

Kutatói e-mail hagyatéék archiválása és feldolgozása

Long-term preservation of e-mail heritage

Alföldi István

Digitális Örökség Nemzeti Laboratórium

alfi@poliphon.hu

ORCID: [0009-0002-9634-0482](https://orcid.org/0009-0002-9634-0482)

Szemigán Dorottya Henrietta

Digitális Örökség Nemzeti Laboratórium

szemigan.dorottya@btk.elte.hu

ORCID: [0009-0007-5116-838X](https://orcid.org/0009-0007-5116-838X)

Palkó Gábor

Digitális Örökség Nemzeti Laboratórium

Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet

palko.gabor@btk.elte.hu

ORCID: [0000-0002-4394-8577](https://orcid.org/0000-0002-4394-8577)

Fellegi Zsófia

Bölcsészettudományi Kutatóközpont, Irodalomtudományi Intézet

fellegi.zsofi@abtk.hu

ORCID: [0000-0001-9199-1759](https://orcid.org/0000-0001-9199-1759)

Abstract

The management of born-digital content is becoming an increasingly relevant task in the different cultural fields, thus the long-term preservation of digitally generated material has been in the focus of the archivist community since the beginning of the new millennium. In the early twenty-first century, the Library of Congress pointed out its concerns about the fragility of the solely digitally created records. Since such material does not have an analog (physical) version, there is a greater risk of either getting lost, or not being available or accessible in their original form as historical records for future researchers.

Preserving digitally generated correspondence between individuals has been an important research area of digital archiving for many years. There are already good practices for archiving e-mails, being a priority especially in the corporate sector. However, merely preserving and making e-mails researchable is not the only issue, since in the case of the cultural heritage materials for instance a critical attitude and a curatorial practice is also essential in the archiving work. Furthermore, one of the reasons why a general solution to e-mail preservation still has not yet been found is that archiving e-mails at different stages of their lifecycle requires radically different solutions.

Our research project at the National Laboratory for Digital Heritage focuses on the long-term preservation of an e-mail heritage, more precisely the electronic correspondence of a Hungarian critic and literary historian from Cluj-Napoca (Transylvania), Lajos Kántor. We

are investigating how e-mail archiving tools can be used in this specific field, how e-mails can be made visible and researchable as an integral part of the overall heritage, and how the experience gained in preserving and processing paper-based correspondence can be used in archiving and publishing e-mails.

Keywords: Born-digital, e-mail preservation, cultural heritage, long-term preservation

Bevezetés

A hálózatba kötött elektronikus levelezés robbanásszerű fejlődése az elmúlt 52 év egyik jelentős technikai és társadalomtörténeti fejleménye volt. A számítógép alapú levelezés kezdeti szakaszában, az 1960-as években kialakított time-sharing¹ révén lehetővé vált az elektronikus üzenetek küldése ugyanazon rendszer/számítógép felhasználói között, majd Ray Tomlinsonnak köszönhetően 1971-ben már ennek hálózatos kiterjesztése is megszületett. Tomlinson az MIT meghívásából az internet elődjének, az ARPHANet (Advanced Research Project Agency Network) hálózatának kiépítésében segédkezett, amelynek rendszerében később létre is hozta az első e-mail programot.² Tomlinson voltaképpen összekötötte a SNDMSG elektronikus levelezésre használt programját és az ARPHANet hálózatát egy kísérleti file-transfer program segítségével, emellett meghatározta a mára közismert címszintaxist³ a @ szimbólum felhasználásával. Bár kezdetben csupán egy szűk szakmai réteg részére gyorsította meg jelentősen az információ továbbítását, mára egy mindenütt jelenlévő kommunikációs eszközzé vált, mely releváns forrásnak tekinthető a történeti vizsgálódások számára.

Az e-mail történetének több mint fél évszázada során számos átviteli protokollt és különböző szabványt határoztak meg, emellett megannyi felhasználói eszköz, e-mail kliens jött létre, melyek nagyfokú diverzitása nagyban megnehezíti az e-mailek archiválási folyamatát, amelyre jelenleg még nem áll rendelkezésre szabványosított megoldás. A problémát tovább bonyolítja, hogy az e-mailek rendkívül heterogén digitálisan keletkezett (born-digital) objektumok, melyek különböző fájlokat és adatokat tartalmazhatnak, a különböző szolgáltatók és kliensek pedig különféle sztenderdeket követhetnek. Mivel a formátumok idővel elavulhatnak, így ezen dokumentumok hozzáférhetősége megszűnik, fontos hosszútávú megőrzésük és integritásuk biztosítása.

