

## Digitális dokumentumok gyűjteménykezelési gyakorlatának támogatása a digitális tartalmak számossága, mérete és féleségeik vizsgálatával

### Where to write the inventory number? Practice of handling born digital documents in recent collection management systems.

Simon András

Monguz Információtechnológiai Kft Szeged  
[andras.simon@qulto.eu](mailto:andras.simon@qulto.eu)

#### Absztrakt

Folyóirat cikkek, tanulmányok, levéltári iratok, magánlevelek, fényképek és a különféle kisnyomtatványok sokasága keletkezik ma már digitálisan. Ezek a dokumentumok különösen számottevő részét teszik ki az egyes közösségek saját kulturális örökségének, melyet a társadalom részéről jelenleg nagy figyelmet kapó terület, a helytörténet kezel. A digitális objektumok látszólag kevés gondoskodást igényelnek, főként a könnyű tárolhatóság, és az olcsó, veszteségmentes sokszorosítás lehetősége miatt, ugyanakkor a közgyűjteményi állományok legsebezhetőbb, legveszélyeztetettebb részét alkotják. Tartós, intakt, hiteles és dokumentált megőrzésük nagy felelősséggel járó munka, amely nagy terhet ró a emlékezetintézményekre. A gyűjteménykezelés korszerű szabályozásához, és a megfelelő gyakorlati megoldások kialakításához elengedhetetlen megfelelő információk szerzése ezen dokumentumok mennyiségéről, elterjedtségéről, sokféleségéről. A tanulmány szerzője a Monguz Kft munkatársaként négy éve, a doktori disszertációjához folytatott kutatásai keretében vizsgálja a cég múzeumi és könyvtári ügyfeleinek adatbázisait, egyebek mellett az őrzött digitális tartalmak számossága mérete és féleségeik tekintetében. A kutatás keretében 2020 őszén lefolytatott egyik vizsgálatot végzi el most el újra, megpróbálva megállapítani a közgyűjteményi tartalomkezelés legfontosabb trendjeit. A tanulmányban ezen vizsgálat eredményéről készül beszámolni.

**Kulcsszavak:** *közgyűjteményi informatika, tartalomkezelés, digitálisan keletkezett dokumentum, tartalomelemzés, fájlformátum*

#### Abstract

Most of the journal articles, studies, archival documents, private letters, photographs and various small print documents are being produced digitally in the recent days. These documents make up a particularly significant part of the cultural heritage of each community, and local history, managing these types of objects, is one of the areas currently receiving considerable attention from society. Digital objects seem to require little care, mainly because of their easy storage and the possibility of cheap, lossless reproduction, but they are the most vulnerable and endangered objects of public collections. Storing them for a long time unharmed, authentic and documenting the interventions is a major responsibility and a big task for heritage institutions. To ensure modern collection management and to develop appropriate practices, it is essential to obtain adequate information on the quantity and diversity of the digital objects and also information is necessary, how widely these documents are used. The author of the study, as the employee of Monguz Ltd. in 2020 examined the databases of the company's museum and library clients, including the size and diversity of the digital contents of the databases. This examination was the part of his doctoral dissertation researches. The study is now being carried out again after almost

three years, trying to establish key trends in content management in public collections. The presentation will report on the results of this investigation.

A közgyűjteményekben tárolt digitális objektumok kezelése és tárolása növekvő jelentőségű feladatot jelent az intézmények számára. A digitális tartalmak száma egyre nagyobb, és méretük is növekszik. Ennek oka, hogy egyrészt egyre szélesebb a digitálisan keletkezett dokumentumok köre, másrészt melléjük sorakoznak a megőrzési célból digitalizált analóg dokumentumok digitalizált példányai. Különösen nagy szerep hárul a közgyűjteményi munkafolyamatok informatikai igényeinek kiszolgálásakor az intézményben használt integrált rendszerekre, melyek tervezésekor és fejlesztésekor az alábbi problémák megoldására kell felkészülni:

