

Kéziratos források digitális feldolgozása a közgyűjteményi szférában

Közgyűjteményi feladatok kéziratos források esetén

A kéziratos forrásokat őrző közgyűjtemények hagyományos feladatait a digitális technológiák fejlődése és rohamos elterjedése miatt egyre sürgetőbb újragondolni, átformálni, kibővíteni. A hagyományos eljárások, mint a források leírása, metaadatolása, gyűjteményezése már nem látszik elegendőnek, illetve ezek módszere is korszerűsítésre szorul. Az archiválás, hosszú távú megőrzés, illetve a kereshetővé és hozzáférhetővé tétel szintén mást jelent digitális közegben, mint az analóg világban.

A források digitális másolatai létrehozásának általános gyakorlatával az is kérdésessé vált, hogy a közzététel mely szintjét nevezhetjük publikációnak. Ha egy kéziratos forrás digitális másolatát közzétesszük, az már annak számít-e?¹

Mindenesetre ma már alapvető feladatnak tekinthetjük, hogy a közintézményekben őrzött kéziratos források valamilyen módon hozzáférhetőek legyenek anélkül, hogy a fizikai példányt a kezünkbe kellene vennünk. Ez viszont azt jelenti, hogy a szövegkiadás – legalábbis bizonyos szintjei – a közgyűjtemények hatáskörébe kerültek. A GLAM szektor kéziratos forrásokat őrző szereplőinek ezen új feladatkör kihívásaival kell szembesülniük, és azokra válaszokat adniuk.

A kéziratos források közzététele több szinten lehetséges:

1. A kéziratok digitális másolatának közzététele

Mivel az archiváláshoz is szükséges egy jó minőségű digitális fakszimile készítése, nem jelent sok többletmunkát az ún. master fájlból egy közzétételre alkalmas (kisebb felbontású, de jól használható) verzió létrehozása. A művelet automatikus konverzióval megoldható. A szkennelésen túl csak az intézmény online katalógusfelületére való feltöltésre van szükség (és természetesen előzetesen az adott forrás gyűjteményezésére). Így a gyűjteményi rekord adatai, tehát a kéziratok metaadatai kereshetővé, szűrhetővé válnak (az adott katalógusfelület lehetőségei szerint), a felhasználó pedig azonnal megkapja olvasásra a kézirat digitális másolatát.

Ez a szint az analóg világban a hasonmás kiadásnak feleltethető meg, bár az online keresőfelület ehhez számos további funkciót ad hozzá. Jó példa erre a módszerre az Országos Széchényi Könyvtár nemrégiben indított új szolgáltatása, amely Babits-kéziratokat tesz közzé.²

¹ Erre a kérdésre itt nincs lehetőség részletesebben kitérni, a téma külön tanulmány tárgyát képezi.

² <https://copia.oszk.hu/levelek> (2022. január 7.)

2. A kéziratok digitális másolatának és átiratainak közzététele

Az előző szinthez képest ez az eljárás egy lépéssel továbbmegy, nemcsak digitális fakszimilét készít, hanem a szövegek átírását is elvégzi. Ez természetesen idő- és erőforrásigényes munka, amelyre nem biztos, hogy mindenhol van kapacitás. A jövőben viszont a nyomtatott szövegek automatikus gépi felismertetéséhez (OCR) hasonlóan egyre több lehetőség nyílik a kéziratos dokumentumok ugyanilyen szövegfelismertetésére is (l. később), amely jelentősen csökkenteni fogja a digitális szövegek létrehozásához szükséges erőforrásokat.

Az eredmény pedig nagy lehetőségeket rejt magában, hiszen a kéziratok teljes szövegállománya kereshetővé válik, amely a kutatók, felhasználók előtt hihetetlen távlatokat nyit meg, illetve létrejön az alapvető nyersanyag egy digitális forráskiadáshoz. (Ezen a szinten készülhet betűhű, betű szerinti, illetve normalizált átirat is, a projekt céljainak megfelelően.)