A Digitális Örökség Nemzeti Laboratórium (DH-LAB) born digital alprojektje a digitálisan keletkezett kulturális örökség hosszútávú megőrzésének,⁴ archiválásának és kereshetővé tételének módszertanát hivatott kidolgozni az adatvesztés minimalizálásának célkitűzésével.⁵ Az aktuális pilot projektünk fókuszába a kutatói e-mail hagyatékok fent jelzett szempontok szerinti vizsgálatát állítottuk.

1 A time-sharing lehetőség voltaképpen a számítási erőforrások megosztott használatát jelentette a számítástechnikában. Mivel a számítási kapacitás és a (A time-sharing történetéhez lásd: J.A.N. Lee, "Claims to the Term 'Time-Sharing,'" *IEEE Annals of the History of Computing* 14, no. 1 (1992): 16–54, <https://doi.org/10.1109/85.145316>.)

2 Az e-mail fejlődésének technikai részleteiről lásd Craig Partridge, "The Technical Development of Internet Email," *IEEE Annals of the History of Computing* 30, no. 2 (April 2008): 3–29, <https://doi.org/10.1109/MAHC.2008.32>.

3 user@host

4 Az alprojektben résztvevő munkatársak: Alföldi István (born digital szakértő), Dobás Kata (ITdata specifikáció), Fellegi Zsófia (digitális filológiai szakértő), Indig Balázs (Python programozás), Szemigán Dorottya Henrietta (born digital szakértő), Palkó Gábor (projektvezető).

5 Emellett hosszútávú terveink közé tartozik ezen anyagok közzététele, illetve oktatói anyagok létrehozása is.

A DH-LAB DANUBE AI programja a határon túli hungarikumok, a magyar nyelvű vagy vonatkozású kulturális örökség felkutatásával foglalkozik, valamint kiemelt célja együttműködések kialakítása a közép-európai régió kutatási és oktatási intézményeivel, a tudáscsere és a tudományos hálózatépítés. E program biztosította továbbá a lehetőséget, hogy a Minerva Művelődési Egyesülettel együttműködésben kutatási és archiválási célra megkapjuk Kántor Lajos elektronikus levelezését.⁶

Kántor Lajos kolozsvári születésű, meghatározó jelentőségű kulturális és közéleti szereplő, termékeny kritikus, szerkesztő volt. Élete során közel 70 önálló és általa szerkesztett kötet jelent meg. Haláláig a *Korunk* kolozsvári kulturális folyóirat főszerkesztője volt, emellett kiterjedt kultúraszervező tevékenységet folytatott Kolozsváron, többek között alapító tagja és elnöke is volt (2006-tól egészen 2017-ben bekövetkező haláláig) a Kolozsvár Társaságnak, mely a város magyar kisebbségének gazdasági, társadalmi és kulturális integrálását tűzte ki célul. Tekintélyes életműve, ezen belül nagy terjedelmű levelezése – elsősorban, de nem kizárólagosan a határon túli magyar irodalomtörténetírás számára – értékes forrás. Elsődlegesen a *Korunk* szerkesztői levelezése számtalan érdekes és eddig ismeretlen adatot tartalmazhat irodalom-, művészet- és társadalomtörténeti vonatkozásban is.

Kántor Lajos e-mail hagyatéka tulajdonképpen hibrid jellegű, hiszen az összesen 6 különböző e-mail fiók leveleit tartalmazó hagyatékrész mellett a korai, csupán papíron megőrzött, kinyomtatott leveleket is az e-mail hagyatéka szerves részeként értelmeztük, ezeket digitalizáltuk.⁷

Az eredeti e-mailek biztosítása

Az első feladat a jogi kérdések tisztázása és a megállapodás aláírása után az örökösöktől és a kolozsvári Minerva Művelődési Egyesülettől megkapott elektronikus levelezés biztosítása volt. Az e-mailek átadása egy Google Gmail fiók hozzáférési adatainak megadásával történt. Kántor Lajos 2017-ben bekövetkezett halála óta a fiókot az örökösök kezelték.