- A digitális és analóg dokumentumokat együttesen kell kezelni, egyelőre ugyanis - főként a múzeumokban - még nincsenek önálló gyűjtemények kialakítva a digitálisan keletkezett, vagy csak digitálisan őrzött dokumentumok számára,
- A tartalomszolgáltatás fejlesztése, összhangban a korszerű felhasználói (fogyasztói) igényekkel, a szerzői jogi szabályozásokkal, a személyiségi jogi elvárásokkal és az emlékezet intézmények saját hosszútávú érdekeivel,
- A mennyiségi növekedés kezelése. Ezek az objektumok óriási mennyiségben keletkeznek, és az analóg tartalmak retrospektív digitalizálása is egyre nagyobb ütemben halad. Az átlagos fájl méret bizonyos mértékben még ma is növekszik, a válogatás feladata tehát egyre kevésbé megkerülhető,
- A formai és tartalmi feltárás egyre költségesebb, a mennyiségi növekedés és az egyre nagyobb kvalifikációt igénylő emberi munkaerő miatt.
- A tartós megőrzés, (műszaki használhatóság, szolgáltatathatóság eredetiség és hitelesség biztosítása) költségei is a kereslet folyamatos bővülése miatt állandóan az inflációt meghaladó ütemben növekszenek.

A korszerű integrált gyűjteménykezelő rendszerekben tehát többféle tartalom kezelésére kell felkészülni. Az őrzött dokumentumok egyrészt lehetnek:

- Analóg,
- szolgáltatási célból digitalizált,
- megőrzési célból digitalizált, illetve
- digitálisan keletkezett dokumentumok.

Másrészt az őrzés szempontjából el kell különíteni a

- Sorozatgyártással tömegesen keletkezett,
- sorozatgyártással kis mennyiségben keletkezett, illetve az
- egyedi dokumentumokat.

Az anyag jellege szerint megkülönböztethető, de könyvtári, múzeumi és levéltári közgyűjteményi ágakban egyaránt kezelt dokumentumfélések az alábbiak lehetnek:

- Könyv,
- periodika,
- különlenyomat,
- térkép,
- audiovizuális felvétel,
- levéltári irat,
- kézirat,

- kisnyomtatvány (pl. gyászjelentés, rendezvény meghívó),
- képeslap,
- fotó,
- történeti dokumentum (bármilyen történeti értékű szöveges anyag),
- történeti tárgy,
- vizuális dokumentum.

A történeti tárgy kivételével ezek mindegyike keletkezhet ma már digitálisan is.

