

A TUDOMÁNY, AZ OKTATÁS ÉS A KÖZGYŰJTEMÉNYI KISZOLGÁLÁS ÚJ INFORMATIKAI SZINERGIÁI

**NETWORKSHOP 2026
35. Országos Informatikai Konferencia**

**2026. március 31–április 2.
Debreceni Egyetem, Debrecen**

Szerkesztette: Tick József, Kokas Károly, Holl András

**HUNGARNET Egyesület
Budapest, 2026**



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Lukács Katalin

Networkshop 2026 konferencia előadásainak közleményei

Debreceni Egyetem, Debrecen

2026. március 31–április 2.

ISBN 978-615-6792-29-7

DOI: <https://doi.org/10.31915/NWS.2026>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2026

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

ELEKTRONIKUS LEVELEZÉSKIADÁS MINT SZEMANTIKUS HÁLÓZAT

ELECTRONIC CORRESPONDENCE PUBLISHING AS A SEMANTIC NETWORK

Szemigán Dorottya Henrietta
Digitális Örökség Nemzeti Laboratórium

szemigan.dorottya@btk.elte.hu

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

Dobás Kata
ELTE HTK Irodalomtudományi Kutatóintézet

dobas.kata@abtk.hu

ORCID: [0009-0009-7632-8276](https://orcid.org/0009-0009-7632-8276)

Palkó Gábor
ELTE HTK Irodalomtudományi Kutatóintézet

palko.gabor@abtk.hu

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

Fellegi Zsófia
ELTE HTK Irodalomtudományi Kutatóintézet

fellegi.zsofia@abtk.hu

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

Nemes-Jakab Éva
ELTE HTK Irodalomtudományi Kutatóintézet

jakabev@gmail.com

ORCID: [0009-0001-4516-0935](https://orcid.org/0009-0001-4516-0935)

Sárközi-Lindner Zsófia Mónika
ELTE HTK Irodalomtudományi Kutatóintézet

zsofia.monika@gmail.com

ORCID: [0000-0002-2558-0633](https://orcid.org/0000-0002-2558-0633)

Absztrakt

A tanulmány Kántor Lajos kolozsvári irodalomtörténész e-mail-hagyatékának feldolgozásán keresztül mutatja be, miként alakítható ki egy born-digital irodalmi hagyaték teljes feldolgozási workflow-ja az archiválástól és a hosszú távú megőrzés biztosításától az MBOX-alapú feldolgozáson át a TEI XML-ben megvalósuló, szemantikus adatbázishoz kapcsolt digitális e-mail-kiadásig. Kutatási projektünk fókuszában Kántor Lajos és Ilia Mihály elektronikus levelezésének digitális tudományos kiadásra való előkészítése áll. Esettanulmányunkon keresztül bemutatjuk a korpusz kijelölésének szempontjait, a metaadatok strukturálását, az ITIdata szemantikus adatbázissal való összekapcsolást, a levelek e-mail-specifikus TEI XML-struktúráját, adatvédelmi és anonimizálási elveinket, valamint a lektorálást támogató ellenőrző- és munkafelületünket.

Kulcsszavak: e-mail, szemantikus adatbázis, colab notebooks, digitális kiadás, e-mail-hagyaték, GDPR, ePrivacy

Abstract

The paper presents how a complete workflow for processing a born-digital literary archive can be developed through the processing of the e-mail legacy of Lajos Kántor, a literary historian from Cluj-Napoca: from archiving and ensuring long-term preservation, through MBOX-based processing, to a digital e-mail edition implemented in TEI XML and connected to a semantic database. The focus of our research project is the preparation of a digital scholarly edition of the electronic correspondence between Lajos Kántor and Mihály Ilia. Through this case study, we present the criteria for defining the corpus, the structuring of metadata, integration with the ITIdata semantic database, the structure of e-mail-specific TEI XML files for the letters, our principles for data protection and anonymization, and the review interface developed for proofreading.

Keywords: e-mail, semantic database, colab notebooks, digital edition, e-mail heritage, GDPR, ePrivacy

1. Bevezetés:

A 20–21. századra az e-mail a kulturális, tudományos és intézményi kommunikáció, így a kortárs hagyatékok egyik meghatározó born-digital történeti forrástípusává vált. Az e-mail, amely egyszerre szöveganyag és metaadatokkal rendelkező rekord, nem tekinthető pusztán a papíralapú levél digitális megfelelőjének, hiszen a levéltest mellett fejléceadatokat, címzési mezőket, időbélyegeket, válaszkapcsolatokat, karakterkódolási sajátosságokat, továbbítási információkat és csatolmányokat is tartalmazhat. Feldolgozása ezért csak a technikai, szerkezeti és relációs rétegek együttes figyelembevételével végezhető el.