Az átirrt szöveg és a digitális fakszimile együttes megjelenítésére célszerű formátum a kétrétegű PDF, ahol a felhasználó az eredeti kézirat fotóját látva kereshet a kép mögött rejtve tárolt szövegátiratokban. Ehhez szükséges a szkennelt kéziratképek ún. szegmentálása is, amelynek során zónákra osztjuk a képet, s a zónák koordinátáinak használatával lesz összekötve a kép az átirrt szöveggel. A kétrétegű PDF-ek szintén feltölthetők az adott intézmény online katalógusfelületére az előző pontban írtak szerint, s itt kell lehetőséget biztosítani az ún. teljes szövegű, tehát a gyűjteményi rekordokhoz kapcsolt PDF-ek tartalmában való keresésre is.³ Ilyen típusú közzétételt elsőként a Digitális Bölcsészeti Központ valósított meg a Petőfi Irodalmi Múzeum (PIM) OPAC-on Kiss József levelezésének feldolgozása során.⁴

3. Digitális forráskiadás

Ezen a szinten hozhatunk létre egy textológiai-filológiai szempontból forráskiadásnak tekinthető publikációt. Az előző pontban született nyersanyagon tovább munkálkodva, vagy azt kihagyva és újonnan létrehozva születhet meg a kézirat annotált, adatgazdagított, részletesen metaadatolt kiadása. Erre a célra alkalmas a TEI XML⁵ formátum, amely a szövegek annotálására nemzetközi szinten is a leginkább elfogadott és elterjedt jelölőnyelv. Az ún. címkék (tagok) és tulajdonságaik (attribútumok) használatával lehetőség nyílik szintaktikai (pl. bekezdés) és szemantikai (pl. név) szövegegységek azonosítására, megjelölésére, ennél fogva azok lekérdezhetővé, kereshetővé és különböző módokon megjeleníthetővé

³ Természetesen lehet átírást készíteni egyéb formátumokban is, pl. sima szöveggént (txt), Wordben, de figyelembe kell vennünk, mik a kitűzött céljaink, szeretnénk-e a digitális fakszimilét összekapcsolni az átirrt szöveggel, illetve kívánunk-e bizonyos alapvető textológiai jelenségeket már ezen a szinten is jelölni, annotálni (pl. törlések).

⁴ <https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/>"Kiss József 1843-1921"

⁵ Text Encoding Initiative: <https://tei-c.org> (2022. január 7.)

válnak. A TEI fejlécben (TEI Header) sokféle metaadatot rögzíthetünk, illetve szerkesztői megjegyzésekkel, külső adattárakra való hivatkozásokkal gazdagíthatjuk a szövegközlést. A TEI XML publikációjára külön platform kialakítása szükséges, amely kiaknázza az annotált, részletesen feldolgozott szövegekben rejlő megjelenítési, keresési és egyéb lehetőségeket. A közgyűjteményi szférában sokak számára megoldást nyújthat egy közösségi platform használata, ezért többek között erre a célra terveztük a Digital Humanities Platform (dHUpLa) felületét (<https://dhupla.hu>).

A digitális forráskiadás textológiai-filológiai munkájának mélysége természetesen különböző mértékű lehet, azt a projekt céljainak megfelelően kell meghatározni, sőt, a feldolgozást további egymásra épülő fázisokra lehet bontani (pl. az alapvető textológiai annotációk elhelyezésétől a tulajdonnevek beazonosításáig). Így már az első fázis elvégzésével publikálhatóvá válnak a szövegek, és a továbbiakban csak frissíteni kell őket.

A forráskiadás alapját képezheti egy későbbi kritikai (tudományos⁶) kiadásnak is, melynek létrehozása már nem feltétlenül a forráskiadást készítő intézmény feladatkörébe tartozik, de adott feltételek mellett természetesen ott is megoldható.

Minden GLAM szektorhoz tartozó intézménynek fel kell mérnie a saját lehetőségeit, kapacitását, és ahhoz mérten választani a fenti lehetőségek közül.

Miért jó a digitális szövegkiadás?

A digitális szövegkiadásoknak ugyanazok az elsődleges előnyei, amelyek a hátrányai is. Rugalmasak, ennél fogva állandóan változnak, így folyamatosan javíthatók is. Esetükben nincs terjedelemhatár, bármeddig bővíthetők, tehát nem szükséges válogatásokat eszközölni – hacsak nem humánerőforrás-hiány miatt.