Az e-mailek biztosítása a következő feladatokat jelenti:

- A Gmail fiók biztosítása
- MBOX mentés készítése
- Az MBOX állomány integritásának biztosítása

A Gmail fiókról fontos tudni, hogy ha egy ideig nem lépünk be, az e-mailek törlődnek. Ezt elkerülhetjük, ha beállítjuk a Beállítások \ Adatok és adatvédelem \ További lehetőségek \ Rendelkezzen digitális hagyatékaról szekció alatt, hogy huzamosabb inaktivitás esetén mi történjen az e-mailekkel és kit értesítsen a rendszer.

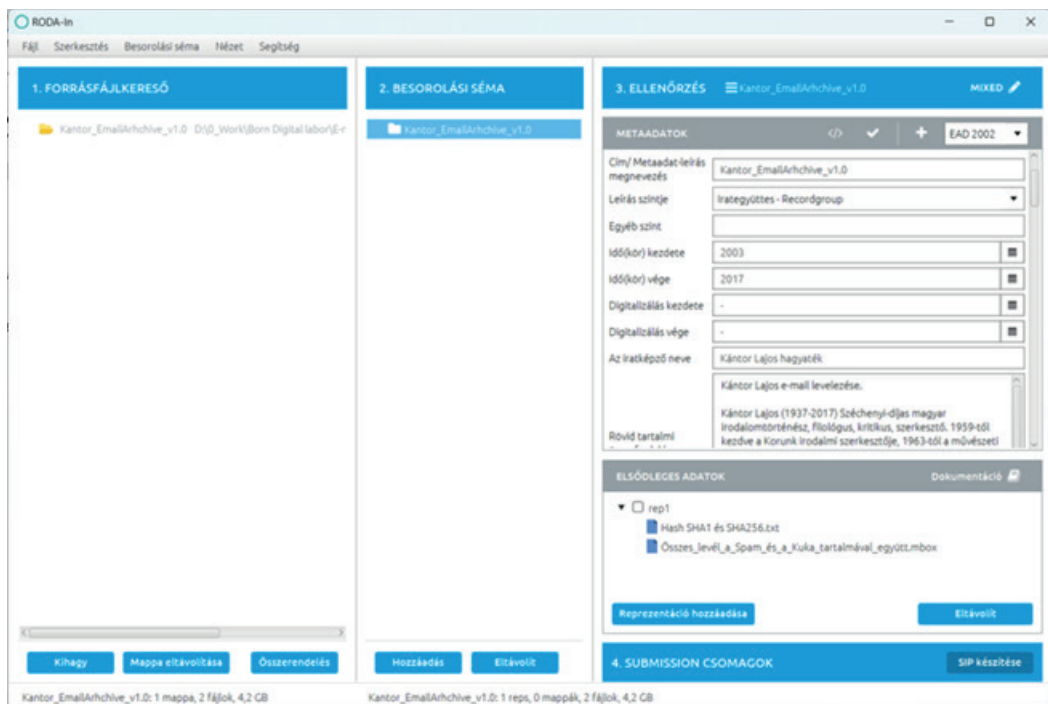
A Gmail fiókból a Google Takeout segítségével készítettünk MBOX mentést. Az MBOX egy széleskörűen támogatott sztenderd e-mail konténer formátum, amely több e-mail strukturált tárolását teszi lehetővé. A további feldolgozás során az MBOX állományokból indultunk ki, ezért fontos volt az állomány integritásának biztosítása. Ehhez a SHA (Secure Hash Algorithm)

⁶ A Kántor Lajos hagyatékat alapvetően a Minerva Művelődési Egyesület gondozza, ennek egy részét, a hagyatéka elektronikus levelezését bízták a DHLAB-ra.

⁷ A nyomtatott e-mailek szkennelését követően, a DH-LAB által fejlesztett másodlagos GUI felület mesterséges intelligencia alapú OCR (Optical Character Recognition) moduljának segítségével felismertettük nyomtatott levelek szövegét, így a továbbiakban a digitális hagyatékkal együtt tudtuk kezelni.

