

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

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



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: A Régi Magyar Költők Tárának 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 szövegtörzseken. 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 Máttyás: Kézírtos dallamlejegyzések 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 névelő-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása)	359

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
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.¹ 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.²

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.³

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.⁴

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.⁵

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.⁶

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.⁷

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.⁸

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 használati célt 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 rendszerek⁹ 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.¹⁰

Az integrált rendszerekben tárolt adatok eddig is fontos források 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.¹¹

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 lehettek 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.¹²

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.¹³

Akatalógusokba 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.¹⁴

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 időbélyeggel, és a módosítási lépések tárolása.¹⁵

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

¹⁵ 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.¹⁶

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.¹⁷

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).¹⁸ 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.

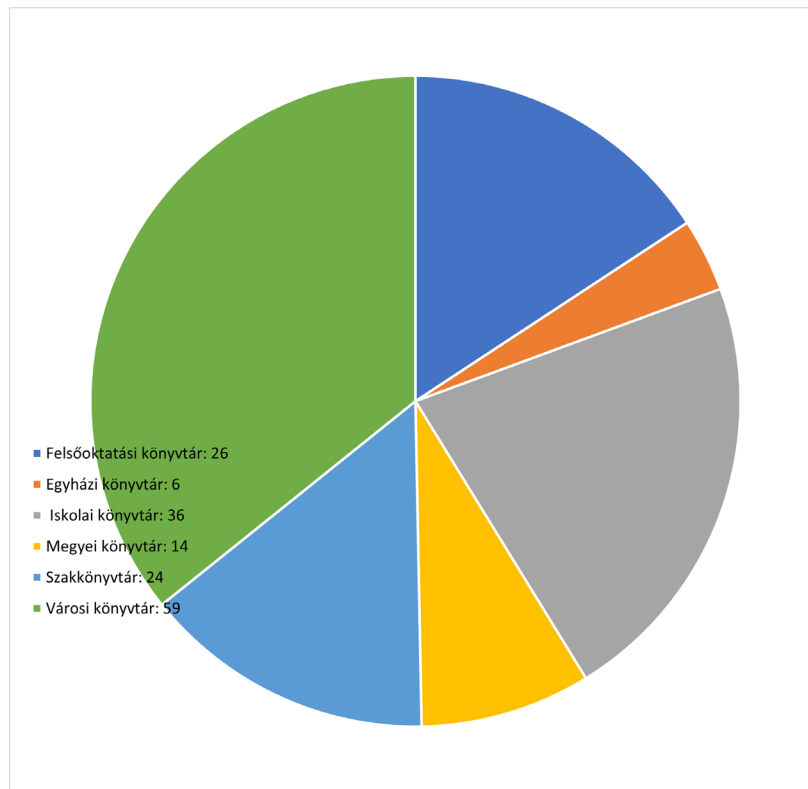
A vizsgálat bemutatása

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.¹⁹

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.

¹⁹ 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/pdte-2021-0007>

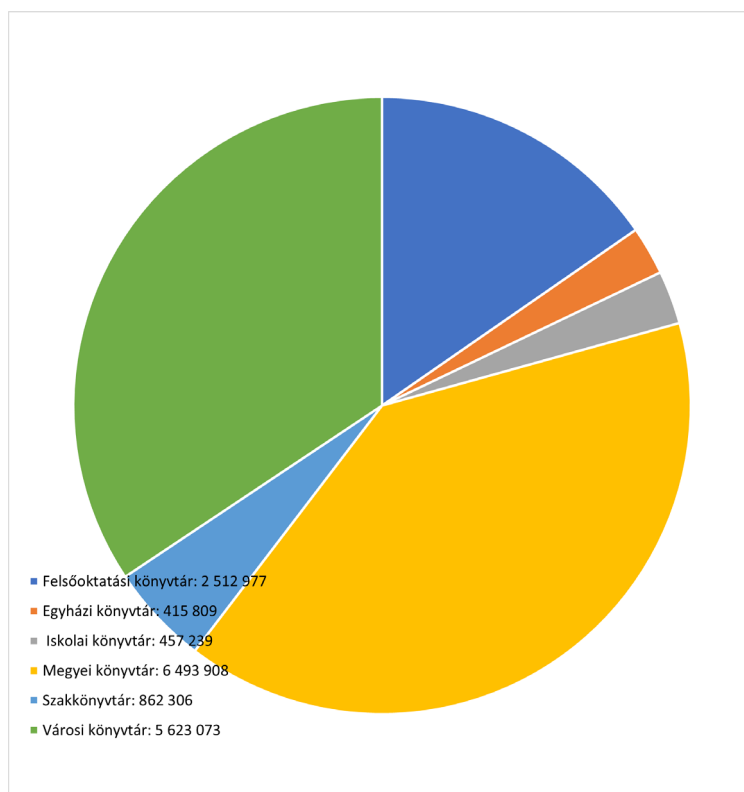
Az 1. ábrán a vizsgálatban végül részt vevő könyvtári adatbázisok számát mutatom be, könyvtártípus szerinti bontásban.



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 feladata²⁰). 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.

²⁰ 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 végezhetők, illetve CSV outputok nyerhetők 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.