Előnyük viszont az analóg kiadásokhoz képest, hogy annotálhatók, összeköthetők egyéb tudástárakkal (pl. névterek, bibliográfiák, teauruszok), több formában és folyamatosan publikálhatók, verziózhatók, illetve a legkülönbözőbb műveletek végezhetők bennük (pl. keresés, szűrés, adatvizualizáció). Egészen új módszerekkel válnak kutathatóvá a szöveges tartalmak (l. „distant reading”⁷), illetve az eredeti források archiválására is alkalmasak.

A digitális szövegkiadások ettől függetlenül nem kell, hogy kizárják a nyomtatott szövegkiadások lehetőségét, sőt. A tendenciák inkább azt mutatják, hogy nagyon is tudnak egymás mellett létezni, más-más felhasználói igényeket kielégítve. A jövő afelé halad, hogy a digitális szövegkiadások lehetnek a nyomtatott kiadások alapjai, s nem fordítva.

⁶ A *kritikai kiadás*, illetve *tudományos kiadás* (digital scholarly edition) terminológiai kérdései szintén egy külön tanulmány tárgyát képezik.

⁷ https://en.wikipedia.org/wiki/Distant_reading (2022. január 7.)

Új feladatok, problémák

Milyen új feladatokkal, problémákkal kell megküzdenie egy közgyűjteménynek, ha – valamilyen formában – vállalja a kéziratos forrásai közlését?

Először is össze kell hangolnia a digitalizálást végző műhely, a kéziratoskat őrző intézmény munkáját, illetve a digitális bölcsészeti feladatokat. Minden esetben meg kell találnia a közös nevezőt, amely az összes fél igényeinek megfelel. Ez korántsem egyszerű, már csak a meglévő eszközök, hosszú idők során kialakult hagyományok és protokollok fényében sem.⁸

A közös munka elvégzéséhez elengedhetetlen a közgyűjteményi és a digitális bölcsészeti szempontok egyeztetése. Ha egy közgyűjtemény kéziratos források közzétételére vállalkozik, elkerülhetetlenül beengedi a digitális bölcsészeti feladatokat az intézmény hatáskörébe, és integrálnia kell azokat. Ugyanígy, bizonyos digitális bölcsészeti eszközöket is be kell építenie a meglévő közgyűjteményi infrastruktúrájába, illetve adott esetben visszafelé, egyes digitális bölcsészeti eszközökbe integrálhatja a már meglévő infrastruktúra elemeit.

Kritikus probléma a többletfeladatokhoz szükséges humánerőforrás biztosítása (pl. szövegek átírása). Amennyiben van rá lehetőség, mindenképpen kézenfekvő és célravezető a kéziratoskat őrző gyűjtemény dolgozóinak betanítása és bevonása a digitális bölcsészeti feladatokba. Ezenkívül feltétlenül szükséges az intézmény különböző területein dolgozó informatikusokkal való folyamatos együttműködés is.

Pilot projekt

A fentiekben felvázoltak miatt a Digitális Bölcsészeti Központ⁹ elindított egy pilot projektet, ahol igyekeztünk a felmerülő kérdésekre választ adni, módszereket, protokollokat, workflow-kat kialakítani, megfelelő eszközöket találni, valamint a digitális forráskiadás architektúrájához tesztmodulokat fejleszteni. Az alábbiakban ezeket mutatom be vázlatosan. A pilot projekt Kiss József¹⁰ levelezésének digitális forráskiadása.

⁸ Jó példa levelezéskiadás esetén a digitális objektum egységének meghatározása. Egy kéziratár a gyűjteményezés során általában teljes levelezési irányokat ír le, amely egy személy másik személynek írott összes levelének metaadatait tartalmazza. A szövegfeldolgozás során az alapegységnek viszont a levélnek kell lennie. Össze kell tehát hangolni a szkenneléstől a kiadásig, mikor mi legyen az alapegység, a létrejött digitális objektumokat hogyan kapcsoljuk össze a katalógusban lévő rekordokkal, és hogyan jelenítsük meg azokat.

⁹ A Digitális Bölcsészeti Központ 2021. december 1-jétől az Országos Széchényi Könyvtár részeként működik.