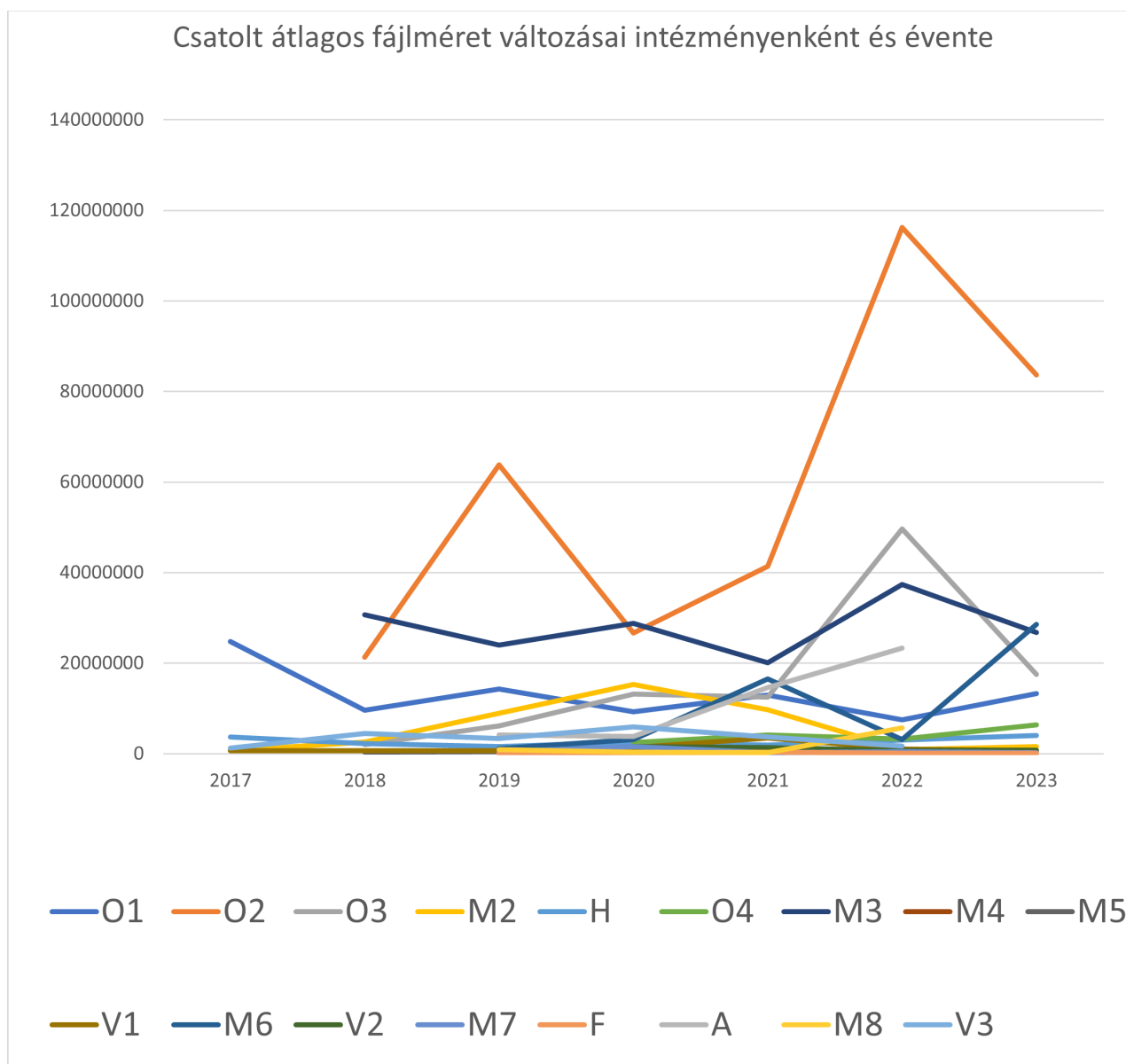
Tekintettel a közgyűjteményekben kezelt digitális dokumentumok hasonlóságára, a közgyűjteményekben hasonló megoldási stratégiák vannak kialakulóban, és emiatt a tartalomkezelést az egyes gyűjteményi ágak esetében párhuzamosan kell vizsgálni. A hasonló elvárásoknak megfelelő hasonló megoldások jelentőségét növeli a közösségek kulturális örökségének mind hangsúlyosabb jelenléte a - mindinkább digitálisan létező - helytörténeti gyűjteményekben. Ezek a tartalmak, melyek korábban nehezen voltak elérhetők, most könnyen közzé tehetők, és a kényelmes és bárholnan bármikor elérhető felületek révén egyre növekvő közönség számára váltak hozzáférhetővé. A digitalizációnak részint okaként, részint következményeként a kulturális örökség fogalmköréhez tartozó, a helyi közösség saját története szempontjából fontos dokumentumok jelentősége egyre nagyobb.

A közgyűjtemények közös történelmi küldetése annak megakadályozása, hogy az emberiség egy évtizede jelentős részben már digitálisan keletkező kulturális örökségét ne nyelje el a „digital black hole”, a digitális fekete lyuk. Ezek az állományok az analóg tartalmaknál sokkal veszélyeztetettebbek, egy részük megfelelő gondosság hiányában örökre elveszhet. Hallatlanul nagy a sérülés, az adatvesztés kockázata. A gyűjtésre, illetve az őrzésre kialakult megoldások, hosszú ideje bevált gyakorlatok a dokumentumok újszerű volta miatt még nem állnak rendelkezésre. Ráadásul a digitális tartalmak hosszútávú, dokumentált és hiteles formában való megőrzése, illetve tovább hasznosíthatóságának hosszú időtartamra való biztosítása igen költséges.<sup>1</sup>