Az e-mailek megőrzése, feltárása és hozzáférhetővé tétele összetett technikai, intézményi és jogi feladat (Task Force on Technical Approaches for Email Archives, 2018). A szakirodalom az e-mailarchiválást életciklus-alapú kurátori folyamatként írja le, amelyben a létrehozás, használat, exportálás, megőrzés, leírás és hozzáférhetővé tétel egymással összefüggő munkafázisokként jelennek meg (Pennock, 2006; Prom, 2019). Jaillant érvelésével összhangban a technikai megmentés azonban önmagában még nem jelent kutatói hozzáférhetőséget. A megfelelő leírás, kereshetőség, jogosultságkezelés és jogilag szabályozott hozzáférési modell¹ nélkül az e-mailgyűjtemény továbbra is zárt, a kutatás és a nyilvánosság számára ténylegesen nem hozzáférhető archívum maradhat (Jaillant, 2019).

Bár a TEI Correspondence SIG és a Correspondence Metadata Interchange Format fontos mintát adnak a levelezési események strukturált leírására (Dumont, 2016; TEI Correspondence SIG, n.d.), az e-mail esetében a pontos időbélyegek, a többes címezés, a válaszláncok, a csatolmányok, a Message-ID és a nem publikálható technikai metaadatok is a kiadás módszertani döntéseinek részévé váltak. Jelen tanulmányunk bemutatja, hogyan dolgoztuk ki a fentiek figyelembevételével a digitális bölcsészeti területen még egyedülálló digitális emialkiadás módszertani alapjait.

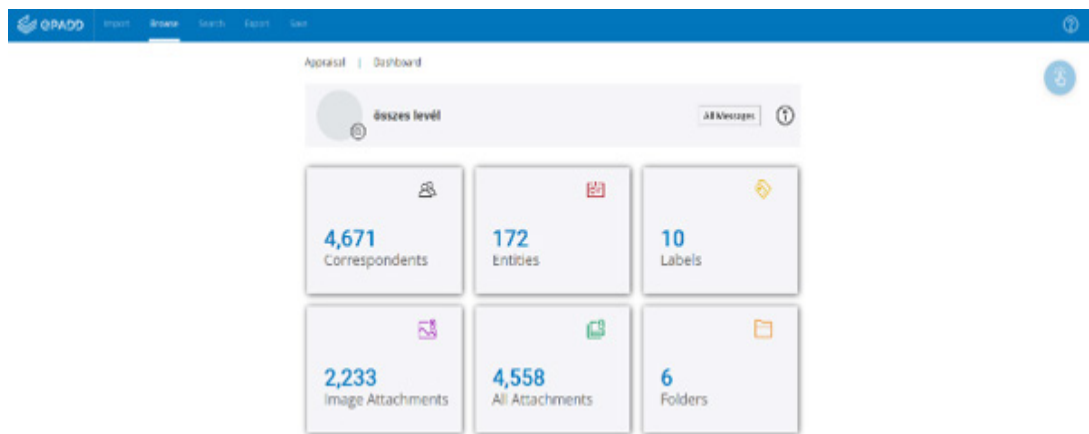
2. A Kántor-e-mail-hagyaték feldolgozás teljes workflow-ja

E-mail-hagyaték feldolgozás lehetséges módszertanának létrehozásával foglalkozó kutatási projektünk Kántor Lajos kolozsvári irodalomtörténész, filológus, kritikus és szerkesztő elektronikus levelezésének archiválását, feldolgozását és részleges forráskiadását célozta meg. A munkafolyamat első szakasza az e-mailállomány exportjára, integritásbiztosítására, szabványos archiválására és hosszú távú megőrzésére irányult; ennek módszertani tanulságait korábbi tanulmányunk foglalta össze (Alföldi et al., 2023a). A második szakasz az MBOX-formátum feldolgozására, kereshetővé tételére, tisztítására és vizualizációs előkészítésére koncentrált, amelynek eredményeit az NLP4DH 2023 konferencián mutattuk be (Indig et al., 2023). A jelen írás erre a két előzményre építve a harmadik munkafázist, a Kántor Lajos és Ilia Mihály közötti elektronikus levelezés TEI XML-alapú, szemantikus adatbázishoz kapcsolt digitális forráskiadásának létrehozását tárgyalja.²

1 A hozzáférés e-mail-hagyatékok esetében jogi és etikai kérdés is, mivel ezek harmadik személyek adatait, nem nyilvános elérhetőségeket, érzékeny információkat és visszaazonosításra alkalmas technikai metaadatokat tartalmazhatnak.

