartozó cikkek, illetve az oldalak közötti hivatkozások lehetőséget biztosítanak az entitások automatikus felismerésére.

Természetesen a Wikipédiának hátrányai is vannak: a cikkek minősége hullámzó lehet, illetve a mennyiségük is változó az egyes nyelveken belül. A cikk írásának idején a legnagyobb Wikipédia, az angol körülbelül 6,5 millió cikket tartalmaz, míg a magyar kicsit többet, mint fél milliót. Ez utóbbi jelentősen korlátozza egy, a magyar Wikipédiára épülő rendszer felidézését (*recall*).

## Algoritmusok

Az itt használt „wikifikáció” szó az első rendszer, a *Wikify!* nevéből származik (Mihalcea és Csomai, 2007). Az évek során több hasonló algoritmust is kidolgoztak, amik mind az alábbi három lépést valósítják meg különböző módszerekkel:

1. Entitás-jelöltek felkutatása a wikifikálandó szövegben.
2. A jelöltek azonosítása Wikipédia-entitásokkal. Mivel egy jelölt több különböző entitásra is utalhat (pl. *Washington* város, állam, vagy személy?).
3. Az entitások szűrése aszerint, hogy mennyire relevánsak a szöveg tematikája szempontjából.

A *Wikify!* kulcsszókineréssel állítja fel a jelölteket, majd a jelölt környezete és a célentitás(ok) cikkei közötti szöveges hasonlóság alapján egyértelműsít. Milne és Witten egyrészt a szöveges helyett egy, a bejövő hivatkozások halmazhasonlóságán alapuló *szemantikus hasonlóság* (*semantic relatedness, SR*) metrikát vezet be (2008a), másrészt gépi tanuló algoritmusokat alkalmaz a jelöltek megtalálására és egyértelműsítésére (2008b).

Hoffart és társai (2011) alkalmaznak először gráfalapú hasonlósági módszert jelöltazonosításra. A jelöltek és a hozzájuk csatolható entitások egy *jelölt-entitás gráfot* feszítenek ki. Minden jelöltnek és lehetséges entitásnak egy-egy csomópont felel meg. Egy jelölt és egy entitás csomópont között akkor húzódik él, ha a jelölt utalhat az adott entitásra; két entitás között pedig akkor, ha azok valamilyen értelemben hasonlóak. Az entitások hasonlósági metrikája algoritmusfüggő, és az élek a metrika értékével súlyozottak.

Hoffart és társai módszere ebben a gráfban keres olyan sűrű algráfokat, amik minden jelölthöz egy entitást rendelnek. Brank és társai (2017) ezt az ötletet vitték tovább *Wikifier*<sup>5</sup> nevű rendszerükben, és a Google PageRank (Brin és Page, 1998) algoritmusával azonosítják

---

4 <https://www.wikipedia.org/>

5 <https://wikifier.org>



a legfontosabb entitásokat. Ehhez a személyre szabott (*personalized*) PageRank algoritmust használják, ahol a sztochasztikus szörföző a jelölt csomópontokból indul, és legmagasabb PageRankkel rendelkező entitásokkal annotál.

Megfigyelhető, hogy a legtöbb wikifikációs módszer a fenti három lépésből főleg az elsőre és a másodikra koncentrál; a relevanciaszűrés kevésbé kutatott terület. A Wikify! A kulcsszavak, míg a Wikifier a PageRank pontszámok alapján sorrendezi az entitásokat, és egy bizonyos küszöb alattiakat eldobja. A küszöb azonban mindkét esetben az algoritmus hiperparamétere, aminek helyes megválasztása jelentősen befolyásolja a rendszer teljesítményét.

## Huwikifier

A mi rendszerünk – ahogy neve is utal rá – a Wikifier újraimplementálásán alapul. Annak ellenére döntöttünk az újraimplementálás mellett, hogy a Wikifier algoritmus a nyelvfüggetlen, és az oldalon (illetve az API-n) keresztül magyarul is használható. Az egyik indok épp a nyelvfüggetlenség: a Wikifier nem veszi figyelembe a magyar nyelv jellegzetességeit, ami jelentősen rontja az eredményt. A másik ok, hogy olyan funkciókat is implementálni tudjunk, amik nem szerepelnek az eredeti, zárt rendszerben.