Ezen tartalmak katalógusokban való jelenlétének vizsgálatakor természetesen nem tekinthetünk el attól, mind szakmai, mind fenntartói oldalról nagy az érdeklődés aziránt, hogy mekkora a kihasználtságuk. Fontos rámutatni arra, hogy a digitális tartalmak interneten keresztül történő használata nehezebben mérhető, mint a hagyományos könyvtári kölcsönzés. Ennek nyomon követése a fizikai kölcsönzés esetében alapvető lehetőségként adódik az integrált könyvtári rendszer működéséből. A könyvtári ügyviteli munkát támogató integrált könyvtári rendszer ugyanis minden kölcsönzési eseményt, annak valamennyi adatát tranzakciós rekordokban őrzi, mint könyvtár és olvasó közötti gazdasági eseményt naplózó elektronikus bizonylatot. Bármilyen webes keresőfelületet érintő megkeresés ezzel szemben természetesen csak a rengeteg műszaki információt tartalmazó naplófájljába kerül beírásra. Mivel ez a fájl óriási, emiatt csak rövid ideig van őrizve. A tartós őrzésre érdemes hasznos statisztikai adatokat ebből a fájlból kell kibányászni, addig amíg a fájl egyáltalán megvan. Mindez inkább a szabadpolcos állomány helyben való használatának méréséhez hasonlítható, amikor a könyvtárosok az olvasótermi asztalokon hagyott könyveket a polcra visszahelyezve, erről az eseményről statisztikát vezetnek. A mérés alapja a fizikai kölcsönzési eseménnyel ellentétben tehát nem egy alapfunkcióként tartósan őrzött tranzakciós rekord, hanem a rendszer naplófájljaiba „mellékesen” rögzített adat.

1 Weber, H. Chrobak, Lennart: Legal Implications of Digital Heritagization. Implications juridiques de la patrimonialisation numérique. = RESET [Online], 6 | 2017, Patrimoine et patrimonialisation numériques. <http://journals.openedition.org/reset/826> ; DOI : [10.4000/reset.826](https://doi.org/10.4000/reset.826)

Mindez alátámasztja azt, hogy nagyon fontos minél pontosabb ismeretekkel rendelkezünk a digitális tartalmakról féleség, méret és kor tekintetében, lehetőleg felvázolva az elmúlt évek trendjeit is. Ezen szándék vezetett az elmúlt évek során végzett kutatásaimban, melyeket munkaadóm a Monguz Információtechnológiai Kft ügyfeleinek adatbázisaiban végeztem. Az adatbázisokban megjelenő csatolt fájlok számosságára, féleségére és méretére vonatkozó kutatásaimat 2020 őszén végeztem el, ezeket ismételttem meg 2023 tavaszán ugyanazokra az adatbázisokra ugyanazokat a lekérdezéseket elindítva. A kutatás körülményeiről és módszertanáról hely hiányában itt nincs módomban beszámolni, de korábbi publikációkban, ezekről részletesen írok.<sup>2</sup>



1. ábra Az átlagos fájl méret változásai

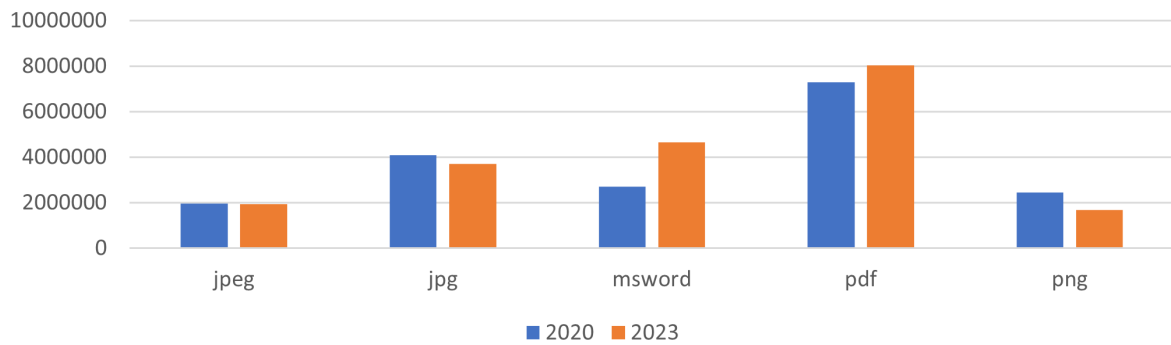
Az első ábrán a digitális katalógus rekordokhoz csatolt fájlok átlagos méretének változásait szemléltettem. A vízszintes tengelyen a fájl a katalógus rekordhoz való csatolásának (nem a létrejöttének) az éve látható. A függőleges tengelyen a fájl mérete olvasható byte-ban kifejezve. Különbőféle színekkel az egyes vizsgált intézményeket jelöltem. A rövidítések