2 Az e-mail TEI-alapú reprezentációjára lásd Syd Bauman és Elisa Beshero-Bondar kísérletét a TEI-L nyilvános LISTSERV-archívumával. A vállalkozás fontos párhuzam, de nem közvetlen előzmény: a két projekt párhuzamosan zajlott, a TEI-L-moddal kísérleti jellegű, és egy nyilvános levelezőlista archiválására, nem pedig személyes irodalmi e-mail-hagyaték digitális kritikai forráskiadására irányul. A jelen projektben a TEI XML-specifikációt a szemantikus adatbázis-kapcsolat, az anonimizálás, a hozzáférés-szabályozás és a technikai metaadatok adatvédelmi szűrése is alakította. Magyar kontextusban Kalcsó Gyula és Szűcs Kata Ágnes tanulmánya, valamint az MNMKGK OSZK DBK Katalist-projektje említendő: ezek fő hangsúlya azonban az e-mailek hosszú távú közgyűjteményi megőrzésén, nem a digitális kiadási modell kialakításán van.

3. A forráskiadás korpusza



1. ábra: a teljes e-mailállomány az ePADD szoftver megjelenítésében

A digitális forráskiadás korpuszát a teljes hagyaték egy főként irodalomtörténetileg releváns része, Kántor Lajos és Ilia Mihály levelezése adja. A hagyaték hibrid jellegű, a benne foglalt teljes e-mailállomány 11 246 digitális objektumot³ és 1424 papír alapon fennmaradt, kinyomtatott e-mailt tartalmaz⁴. A forráskiadásra kijelölt Ilia Mihály és Kántor Lajos levelezésére vonatkozó korpusz 69 nyomtatott e-mailből, 5 kéziratból és 360 digitális levélből, e-mailből áll, a kiadás így összesen 434 levelet közöl 2003–2017 közötti időszakból. A korpusz kiválasztását a technikai vagy mennyiségi szempontok mellett az határozta meg, hogy a Kántor és Ilia levelezése egy adott történeti pillanatban a magyar⁵ irodalmi közeg, szerkesztői és tudományos kapcsolatrendszer olyan szeletét teszi láthatóvá, amely egyszerre kínál irodalom- és intézménytörténeti, valamint művelődés- és politikatörténeti kutatási lehetőségeket.

4. Korpusz feldolgozása, tömeges betöltés, TEI XML e-mail-specifikációk

4.1 Adatfeldolgozás

A hagyaték hibrid jellege az adatfeldolgozás szempontjából külön módszertani kihívást jelentett. Az e-maileket ezért három különböző bibliográfiai leírással rendelkező e-mail-típusként kezeltük, melyek metaadatait ennek mentén tabuláris formában rendszerez-

3 A 2002–2023 közötti időszakból, ti. az e-mailcím a Kántor 2017-ben bekövetkezett halála után is, az archiválásig aktív volt.

4 Az 1997–2016 közötti évekből származó e-mailek még kiegészülnek adott esetben a nyomtatott e-mailekre rögzített válaszevelek kéziratával is.

5 Magyar nyelvű, azaz belföldi és határon túli egyaránt.

tünk.⁶ Ennek több előnye is kirajzolódott, egyfelől a katalogizálás/leltározás funkciója, az adatok strukturálása, valamint a tömeges betöltéshez való előkészítése.

A kinyomtatott e-mailek szkennelt másodpéldányainak optikai karakterfelismerése a Transkribus szoftverével történt, az MBOX-állomány feldolgozásához pedig a Stanford University által fejlesztett, open source mbox-megjelenítő ePADD és a full MIME supporttal rendelkező Windows MBox Viewer szoftvereket használtuk. Ez utóbbival több karakterkódolási problémát is sikerült áthidalni.

4.2 Tömeges betöltés, ITIdata

A kinyert metaadatok strukturálását két célkitűzés mentén végeztük el a projekt során: 1. a létrehozott specifikáció mentén megtörténhessen a félautomatikus betöltés a QuickStatements eszköz⁷ segítségével az ITIdata⁸ elnevezésű, az ELTE HTK Irodalomtudományi Kutatóintézet szemantikus adatbázisába, 2. a TEI fájlok headerjeinek félautomatikus létrehozását is támogassa.