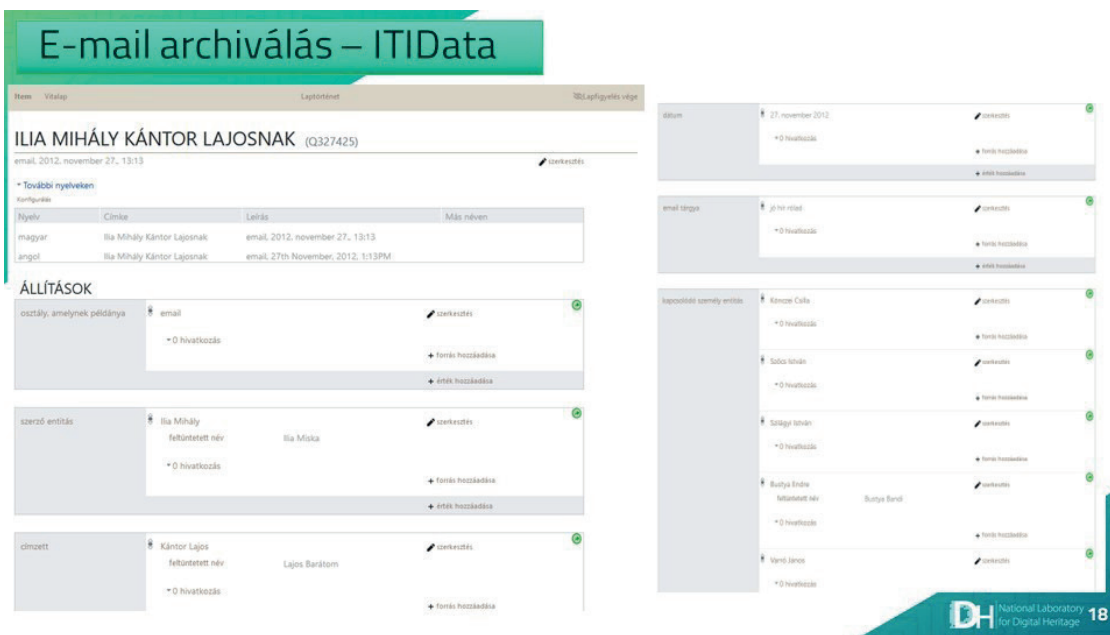
256 bites hash formátumát használtuk. Az algoritmus egy úgynevezett „visszafordíthatatlan és egyedi kivonatot” készít, amellyel a későbbiek során bitszinten ellenőrizhető, hogy nem történt változás az állományban. Az SHA-256 hash-t a Windows Power Shellbe beépített Get-FileHash eszközzel hoztuk létre.

Az MBOX fájlból ezek után készítettünk egy E-ARK SIP formátumú beadási csomagot. A Born-digital alprojekt kimondott célja az Európai Bizottság E-ARK/eArchiving programja által készített és támogatott digitális archiválási specifikációk és eszközök tesztelése, használatuk magyarországi előmozdítása. A digitális archiválás elméleti alapjait lefektető Open Archival Information System (OAIS) referenciamodell definiálja a beadási csomagot, de nem rendelkezik annak pontos tartalmáról és struktúrájáról. Az E-ARK projekt során kidolgozott E-ARK SIP egy szabványos csomag formátum az archívumnak átadandó állományok és metaadatok kezelésére. Az E-ARK SIP csomagot a RODA-in SIP-készítő eszközzel állítottuk elő.



1. ábra: E-ARK SIP készítése

A csomagot végül a CERN nyílt adatrepozitórium szolgáltatásának (Zenodo.org) segítségével archiváltuk. Ezekkel az eszközökkel biztosítottuk a hagyaték hosszútávú szabványos megőrzését, és a továbbiakban a feldolgozásra koncentrálnhattunk.



2. ábra: A Kántor hagyaték a Zenodo repozitóriumban

MBOX állományok feldolgozása

Az előbbieken archivált MBOX konténer a feldolgozás nyers bemeneti állománya. Ebben megtalálható az összes levél, függetlenül annak fejléceadataitól és tartalmától. A feldolgozás előtt el kellett végezni a levelek szűrését. A nyers állomány összesen 11245 e-mailt tartalmazott.