<sup>2</sup> Simon András: Integrált Könyvtári Rendszerben tárolt tranzakciós rekordok felhasználása a könyvtárhasználat statisztikai elemzésére= TMT 66. évf. 2019. 12. sz.

feloldása: „O” (országos intézmény 4 db), „M” (Megyei hatókörű városi intézmény 7 db), „H” (Helytörténeti intézmény), „V” Városi intézmény 3 db), „F” (Felsőoktatási intézmény 1 db), „A” (archívum 1db). Az ábra jól mutatja, hogy átlagos növekedés még enyhe mértékben sem feltétlenül állapítható meg három év távlatából. A két kiugró értéket mutató országos intézmény egyébként nagy mennyiségben gyűjt és kezel, audiovizuális anyagokat.

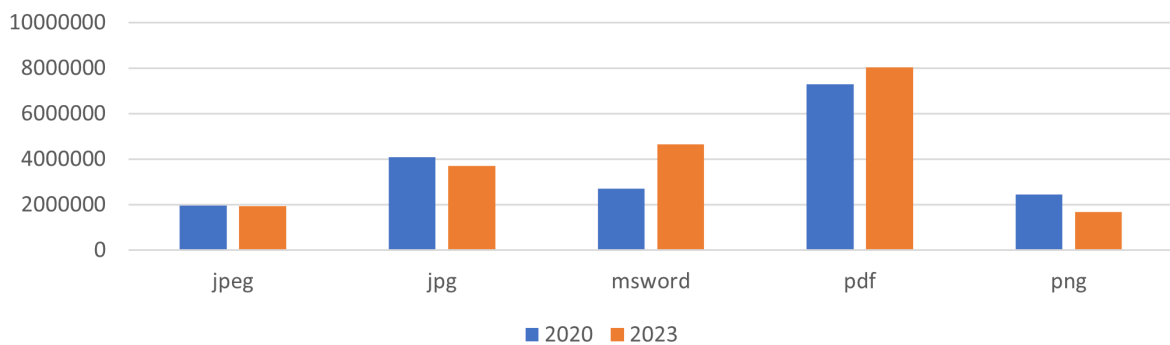
Az 1. számú ábrán látható diagramon felsorolt intézmények adatai láthatók a további ábrákon is.

Gyakoribb fájlok átlagos méretének méretváltozása, kisebb méret



2. ábra Fájlméretek méretváltozásai, (kisebb méretű fájlok)

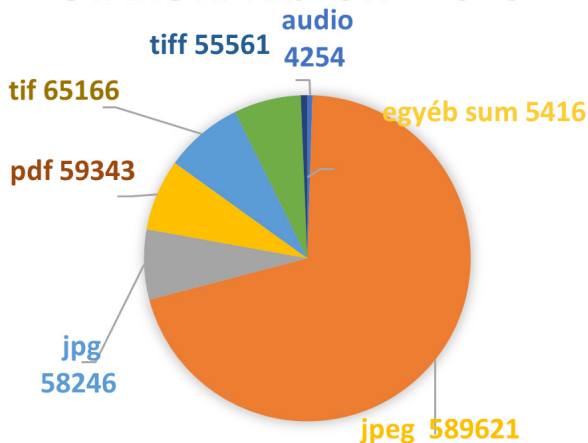
Gyakoribb fájlok átlagos méretének méretváltozása, kisebb méret



3. ábra Fájlméretek méretváltozásai, (nagyobb méretű fájlok)

A 2. és 3. ábrán a leggyakoribb fájltypusok átlagos méretváltozását mutatom be. Az áttekinthetőség kedvéért bizonyos fokú egységesítést végeztem a fájlnevekben, összevontam kezelem pl. a tif és tiff formátumot. A nagyobb és kisebb méretű fájlokat a megfelelő személtetés céljából különválasztottam. Az 1. számú ábrához hasonlóan itt is szignifikánsan kimutatható átlagos méretnövekedésről nem beszélhetünk.