A specifikáció kidolgozásánál a schema.org e-mailekre kidolgozott leírását vettük alapul, valamint figyelembe vettük az ITIdata már meglévő adatainak struktúráját, elsősorban a klasszikus, papíralapú levelezéskiadásokhoz (Arany János és Jókai Mór levelezése) készült tulajdonságok (property) listáját. A metaadatokat excel táblázatokban rögzítettük, formázásukat a QuickStatements-el való betöltéshez igazítottuk.

A betöltés előkészítésekor megtörtént a nevek azonosítása, névtérre fektetése, a részletes annotációk elkészítése, amelynek során az e-mailekben említett nevek, művek, földrajzi nevek és intézmények entitásazonosítása is végbement.

Mivel a Kántor-Ilia-e-mail-hagyaték az e-mailezésnek egy köztes állapotát rögzíti, amelyben az e-mailek még kinyomtatásra is kerültek, és feladójuk sokszor ezekre rögzítette válaszáát (kézírással), ezért a szemantikus kapcsolatok kidolgozásánál ezt is figyelembe kellett vennünk.⁹

6 A digitális formában fennmaradt leveleket e-mailként, a kinyomtatott e-maileket e-mailmásolatként, míg a kinyomtatott e-mailekre kézírással írt válaszleveleket kéziratként kezeltük.

7 <https://www.wikidata.org/wiki/Help:QuickStatements/hu>

8 Az adatbázisról: Dobás, Kata és Fellegi, Zsófia és Palkó, Gábor (2023) *A kis gömböc meséje – az ITIdata irodalomtudományos adatbázis fejlesztése 2022–2023-ban*. In: Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé. HUNGARNET Egyesület, Budapest, pp. 192–198. Online hozzáférés: <https://real.mtak.hu/182937/> Az adatbázis elérhetősége: https://itidata.abtk.hu/wiki/Main_Page

9 <https://itidata.abtk.hu/wiki/Item:Q492429> – az e-mail nyomtatott és elektronikus változata közötti szemantikus kapcsolatokat létrehoztuk, a kapcsolat fajtáját minősítőben rögzítettük. A teljes korpusz elérhetősége az ITIdata-ban: <https://tinyurl.com/27dyceu6>

kapcsolatban áll	Ilia Mihály – Kántor Lajosnak, 2009. január 06. 11:44
típus	nyomtatás
▼ 0 hivatkozás	

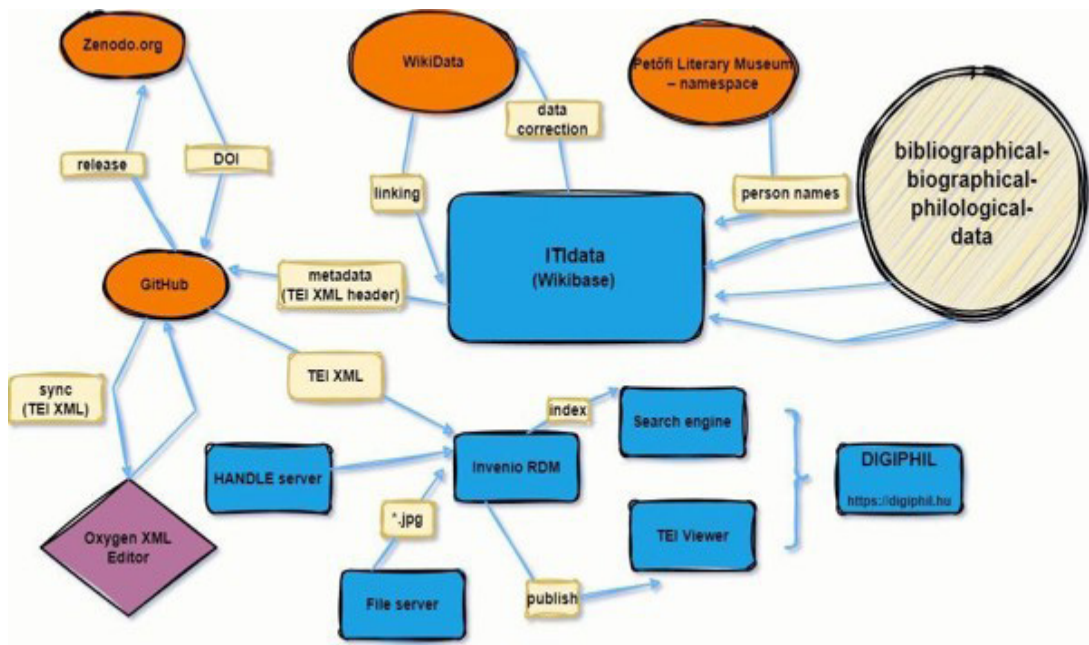
2. ábra részlet az ITIdata egyik rekordjának szemantikus kapcsolatairól

