

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 kötelezpéldányok 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 szövegtörzseten. 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 névelim-azonosítás magyar nyelvű szövegeken (a HuWikifier bemutatása) | 359 |

Digitálisan született kutatási anyagok megőrzése: a relációs adatbázis mint born-digital objektum

Alföldi István

Digitális Örökség Nemzeti Laboratórium
alfi@poliphon.hu

Dióssy Anna Laura

Digitális Örökség Nemzeti Laboratórium
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.¹ 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.² 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.³ 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.

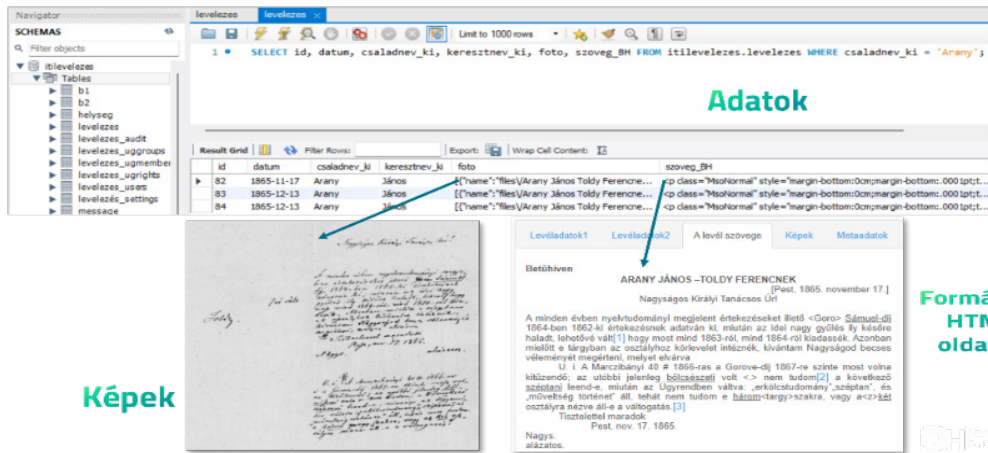
2. Az adatbázis archiválás kihívásai

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 tartalmazták. 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 tartalmazza.

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.

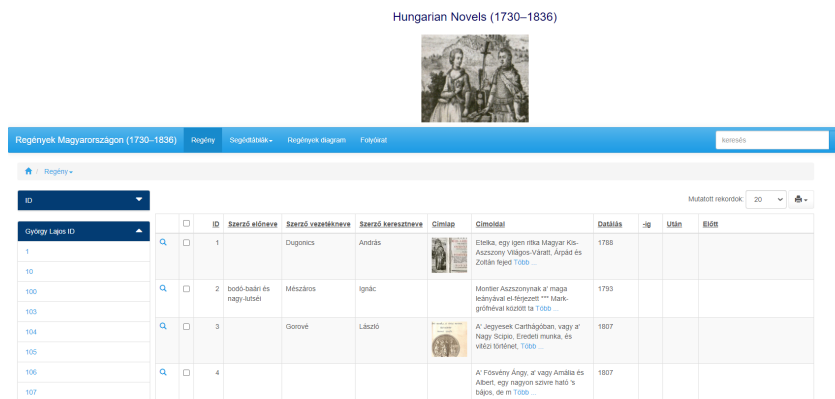


1. ábra. Képek és HTML mezők az adatbázisban

Az adatbázisok használati módja szintén kihívások elé állította a labort. Az Irodalomtudományi Intézetől 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.



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.⁶

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.⁷

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.⁸

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 lehetetlenné 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... | | A Barna Péter-ponyvák repertoriája nem |
| 10 | 248 | A British Library (11585.aa.52 jelzetű) példány... | | |

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

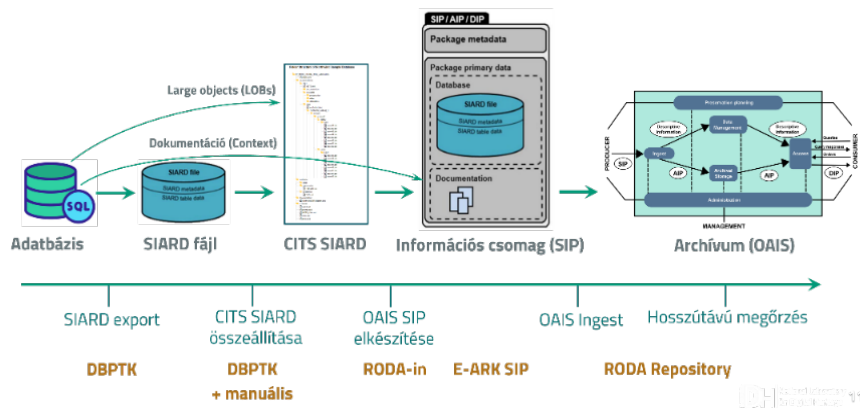
fájl sok kép esetén is kezelhető marad, de figyelni kell rá, hogy a képek a SIARD fájlal mindig együtt mozogjanak, és ne változzanak a könyvtár és fájl elnevezések. A harmadik megoldás szerint az eredetileg is külön tárolt képeket a SIARD mellett archiváljuk. Ez szabványos Data Link mezők esetén automatikusan 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 belülre tegye. Az 1. pilot esetében belülre 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)

E-ARK SIP specifikáció

(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)
- Case Study 2 (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>.