¹⁰ Kiss József (1843–1921), *A Hét* című hetilap alapítója, „olyan szerkesztői tehetség, akit csak az egy Osvát Ernő múlt felül. Lapja rövidesen orgánuma lett mindazoknak, akikben az új polgári szellem irodalmi formát öltött, de hordozta a régebbi nemzedékek kritikusabb elméit is.” – írja Szerb Antal [Szerb Antal: *A magyar irodalom története*. Budapest, Magvető. 1972., 420. p.]. *A Hét* című folyóirat a Nyugat elődjének tekinthető, amelyre az irodalomtörténeti kutatások eddig méltatlanul

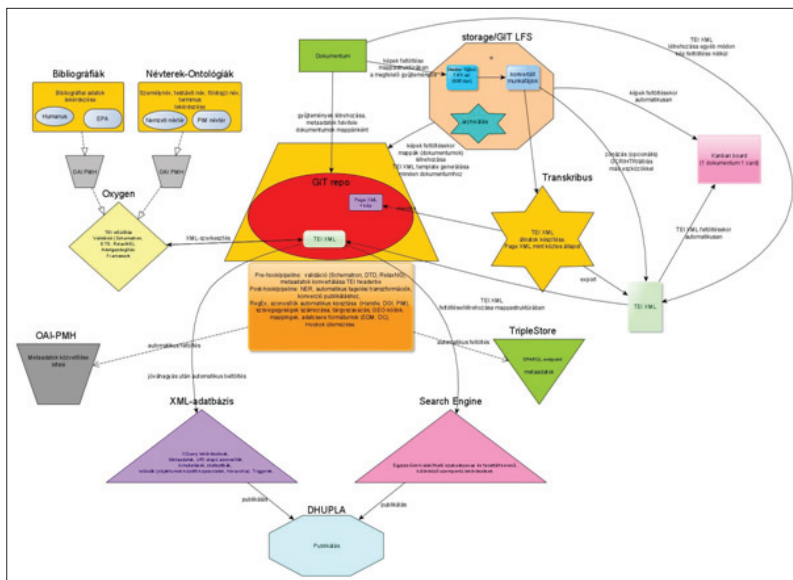
Első lépések

Megtettük a szükséges első lépéseket:

- a szkennelés szabályainak kialakítását;
- a névkonvenció meghatározását (fájlok elnevezése, strukturálása);
- a nyilvántartások, kimutatások elkészítését a projekt forrásairól;
- ütemeztünk;
- kialakítottuk a content management környezetet;
- kiválasztottuk az eszközöket;
- kiosztottuk a szerepeket;
- megterveztük a workflow-t.

Workflow, infrastruktúra

A platformhoz kialakított infrastruktúra a következő ábrán látható:



1. ábra

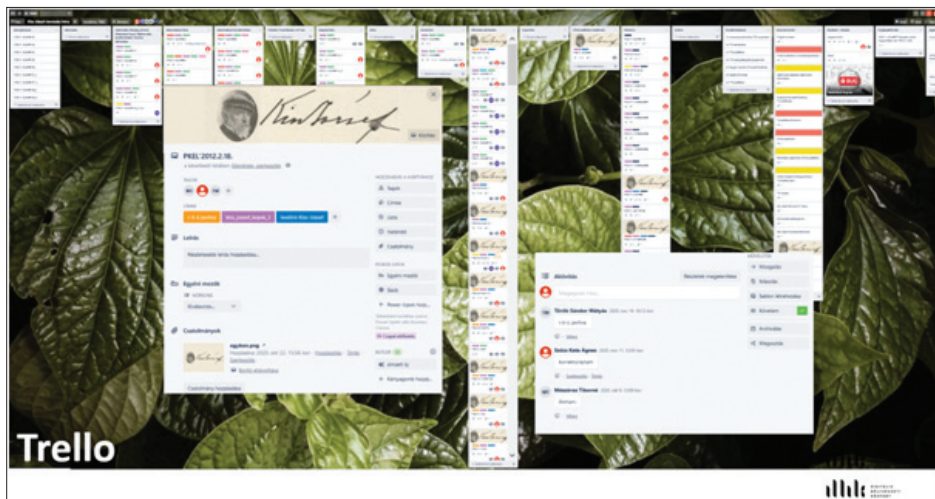
A szkennelés-szövegátírás-publikálás hármasa helyett sokkal összetettebb rendszerre volt szükségünk, ugyanakkor ez mégsem olyan bonyolult, mint amilyennek első ránézésre tűnik. A középpontban egy verziókövető rendszer áll, amelyben a tárolt anyagok minden változása visszakövethető, s innen indul ki minden automatizált folyamat, ezzel áll kapcsolatban közvetlenül vagy közvetve az összes használt eszköz.

kis figyelmet fordítottak. Kiss József szerkesztőségi és egyéb levelezésének közzététele eddig fel-táratlan közeget tesz hozzáférhetővé mind a szakmai közönség, mind az olvasók számára.