A következő leveleket szűrtük ki a feldolgozás első lépéseként:

Szűrési feltétel	Levelek száma
Összes levél	11245
Érzékeny tartalmú levelek (családon belüli levelezés)	779
Levélszemét és egyéb lényegtelen e-mail	1710
Duplikátumok, tömeges levelek	0
Maradt	8756

3. ábra: Elektronikus levelezés szűrésének feltételei

Először az érzékeny tartalmú levelek kerültek kiszűrésre. Az örökösök megnevezték azokat a családtagokat, akik sem címzettként, sem feladóként nem szerepelhettek a publikálandó levelek között. Az összes többi levelet elvben publikálhatónak nyilvánították. A következő csoport a levélszemét volt. A levélszemetet vagy eleve el sem mentették, vagy később törölték, mert alig találtunk biztosan kéretlen leveleket. Viszont sok olyan körlevél és hírlevél volt, ami Kántor Lajos halála után keletkezett. Ezeket is kiszűrtük. Megtartottuk viszont a halála előtt keletkezett hasonló leveleket, mert olyan elvek definiálása lehetetlennek tűnt, amelyek mentén egyértelműen lehetett volna dönteni arról, hogy melyik levél nem releváns a hagyaték összefüggésében. A duplikátumok azonosítására sem áll rendelkezésre megkérdőjelezhetetlen módszer, ezért a válaszlevélben megismételt, vagy különböző feladótól többször megkapott leveleket sem szűrtük ki.

A következő lépés a metaadatok (voltaképpen az e-mail fejléc [header] adatainak) kinyerése és elemzése volt. A legfontosabb e-mail metaadatokat a feldolgozás ezen korai szakaszában a Python programnyelven készült `email.parser` alkalmazással⁸ nyertük ki és rendeztük táblázatos adatformátumba.

A csak a metaadatok és a csatolmányok neveit tartalmazó táblázat segítségével különböző bontásokban (dátum, szerző, címzett, tárgy) lehet elemezni az e-maileket.

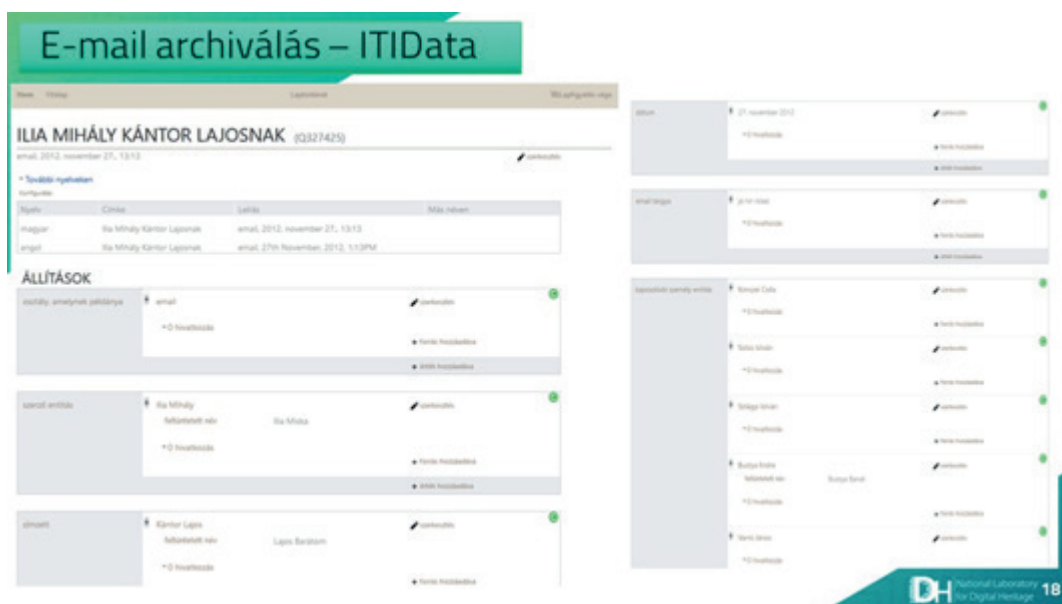
Papíralapú levelezés tapasztalatainak felhasználása

A pilot keretei között arra tettünk kísérletet, hogy a digitális kritikai levelezések kiadása során a DigiPhil projektben összegyűjtött tapasztalatokat összevevessük az elektronikus levelezések publikálásának gyakorlati kérdéseivel. A levelek szemantikus adatgazdagítására az azokban található bibliográfiai és filológiai adatokat az Irodalomtudományi Intézet szemantikus web alapú adatbázisát, az ITldata-t használtuk. Megvizsgáltuk annak a lehetőségét is, hogy a papír alapú levelezések digitális kiadásánál széles körben használatos jelölőnyelvi ajánlás, a Text Encoding Initiative alapján írjunk le e-maileket.