4.3 TEI XML e-mail-specifikációk

Az e-mailek publikálásának PDF-formátumú általános gyakorlatával szemben digitális kiadásunk – a tudományos igényesség jegyében – a TEI XML szabványos, interoperábilis, ember és gép számára egyaránt olvasható, a hosszútávú megőrzést, összetett keresést is biztosító formátum mellett döntött.

A teiHeaderben szerepet kap több az e-mailre vonatkozó metaadat, az egyes levelek részletes bibliográfiai adatait tartalmazó ITIdata rekodok azonosítói, a küldés pontos idejére, valamint a levelek közötti kapcsolatokra vonatkozó információk és az egyes e-mailek PID azonosítója. Az elektronikus levelek címét egységes szintaxis alapján alakítottuk ki: a cím a küldő nevéből, a címzett nevéből, valamint a küldés dátumából és pontos idejéből áll – a küldő és a címzett nevét gondolatjel választja el egymástól. A TEI XML-struktúrában a levél címe, az e-mail eredeti tárgya, valamint a levél szövege a <body> elemekben kapott helyet. A levelek adatgazdagítása során a szövegben szereplő személyneveket, településeket, intézményeket, műveket, műcímeket és dátumokat, továbbá a bibliográfiai adatokhoz, a kritikai apparátushoz és a különböző szövegváltozatokhoz kapcsolódó információkat a megfelelő TEI-elemekkel jelöltük.

A korpusz részletes annotálása számos lehetőséget kínál a levelezés további vizsgálatára, és új dimenziókat nyit meg annak digitális bölcsészeti eszközökkel és módszerekkel történő tartalmi feltérképezésében. A két szerző levelezésében említett, irodalomtörténetileg releváns szereplők azonosítása révén létrehozhatóvá válhat egy meghatározott korszak irodalmi közegének hálózati reprezentációja, hasonlóan a levelezésben említett művek kapcsolatrendszeréhez. A törzsszövegben előforduló numerikus dátumok és szöveges időutalások annotálása szintén fontos eleme volt a feldolgozásnak, mivel ezek az e-mail időbélyegével és a küldés dátumával összevetve lehetővé teszik annak vizsgálatát, hogy a felek milyen gyakorisággal tematizálnak aktuális eseményeket. Ezáltal elkülöníthetővé válhat a jelenre és közelmúltra, illetve a távolabbi múlttra vonatkozó párbeszéd.



3. ábra: DigiPhil infrastruktúra

5. Adatkezelési stratégiák

Az Ilia Mihály és Kántor Lajos közötti e-mailváltás kritikai forráskiadása – a levéltest és a metaadatok adatszerkezetének feltárása, szemantikus adatbázisban való rögzítése, valamint TEI XML-kódolása mellett – az adatkezelés jogi és etikai kereteinek újragondolását is szükségessé tette.

Az e-mail metaadatokat tartalmazó réteg (pl. e-mail-címek, szerverinformációk, IP-címek, továbbítási láncok, időbélyegek, CC/BCC mezők, csatolmányok és rejtett címzettek) adott esetben érzékenyebb információkat hordozhatnak, mint a tartalmi szöveg. Az Európai Unió jogértelmezése szerint ezek az adatok az elektronikus hírközlési adatok körébe tartoznak, így a GDPR mellett az ePrivacy-szabályozás hatálya alá is esnek.

Ebből következően az e-mailkiadás során nem volt elegendő a levélszöveg klasszikus anonimizálása vagy pszeudonimizálása, hanem különös figyelmet kapott a Message-ID, a routing lánc, az IP-címek, a levelezőszerver-információk, valamint a kézbesítési és fejléc-metaadatok kezelése, és ezek publikálhatóságának korlátozása.