Jelen (béta) állapotában a Huwikifier az eredeti Wikifier algoritmusának minden lépését megvalósítja. A rendszernek két végpontja van: egy REST API, ami a Wikifier API nekünk szükséges részalmazát támogatja, és egy HTML tesztinterfész, ami egy megadott dokumentumon lefuttatja az annotációt, és különféle tesztelési információkat jelenít meg. A Wikifierrel való API kompatibilitás fontos szempont volt, hogy az azt használó szemantikus keresőt könnyen, a végpont átírásával egyszerűen át tudjuk állítani a Huwikifier használatára.

A tesztüzem alatt a rendszer nem elérhető a nyilvánosság számára, de a kész rendszert nyilvánosságra tervezzük hozni.

## Wikifikálás magyarul

Az eredeti Wikifierrel magyar mondatokkal tesztelve, két fő hibaosztályt ismertünk fel. Ezek közül az első a morfológia figyelmen kívül hagyása. A Wikifier a jelöltek megtalálásához a szövegben az egyes entitások valamilyen ismert, Wikipédiában látott említését keresi. A keresés azonban szigorúan a felszíni formára korlátozódik. Ez a módszer jól működik angolra, ahol egy lemmához kevés lehetséges felszíni alak tartozik; azonban egy, a magyarhoz hasonlóan gazdag morfológiájú nyelvben az *adatrítkaság* problémákat okozhat. Egy valós példa erre Székesfehérvár, aminek említései között előfordulnak a „Fehérvár”, „Fehérvárról” alakok, de pl. a „Fehérvárnál” nem, ezért ha az annotálandó szöveg ezt tartalmazza, az algoritmus nem ismeri fel azt jelöltként.

A másik hiba a szófajok figyelmen kívül hagyása. A „Tüdejét is kiköpi, ahogy szív egy kis levegőt.” mondatban a Wikifier mind a „tüdő”-t, mint a „szív”-et annotálja a megfelelő testrészek oldalával, függetlenül attól, hogy az utóbbi ebben mondatban egy ige.

A Huwikifier ezeket a problémákat az emtsv<sup>6</sup> (Indig és társai, 2019) segítségével oldja meg. Egyrészt mind a szöveget, mind az említéseket lemmatizálja, így az első példában mind a „Fehérvárról”, mind a „Fehérvárnál” „Fehérvár”-ra egyszerűsödik. Ez nemcsak abban segít, hogy az adatrítkaságot megszüntesse, hanem az említésadatbázis méretét is jelentősen lecsökkenti, hiszen elég a lemmákat eltárolni a felszíni alakok helyett. Másrészt az emtsv a szöveg minden szavához visszaadja a szófajt is, amit felhasználunk a jelöltek szűréséhez: minden jelölt, ami nem főnévre végződik (vagyis nem főnévi csoport) törlésre kerül.

6 <https://github.com/nytud/emtsv>

A Huwikifier az emtsv-t REST API-n keresztül éri. A Huwikifier teljesítményét szemlélteti, hogy az emtsv felel egy átlagos online média cikk feldolgozási idejének 2/3–3/4-ért.

## Új funkciók

A Huwikifiert, bár még béta állapotban van, elkezdtük bővíteni az eredeti rendszeren túlmutató funkciókkal. Ezek egyike a TEI XML<sup>7</sup> kezelése.

A digitális bölcsészet nagy hangsúlyt fektet a dokumentumokhoz tartozó metaadatokra, emiatt nyers szöveg helyett a repozitóriumok tipikusan TEI XML formátumú dokumentumokat tartalmaznak. Ennek megfelelően e formátum támogatása alapvető cél volt. A Huwikifier képes fogadni TEI XML formátumú bementet az API-n keresztül, és az annotációkat a szavak XML ID-jaival indexelni. Ezt utána a kliens mint külön *standoff annotation* mezőt adhatja hozzá a dokumentumhoz (hasonlóan egyéb, nem tokenszintű annotációkkal, mint pl. névelemek).