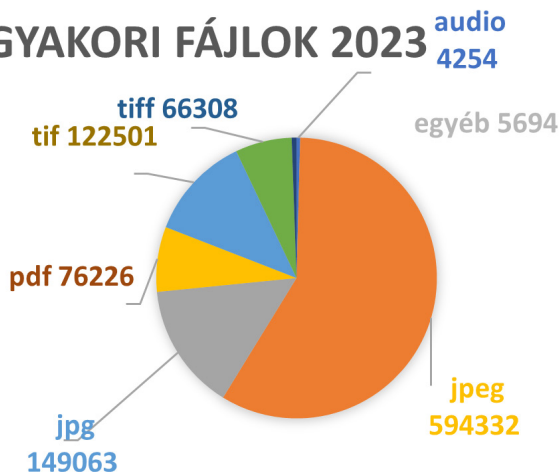
## FÁJL FÉLESÉGEK: GYAKORI FÁJLOK - 2020



4. ábra Fájl féleségek 2020 – gyakrabban előforduló fájlok

A 4. és 5. ábrán jól látható, hogy a jpeg és jpg formátumok együtt háromnegyed körüli arányt képviselnek.

## FÁJL FÉLESÉGEK: GYAKORI FÁJLOK 2023

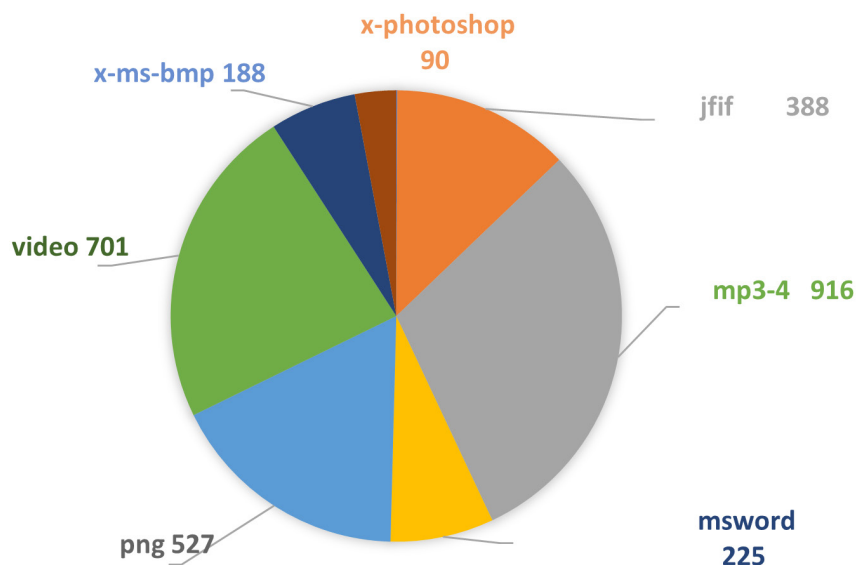


5. ábra Fájl féleségek 2023 – gyakrabban előforduló fájlok

képest kicsit nagyobb sokféleség, de ez csak a ritkábban előforduló fájlok viszonylagosan nagyobb számával magyarázható. Így pl. a bmp formátum a 2023-ra 625 előfordulással már felkerült a diagramra. Egészen új fájl típusok az elmúlt három évben még kisebb számban sem jelentek meg. Egyes fájlok esetében számszerű csökkenés is észlelhető. Ha intézményekre lebontva vizsgálnánk a fájlok összesített darabszámát, azt látnánk, hogy még inkább tetten érhető az egyes intézményeknél bizonyos fájl típusokban a darabszám csökkenése, ez a csatolt tartalmak rendszeres, minőségi szempontokat követő cseréjével magyarázható.