A kutatás adatkezelési protokollja a tudományos feldolgozás céljainak és a személyes adatok védelmének egyensúlyára épül. Ennek során figyelembe vettük a 2016/679/EU rendeletet (GDPR), különösen a tudományos kutatási célú adatkezelésre vonatkozó 89. cikket, a 2002/58/EK irányelvet (ePrivacy), valamint az információs önrendelkezési jogról szóló 2011. évi CXII. törvényt (Infotv.). A jogi keretet a hagyaték jogtulajdonosaival – Ilia Mihály

esetében, valamint a Kántor Lajos-hagyatékot gondozó Minerva Művelődési Egyesülettel – kötött szerződéses megállapodások egészítik ki.

A hozzáférés szabályozása két szinten történt. A nyilvános repozitóriumba kizárólag válogatott és anonimizált e-mailek kerültek, míg a teljes, feldolgozatlan korpusz zárt környezetben marad, amelyhez kizárólag a jogtulajdonosok és a kutatásban részt vevő szakemberek férhetnek hozzá, kutatási engedély alapján.

Az anonimizálási és adatminimalizálási elvek egyrészt a levéltestre, másrészt a TEI-rekord metaadatokat tartalmazó <teiHeader> elemére terjednek ki. A TEI fájlban a levél szövegét közlő <body> elemében strukturált módon, a <gap> elem két megkülönböztetett típusát használtuk e célra: személynevek esetében <gap reason="anonymised" unit="persName" extent="1"/>, elérhetőségi és kommunikációs adatok (e-mail, telefonszám, lakcím) esetében <gap reason="anonymised" unit="email/phoneNumber/address" extent="1"/>, míg érzékeny vagy nem rekonstruálható tartalmi egységek esetében <gap reason="anonymised" unit="word" extent="2"/> került alkalmazásra.

A <teiHeader> sem az eredeti e-mail fejléceket reprezentálja, hanem azok kutatási célú, adatvédelmi szempontok szerint szűrt és normalizált megfelelőjét. A nyilvánosan elérhető rekordokban a kommunikációs infrastruktúrára vonatkozó érzékeny vagy potenciálisan visszaazonosításra alkalmas adatok (például szerver-útvonalak, IP-alapú információk) nem jelennek meg.

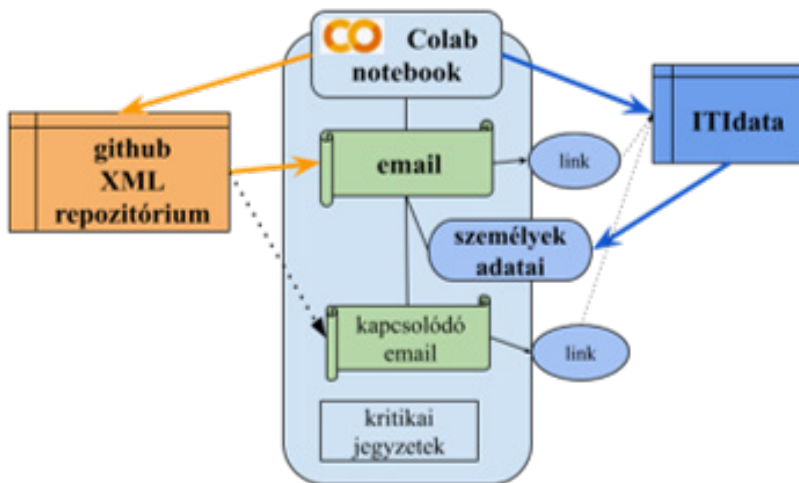
A Kántor Lajos és Ilia Mihály közötti e-mailváltás kritikai forráskiadása ezen gyakorlati szempontok bevezetésével Magyarországon elsőként valósít meg olyan publikációs modellt, amelyben a filológiai, információtechnológiai és adatvédelmi szempontok egységes rendszerben kerülnek integrálásra.