Mivel a TEI XML-eink már eleve tartalmazzák a legalapvetőbb nyelvi annotációkat (lemma, szófaj), ezért ebben az esetben a Huwikifier maga már nem hívja meg az emtsv-t, így egy TEI XML dokumentum feldolgozása gyorsabb, mint a nyers szövegé lenne.

A másik új funkciót a fenti COST Action projekt inspirálta. A Huwikifierbe felvettük a WikiData<sup>8</sup> ontológia IS-A kapcsolatait. Amennyiben ez az adat elérhető, a rendszer nem csak az egyes entitások WikiData azonosítóját adja vissza, hanem azok közvetlen taxonómiai szülőkategóriáit is<sup>9</sup>. Ezen kívül a kliens kikötheti, hogy csak bizonyos kategóriába eső entitásokra kíváncsi; ez esetben a többi entitást töröljük a kimenetből. Bár a fenti kutatás idején ez a funkció még nem volt elérhető, a további, hasonló projektek végrehajtását jelentősen meg fogja könnyíteni.

Terveink közé tartozik ennek a funkciónak a bővítése az IS-A-n kívüli, tetszőleges relációk támogatásával.

## Távlati tervek

A Huwikifier már jelenlegi állapotában is használható, és alkalmazható digitális bölcsészeti kutatások támogatására. Mielőtt azonban késznek nyilváníthatnánk, fontos, hogy az algoritmus robusztusabb legyen: kevesebb irreleváns vagy téves entitást annotáljon. Jelenleg ennek tesztelése és a talált hibák javítása folyik.

Hosszabb távon a rendszert kétféleképp tervezzük továbbfejleszteni. Ahogy láttuk, a különféle wikifikáló algoritmusok más és más módszerekkel oldották meg a főbb lépéseket. Tervezzük a Huwikifier algoritmusát más rendszerekben alkalmazott megoldásokkal, úgymint kulcsszó-kereséssel és gépi tanulási módszerekkel kiegészíteni. Természetesen utóbbiak ma neurális, ún. mélytanulós modellek lennének. Ehhez szükségünk lehet megfelelő tanító és tesztadat létrehozására is.

A másik irány a Wikipédián kívüli, további tudásbázisok bevonása. Olyan adatokra gondolunk itt, mint más wiki oldalak, enciklopédiák, ember- (ki-kicsoda) vagy cégnyilvántartások, stb. Mivel ezekben nincs feltétlenül meg a Wikipédia interkonnectivitása, vagy az entitások szöveges körülírása, automatikus annotációra való alkalmazhatóságuk külön kutatás tárgya lesz.

---

7 <https://tei-c.org/>

8 <https://www.wikidata.org/>

9 Hasonló funkció létezik az eredeti Wikifierben is, de az DBPedia kategóriákat ad vissza.



## Bibliográfia

- Brank, Janez, Gregor Leban, and Marko Grobelnik. „Annotating documents with relevant wikipedia concepts.” *Proceedings of SiKDD* 472 (2017).
- Sergey Brin and Larry Page (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 30 (1–7): 107–117.  
[https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, et al. 2011. [Robust Disambiguation of Named Entities in Text](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 782–792, Edinburgh, Scotland, UK.. Association for Computational Linguistics.
- Balázs Indig, Bálint Sass, Eszter Simon, Iván Mittelholcz, Noémi Vadász, and Márton Makrai: One format to rule them all – The emtsv pipeline for Hungarian. In: *Proceedings of the 13th Linguistic Annotation Workshop*. Association for Computational Linguistics, 2019, 155-165. <https://doi.org/10.18653/v1/W19-4018>
- David Milne and Ian H. Witten. 2008a. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links (AAAI 2008). Association for the Advancement of Artificial Intelligence. Chicago, IL, USA, 25-30.
- David Milne and Ian H. Witten. 2008a. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management (CIKM ,08)*. Association for Computing Machinery, New York, NY, USA, 509–518.  
<https://doi.org/10.1145/1458082.1458150>
- Rada Mihalcea and Andras Csomai. 2007. Wikify! linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM ,07)*. Association for Computing Machinery, New York, NY, USA, 233–242. <https://doi.org/10.1145/1321440.1321475>