ITldata

Elsőként az ITldata hagyományos levelezések adatainak leírására készített specifikációját optimalizáltuk, figyelmet fordítva az e-mailek sajátosságaira. Az ITldata rendszerében minden e-mailt különálló entitásként képeztünk le, a rekordok egyedi azonosítót kaptak, az entitáshoz olyan adatokat rendeltünk, mint a szerző, címzett, küldés időpontja, fájlformátum stb. A szerző és címzett esetében, az adatbázisban entitásként szereplő rekordok adott e-maillal való összekapcsolásán túl feltüntettük a szerző aláírását és a címzett megszólítását, ezek sokszor a beceneveiket reprezentálják. Az egyes e-mailek rekordjainál kapcsolódó rekordként feltüntettük továbbá a levélben említett személy- és helyneveket. Az ITldata rendszerében új tulajdonságokat is létre kellett hoznunk a hagyaték adatainak kifinomult leírásának érdekében, ilyen például az e-mail tárgyának leírására szolgáló tulajdonság, ennek segítségével váltak sorrendezhetővé a levelek. Az e-mailekben említett művek bibliográfiai adatait szintén feltöltöttük a rendszerbe, azokat az e-mailt reprezentáló rekorddal összekapcsoltuk.

8 Lásd: <https://docs.python.org/3/library/email.parser.html>



4. ábra: ITIData: egyedi e-mail entitás és kapcsolódó állításai

TEI jelölőnyelv és XML séma felhasználása

A DigiPhil szakmai irányításával készült a levelezés digitális szövegkiadását reprezentáló TEI XML átírat specifikációja. A DigiPhilnek kéziratos levelezés kritikai kiadásainak digitális közzétételében volt gyakorlata,¹ az e-mailek specifikációjának kidolgozásakor ezeket a tapasztalatokat vettük alapul. Ahogyan arról Fellegi Zsófia beszámolt,² a DigiPhil a korábbi gyakorlatához képest változtatott a TEI XML specifikációin, a metaadatokat az ITIDatá-ban írja le, a TEI XML fájlok meghatározott pontjain pedig az adatbázisra hivatkozik. A specifikáció kialakítása ennek az új gyakorlatnak a figyelembevételével történt.

A digitális kiadás címét, a korábbi, a hagyományos filológiai gyakorlatból eredeztethető módon, a levél szerzőjének, címzettjének és a levél megírásának dátumából hoztuk létre. A TEI XML fájl fejlécében, amely a levélre vonatkozó metaadatokat tartalmazza, a „correspAction” jelölő alatt tüntettük fel a feladó és a címzett nevét, a levél küldésének pontos idejét, ezeket összekapcsoltuk az ITIData megfelelő entitásaival. Fontos kiemelni, hogy bár a DigiPhil gyakorlatában a metaadatok kifinomult leírása az ITIDatában történik, nem várható el a felhasználótól, hogy minden esetben együtt vizsgálja a szövegekölést és az adatbázist, így a leveleket reprezentáló TEI fájlokban is, karaktorsorok szintjén közölnünk kell a releváns adatokat.

1 Ehhez részletesebben lásd Móricz Zsigmond levelezésének kritikai kiadásáról szóló beszámolót: Cséve Anna, Fellegi Zsófia, Kómár Éva: *Móricz Zsigmond levelezésének (1892–1913) digitális kritikai kiadása*, Digitális bölcsészet 1. sz. (2018.) 159-174 <https://doi.org/10.31400/dh-hun.2018.1.227>, valamint Fellegi Zsófia, Palkó Gábor: *Arany-kéziratok és kritikai kiadások közzététele az Arany János Emlékévben*, Helikon 66. Sz. (2020) 82-98.