6. Lektorálási munkafolyamat

A lektorálási folyamathoz egy összetett, kollaboratív környezetet kellett létrehozni, ugyanis rugalmasan és egyidőben rendelkezésre kellett állnia a levelekről készült friss TEI XML-fájloknak, a levelek Wikibase reprezentációjának (ITIdata), valamint az azonosított entitások adatlapján szereplő tulajdonságoknak. Ennek összefogására és a lektor felé való világos és letisztult előkészítésére a Google Colab-ot mint ingyenes és testreszabható eszközt alkalmaztuk. A Google Colab notebookok böngészőből futtatható, felhőalapú Python-környezetet¹⁰ biztosítanak, amelyben a kódcellák, szöveges magyarázatok és vizualizációk egyetlen dokumentumban kezelhetők. A felhasználók telepítés nélkül futtathatnak

¹⁰ A Python egy könnyen tanulható, magas szintű programozási nyelv, amely egyszerű szintaxisával és jó olvashatóságával támogatja a gyors fejlesztést. <https://docs.python.org/3/>

Python-kódot, külső könyvtárakat használhatnak, fájlokat tölthetnek be.¹¹ Az e-mailek XML-jeit tartalmazó GitHub repozitórium integrációja lehetővé tette, hogy a lektor közvetlenül az eredeti fájlok legfrissebb verziójával dolgozhasson. Minden munkamenet a repozitórium installálásával és a Python programkód törzsének aktiválásával kezdődik, amely egy-egy kódcella futtatását jelenti, egyszeri műveletként elvégezhető. A notebookkal való további interakciók sem igényelnek programozói tudást, az egyes levelek lektorálásához szükséges információk előkészítését ugyanis az aktuális XML-fájlnev megadása után elvégzi a program.

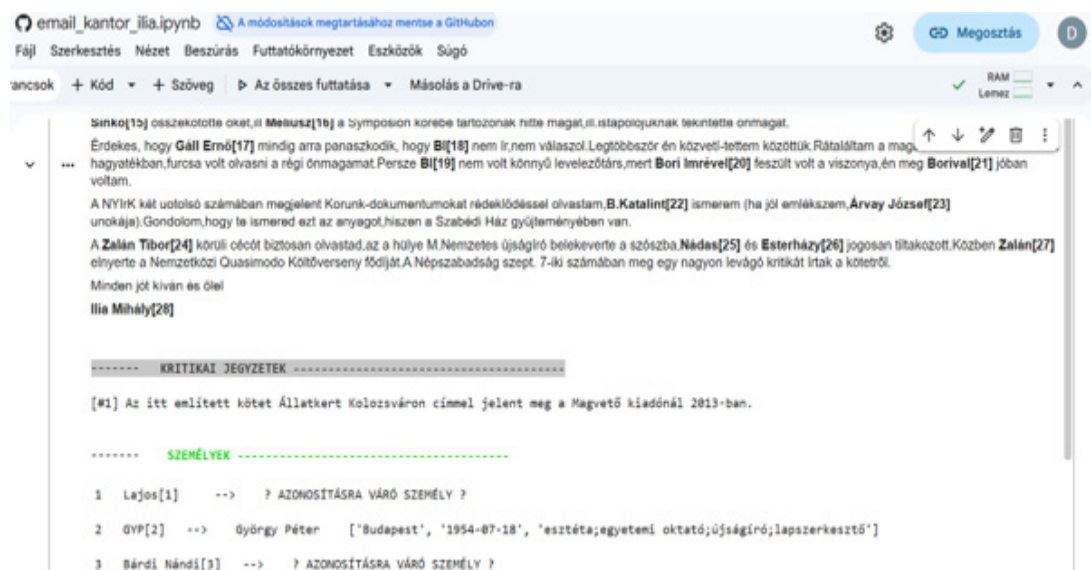


4. ábra: Colab notebook infrastruktúra

Futtatás után a lektor számára a notebookban megjelenik a levél szövege, illetve az ITIdata adatlapjára mutató kattintható link. A szöveg az XML-címkéket elhagyva látható, ugyanakkor az azonosított entitások félkövérrel és számozással, a kritikai jegyzetek számozással, az anonimizált részek külön kiemeléssel jelennek meg benne. A kritikai jegyzetek szövegét – amennyiben tartozik a levélhez – közvetlenül a levél alá írja ki a notebook, így ellenőrizhető a helyességük, illetve a lektor további jegyzeteket, vagy ezek kiegészítését javasolhatja. Az azonosítandó személyekről táblázatos formában olvashatóak az ITIdata adatlapjukról valós időben lekérdezett, kulcsfontosságú állítások (születési hely, születési idő, foglalkozás), melyek alapján a lektor értékelheti, hogy helyes volt-e az előzetes azonosítás, avagy javasolhat módosítást, azonosíthatja az entitást, amennyiben előzetesen üresen lett

¹¹ Továbbá igény szerint GPU- vagy TPU-erőforrásokat is igénybe vehetnek. Ez különösen alkalmassá teszi a Colabot például reprodukálható kísérletek, adatfeldolgozási feladatok bemutatására. Kifejezetten támogatja a kollaboratív munkát, hiszen hasonlóan a Google Docs-hoz a notebookok megoszthatók a jogosultságok differenciálásával. Az alkalmazás további lehetőségeiről lásd: <https://research.google.com/colaboratory/faq.html>

hagyva. Mindezeket kontextusként segíti a kapcsolódó levelek szövegének egyszerű megjelenítése, illetve ITIdata rekordjukra mutató link megjelenítése.