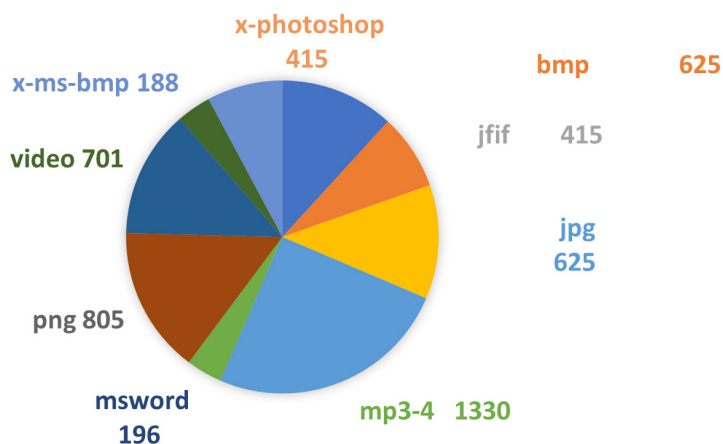
A 4., 5. 6. és a 7. ábrákon látható kördiagramokon a fájl féleségeket, a különféle fájl típusok előfordulásait vizsgálom. Külön választottam a gyakrabban és a ritkábban előforduló fájlokat, hogy a diagramon a féleségeket, a két adott évben egymáshoz viszonyított méretbeli arányukat jobban szemléltethessem. A gyakoribb előfordulású fájlokat bemutató diagram „egyéb” kategóriája van kibontva a ritkább fájlokat bemutató diagramokon. Itt a fájl név kiterjesztéseket a diagram készítésekor nem módosítottam, úgy láthatók itt, ahogyan a fájl csatolásakor az adatbázisba a program regisztrált ezeket.

## FÁJL FÉLESEGEK: RITKÁBB FÁJLOK - 2020



6. ábra Fájl féleségek 2020 - ritkábban előforduló fájlok

## FÁJL FÉLESEGEK: RITKÁBB FÁJLOK - 2023



7. ábra Fájl féleségek 2023 - ritkábban előforduló fájlok

Három év nem nagy idő, ez alatt komoly trendváltozást láthatóan nem tapasztalhatunk. Az adatok zöme múzeumi adatbázisokból származik, könyvtári katalógusokban lévő bibliográfiai rekordokhoz kevés tartalom lett egyelőre még csak csatolva. A felsőoktatási és szakkönyvtárak ugyanis sok esetben önálló a könyvtári katalógustól független intézményi repozitóriumot vezetnek. Az intézmények egy részében ma is tetten érhető a kampányszerű digitalizálás jelensége, a csatolt fájlok darabszáma, az egyes években lényegesen különböző mértékben növekszik. Mindezek miatt a kutatást mindenképpen érdemes lesz megismételni néhány évvel később, amennyire lehet figyelmet fordítva a párhuzamosan épített repozitóriumokra is.

The background is a complex digital artwork. It features a grid of squares, each containing a different texture or color, ranging from warm oranges and yellows on the left to cool blues and teals on the right. A bright, glowing light source is positioned in the center, creating a lens flare effect that radiates across the grid. The overall composition is symmetrical and has a high-tech, futuristic feel.

# ÚJ TECHNOLÓGIÁKKAL, ÚJ TARTALMAKKAL A JÖVŐ DIGITÁLIS TRANSZFORMÁCIÓJA FELÉ

32. Networkshop: országos konferencia

2023. április 12–14.  
Pannon Egyetem, Veszprém



# ÚJ TECHNOLÓGIÁKKAL, ÚJ TARTALMAKKAL A JÖVŐ DIGITÁLIS TRANSZFORMÁCIÓJA FELÉ

**32. Networkshop: országos konferencia**

2023. április 12–14.  
Pannon Egyetem, Veszprém

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület  
Budapest, 2023



Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Workshop

2023. április 12–14. Pannon Egyetem, Veszprém konferencia előadásainak közleményei

ISBN 978-615-82243-1-4

DOI: [10.31915/NWS.2023](https://doi.org/10.31915/NWS.2023)

Kiadja a HUNGARNET Egyesület  
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2023