2 Fellegi Zsófia: *A digitális filológia infrastruktúrái. A DigiPhil megújulásáról*, Valós térben - az online térért, NIIF Networkshop konferenciák; 31., szerk. Tick József, Kokas Károly, Holl András, MTA Könyvtár és Információs Központ (Budapest:2022) 338-344. Ld. Még: Fellegi, Zsófia, & Palkó, Gábor. (2023, June 14). Publishing Digital Text Editions on the Semantic Web. Open Repositories 2023 (OR2023), Stellenbosch, South Africa. Zenodo. <https://doi.org/10.5281/zenodo.8091659>

E-mail archiválás – TEI XML

```

107 <div style="typed" type="email" xml:id="d.1">
108 <div style="typed" type="text" xml:id="p.1"> Kedves <persName>Lajos Barátom</persName> <idno corresp="Kántor Lajos"
109 type="ITIdata">Q146919</idno>
110 <persName>...
111 <p>nagyon jó hangulatot okozott leveled, a terved, hogy
112 kevesebbet fogsz utazni és többet otthon munkálkodni.
113 Ez jót fog tenni az egészségednek is.
114 Ma reggel olvastam <persName>Könczei Csilla</persName> <idno corresp="Könczei Csilla" type="ITIdata">Q330799</idno> leveleit és a
115 székenévsort, végül leíródott, hogy <persName>Szécs István</persName> <idno corresp="Szécs István" type="ITIdata">Q27515</idno>
116 milyen szerepet vállalt. Persze attól tartok, hogy SZI
117 van annyira cinikus, hogy ezt is mellőzi és tovább írja
118 marhaságait a <persName>Munkácsy Mihály</persName> <idno corresp="Munkácsy Mihály" type="ITIdata">Q23736</idno> <persName>Szilágyi Pista</persName>
119 Nekem szomorúság, hogy <persName>Bustya Endre</persName> <idno corresp="Bustya Endre" type="ITIdata">Q240437</idno> nevét is ott olvasom
120 <persName>Könczei Csilla</persName> <idno corresp="Könczei Csilla" type="ITIdata">Q330799</idno> jegyzékében, de <persName>Varró</persName> <idno corresp="Varró" type="ITIdata">Q330799</idno>
121 évvel ezelőtt <persName>Bodor Ádám</persName> <idno corresp="Bodor Ádám" type="ITIdata">Q41694</idno> beszélt arról, hogy <persName>Veress Zoltán</persName> <idno corresp="Veress Zoltán" type="ITIdata">Q60070</idno>
122 árulta el, ill. tőle tudták a szekunál a legbensőbb dolgokat kis csoportjukról, de <persName>BÁ</persName> <idno corresp="Bodor Ádám" type="ITIdata">Q41694</idno> azt is mondta, hogy többször szólította föl
123 <persName>Veress Zoltán</persName> <idno corresp="Veress Zoltán" type="ITIdata">Q60070</idno>, hogy mondja el a történeteket, de soha választ nem
124 kapott. Kérésére. A romániai magyar szellemi élet egészét érintik,
125 biztos, hogy még sok minden más megvilágítása kerül
126 írtam tán, hogy nov. 23-án <placeName>Belgrádban</placeName> <idno corresp="Belgrade" type="ITIdata">Q208</idno> meghalt <persName>Szava Babic</persName> <idno corresp="Sava Babic" type="ITIdata">Q161563</idno>
127 módosító, aki egy könyvtárnyi magyar irodalmat ültetett
128 át szerbre. Az egyik utolsó nagy munkája Madách fő mű-
129 vének a lefordítása.
130 Olvastam, hogy <placeName>Kolozsvár</placeName> <idno corresp="Cluj-Napoca" type="ITIdata">Q258</idno> lesz az ifjúsági kulturális főváros.
131 Ez jó előkép <placeName>Zsuzsanna</placeName> <idno corresp="Zsuzsanna" type="ITIdata">Q15217</idno>
132 legyen belátható időn belül. Igaz, Romániában <placeName>Temesvár</placeName> <idno corresp="Temesvár" type="ITIdata">Q15217</idno> is
133 tervezeti ezt, nagy versenytársa lehet <placeName>Kolozsvárnak</placeName> <idno corresp="Cluj-Napoca" type="ITIdata">Q258</idno>.
134
135 <p xml:id="p.3">Ölel és üdvözlök!</p>
136 <p xml:id="p.4"> <seg type="signature"><persName>Ilia Miska</persName> <idno corresp="Ilia Mihály" type="ITIdata">Q146474</idno></persName></seg>
137

```

5. ábra: TEI jelölőnyelv XML séma kódrészlet, kapcsolódó entitások kódolása, „idno” identifier használata

Azöveggközlés betűhű, tehát karakterszinten követi az e-mailben található szöveget, azonban a levelek szövegében is elvégeztük az adatgazdagítást, tehát például a személyneveket jelölővel láttuk el és összekötöttük az ITIdatával.