5. ábra: Lektorálási felület

Összegzés

A Kántor–Ilia-levelezés digitális forráskiadása több szinten is modellkísérletnek tekinthető. A részletes annotáció, az ITIdata adatbázissal való összekapcsolás, a TEI XML-alapú reprezentáció és a Colab-alapú lektorálási munkafolyamat együttesen egy módszertani javaslatot kínál arra, hogyan válhat az elektronikus levelezés a kortárs hagyatékok tudományosan kiadható, szemantikusan kereshető és adatvédelmi szempontból felelősen kezelhető forrástípusává.

Bibliográfia

- Alföldi István – Szemigán Dóra Hédi – Palkó Gábor – Fellegi Zsófia. „Kutatói e-mail hagyatékok archiválása és feldolgozása.” In *Új technológiákkal, új tartalmakkal a jövő digitális transzformációja felé*, szerk. Tick József – Kokas Károly, 199–207. Budapest: HUNGAR-NET Egyesület, 2023.
- Alföldi István – Szemigán Dóra Hédi – Palkó Gábor – Fellegi Zsófia. *Kántor Lajos e-mail levelezése*. Zenodo, 2023. DOI: [10.5281/zenodo.7825184](https://doi.org/10.5281/zenodo.7825184).
- Bauman Syd – Beshero-Bondar Elisa. „Preserving a LISTSERV Archive in TEI: The Case of TEI-L.” *Journal of the Text Encoding Initiative* 19 (2025). DOI: [10.4000/15pss](https://doi.org/10.4000/15pss).
- Dumont Stefan. „Correspondence Metadata Interchange Format: The Correspondence Metadata of Digital Editions.” *Journal of the Text Encoding Initiative* 10 (2016).

- Indig Balázs – Horváth László – Szemigán Dóra Hédi – Nagy Mariann Eszter. „Emil.RuleZ! – An Exploratory Pilot Study of Handling a Real-Life Longitudinal Email Archive.” In *Proceedings of the 3rd International Conference on Natural Language Processing for Digital Humanities*. Association for Computational Linguistics, 2023.
- Jaillant Lise. „After the Digital Revolution: Working with Emails and Born-Digital Records in Literary and Publishers’ Archives.” *Archives & Manuscripts* 47, 3. sz. (2019): 285–304. DOI: [10.1080/01576895.2019.1640555](https://doi.org/10.1080/01576895.2019.1640555).
- Kalcsó Gyula – Szűcs Kata Ágnes. „Az e-mailek hosszú távú megőrzése.” *Könyvtári Figyelő* 68, 3. sz. (2022): 403–410.
- MNMKK OSZK Digitális Bölcsészeti Központ. „E-mailek archiválása az MNMKK OSZK Digitális Bölcsészeti Központjában.” *dHUpLa – Digitális Bölcsészeti Platform*. <https://dhupla.hu/page/kreativ/email-archivalas>. Hozzáférés: 2026. május 10.
- Pennock Maureen. „Curating E-mails: A Life-Cycle Approach to the Management and Preservation of E-mail Messages.” *DCC Digital Curation Manual*. Digital Curation Centre, 2006.
- Prom Christopher J. *Preserving Email*. 2. kiadás. Digital Preservation Coalition Technology Watch Report. Digital Preservation Coalition, 2019. DOI: [10.7207/twri9-01](https://doi.org/10.7207/twri9-01).
- Schneider J. et al. „Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives.” *Archives & Manuscripts* 47, 3. sz. (2019): 305–326. DOI: [10.1080/01576895.2019.1622138](https://doi.org/10.1080/01576895.2019.1622138).
- Task Force on Technical Approaches for Email Archives. *The Future of Email Archives: A Report from the Task Force on Technical Approaches for Email Archives*. Washington, DC: Council on Library and Information Resources, 2018.
- TEI Correspondence SIG. *Correspondence Metadata Interchange Format / TEI Correspondence SIG Documentation*. é. n.