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

## TARTALOMJEGYZÉK

Előszó.....	5
Király Sándor, Balla Tamás: Flipped classroom az sqlsuli.hu-ban.....	7
Wirágh András: Abaújszántótól Zombolyáig. Megjegyzések egy új sajtóadatbázishoz .....	14
Albert Ágota Katalin: Az EGT-tagállamok adatvédelmi felügyeleti hatóságainak szankcionálási gyakorlata az oktatási szektorban a GDPR alkalmazása óta .....	19
Simon András: Digitális dokumentumok gyűjteménykezelési gyakorlatának támogatása a digitális tartalmak számossága, mérete és féleségeik vizsgálatával .....	24
Bódog András: Az Annif gépi tárgyszavazó rendszer magyarországi adaptációjának feltételei és lehetőségei .....	31
Dezső Krisztina: A Pécsi Egyetemtörténeti Gyűjtemény online adatbázisai és digitális gyűjteményei .....	36
Ungváry Rudolf, Király Péter: Nemzeti könyvtárak és az OSZK MARC21 állományainak összehasonlító elemzése néhány adatmező alapján .....	42
Szemes-Révész Enikő Evelin: Kapocs a tudáshoz – A könyvtár szerepe a civilek és a tudomány kapcsolatában .....	50
Tóth Zoltán: Az RO-Crate alapú kutatási objektum csomagolás keretrendszere az ELKH ARP platformban .....	54
Király Roland, Király Sándor, Palotai Martin Marcell: Neurális hálózatok oktatási alkalmazását támogató keretrendszer Virtual (VR) és Augmented Reality (AR) eszközökkel .....	60
T. Nagy László: Mesterséges intelligencia, multimédia, tanulástámogatás .....	69
Horváth Péter: Egy automatikusan generált rímshótár fejlesztése és a magyar kanonikus költészet rímshavainak néhány jellemzője .....	77
Héjja Balázs, Tóth-Jávorka Brigitta, Tóth Máté: Digitális tartalomfejlesztés közkönyvtári környezetben .....	85
Koczka Ferenc: Szemelvények egy felsőoktatási rendszer informatikai védelmének tapasztalataiból .....	91
Bolya Mátyás: A digitális gyűjtésrekonstrukció lehetőségei: az Ethiofolk projekt .....	99
Dobás Kata, Sidó Zsuzsa, Szabó-Reznek Eszter: A Kolozsvári Állami Magyar Színház jelmezterveinek digitalizációja és felvitele az ITldata adatbázisba .....	108
Köpösdí Zsuzsa: H5P-ben létrehozható interaktív és adaptív tananyagok .....	116
Fülöp Tiffany, Molnár Tamás, Hoczapán Szabolcs: Komplex kutatástámogató szolgáltatási portfólió az SZTE Klebelsberg Könyvtárban .....	122
Vass Johanna: Az Open Science könyvtári vonatkozásai .....	129
Antal Péter, Czeglédi László: A digitális oktatás módszertana a gyakorlatban .....	135
Máray Tamás: A szuperszámítástechnika mint európai stratégiai ágazat .....	143
Frankó Máté, Zeller Rozália: Szoftveres Cutter-keresés az SZTE Klebelsberg Könyvtárban .....	151
Zsiborács Judit, Dési Ádám Dániel, Nagy Attila Árpád, Urbán Katalin: Tudományometriai műhely könyvtári környezetben .....	157



Palkó Gábor, Szekrényes István, Bobák Barbara: A Digitális Örökség Nemzeti Laboratórium webszolgáltatásai automatikus kézírás-felismertetéshez .....	164
Szűcs Kata Ágnes: Adatvizualizációs lehetőségek a bölcsészettudományban .....	170
Leitgéb Mária: A BME Építészettörténeti és Műemléki Tanszék repozitóriuma .....	178
Mihály Eszter, Micsik András: Szerkesztői környezet TEI-alapú szövegkiadásokhoz .....	186
Dobás Kata, Fellegi Zsófia, Palkó Gábor: A kis gömböc meséje - az ITIdata irodalomtudományos adatbázis fejlesztése 2022–2023-ban .....	192
Alföldi István, Szemigán Dorottya Henrietta, Palkó Gábor, Fellegi Zsófia: Kutatói e-mail hagyaték archiválása és feldolgozása .....	199