Az egyéb technikai adatok (*paradata*) rögzítése még nem definiált. Egy lehetséges megoldásként javasoljuk a fájl méretének feltüntetését a „measure” elemben a @type=„quantity” és a @unit=„kb” attribútum-értékekkel, emellett még a fájlformátum, illetve a hordozó fájl típus feltüntetése is szükséges lesz.

További tervek

Ahogy azt fentebb is jeleztük, a Born-digital alprojekt pilotjának fontos célja volt megvizsgálni, hogy a papíralapú levelezés feldolgozása és publikálása során szerzett tapasztalatokat hogyan lehet hasznosítani az elektronikus levelek feldolgozásában. A Digitális Örökség Nemzeti Laboratórium egy másik kiemelt projektje a webaratás. A labor folyamatosan végzi a kutatási és innovációs szempontból releváns webes források kiválasztását, aratását és az ehhez szükséges technológiák fejlesztését. A webaratás projekt célja az archiváláson kívül az összegyűjtött anyagok kutathatóvá tétele. A webaratás és az adatok publikálása saját fejlesztésű Python-nyelvű szoftverekkel történik. Az e-mailek feldolgozása közben jöttünk rá, hogy valószínűleg ez az eszközkészlet is jól hasznosítható az e-mailek további tartalmi feldolgozása és publikálása céljából.

A tervezett feldolgozás során először a teljes metaadat-készletet nyerjük ki egy saját fejlesztésű Python programmal. Az e-mailek elvben szabványos metaadatkészlete az évtizedek alatt felmerülő új igényekre válaszul úgy változott, hogy sokszor ugyanazt a metaadatot a különböző szolgáltatók és szoftverképzítők máshogy nevezik. Ez azt jelenti, hogy a továbblépés előtt az e-mail meta- és para-adatok gondos elemzése szükséges. A metaadatok elemzése és szűrése után a levelek szövegét tartalmazó levéltörzset kell szükség esetén dekódolni és további feldolgozásra alkalmas formára hozni. Ezt az állományt töltjük be végül egy relációs adatbázisba, illetve további tervünk, hogy egy e-mailek elemzésére testre szabott keresőalkalmazás segítségével meg tudjuk jeleníteni az adatokat, így a webaratás projekt keretei között létrehozott kereső és vizualizációs alkalmazáshoz hasonlóan statisztikák, kimutatások készülhetnek az e-mailekről is.

Bibliográfia

- Balázs Indig, Árpád Knap, Zsófia Sárközi-Lindner, Mária Timári, Gábor Palkó. "The ELTE.DH Pilot Corpus – Creating a Handcrafted Gigaword Web Corpus with Metadata" In the Proceedings of the 12th Web as Corpus Workshop (WAC XII), pages 33-41 Marseille, France 2020.
- Kirschenbaum, Matthew G., Richard Ovenden, Gabriela Redwine, and Rachel Donahue. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. CLIR Publication, no. 149. Washington, D.C: Council on Library and Information Resources, 2010.
- Lee, J.A.N. "Claims to the Term 'Time-Sharing.'" *IEEE Annals of the History of Computing* 14, no. 1 (1992): 16–54. <https://doi.org/10.1109/85.145316>.
- Partridge, Craig. "The Technical Development of Internet Email." *IEEE Annals of the History of Computing* 30, no. 2 (April 2008): 3–29. <https://doi.org/10.1109/MAHC.2008.32>.
- Prom, Christopher. "Preserving Email (2nd Edition)." Digital Preservation Coalition, May 2019. <https://doi.org/10.7207/twr19-01>.
- Radicati, Sara. *Electronic Mail: An Introduction to the X.400 Message Handling Standards*. Uyless Black Series on Computer Communications. New York: McGraw-Hill, 1992.
- Task Force on Technical Approaches to Email Archives, Andrew W. Mellon Foundation, and Digital Preservation Coalition, eds. *The Future of Email Archives: A Report from the Task Force on Technical Approaches for Email Archives, August 2018*. CLIR Publication, no. 175. Washington, DC: Council on Library and Information Resources, 2018.

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
Koczká 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 ITIdata 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