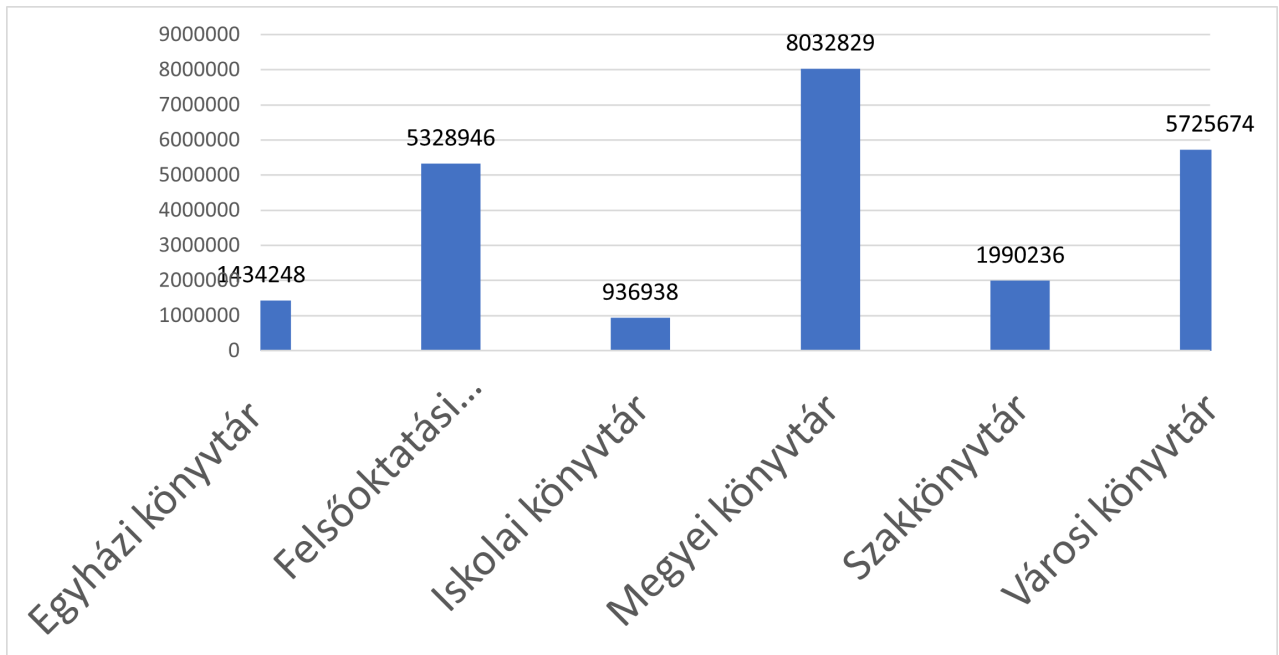
A vizsgálat tárgyát képezte ellenőrzési célból:

- 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

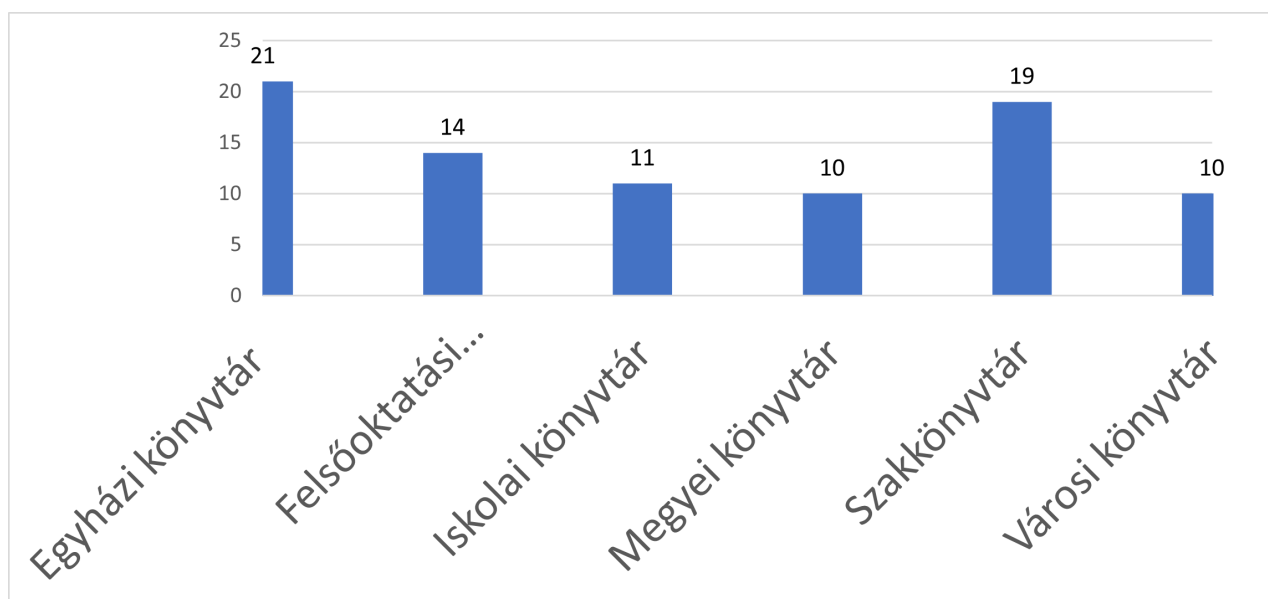
A 3. ábrán a vizsgálatban végül részt vevő könyvtárakban a kölcsönözhető példányok számát mutatom meg, intézménytípusok szerinti csoportosításban



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.²¹

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,

²¹ 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.²² 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.²³

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.²⁴ 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.²⁵ 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.²⁶

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).²⁷

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/pdtc-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.