A TEI XML-ek publikációja több fázisra osztható a feldolgozás mélységének megfelelően, hogy minél előbb publikussá válhassanak a szövegek. A névterekkel, bibliográfiákkal való összekapcsolás például csak egy későbbi fázisban történik. Beépítettük a rendszerbe a metaadatok automatikus továbbításának lehetőségeit is, illetve a különböző lekérdezési módszerekhez szükséges eszközöket.

Content management

Az egyes dokumentumok munkafolyamatban történő haladását a Trello¹¹ felületén követjük.



2. ábra

Az ún. Kanban-módszer¹² testreszabása eredményeképpen minden oszlop a workflow egy állomásának felel meg, s minden dokumentumnak van egy külön kártyája. A kártya tartalmazza az összes szükséges információt: ki dolgozott már a dokumentummal, jelenleg ki dolgozik rajta, illetve minden egyéb, a projekt szempontjából szükséges metaadatot, amelyek szűrhetők is. Ezenkívül hozzászólások formájában lehet megjegyzéseket írni, kérdéseket megvitatni.

A kéziratok digitális másolatának és átiratainak közzététele kétrétegű PDF formájában

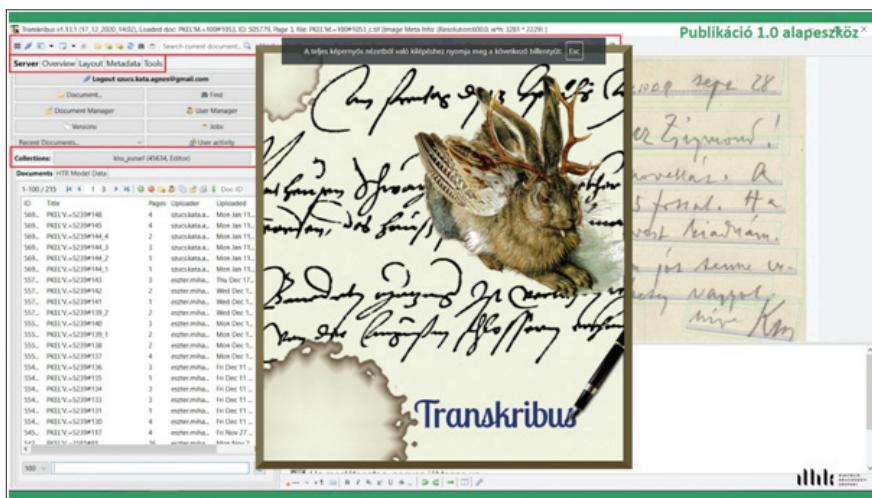
Szerkesztőprogram

A kétrétegű PDF fájlok előállításához a Transkribus¹³ programot használjuk.

¹¹ <https://trello.com> (2022. január 7.)

¹² <https://en.wikipedia.org/wiki/Kanban> (2022. január 7.)

¹³ <https://readcoop.eu/transkribus/?sc=Transkribus> (2022. január 7.)



3. ábra

A programot kifejezetten kéziratos dokumentumok átírására fejlesztették, felhasználóbarát felülettel rendelkezik, s nem mellesleg automatikus kézírás-felismerető, ún. HTR (Handwritten Text Recognition) modellek létrehozására is alkalmas. Az átírt dokumentumok többféle formátumban exportálhatók, például a PDF-be vagy TEI XML-be.

A programmal a saját számítógépen való szerkesztés is lehetséges, de az összes funkció kihasználásához a Transkribus szerverére kell feltölteni a dokumentumokat, amely szintén verziókezelővel biztosítja a tárolásukat.

A felhasználói funkciók teljes körű igénybevétele miatt a projekt során a szerveren dolgozunk, ezért egyik modulfejlesztésként a Transkribus-gyűjtemények automatikus napi mentését tűztük ki célul, amely mára meg is valósult.

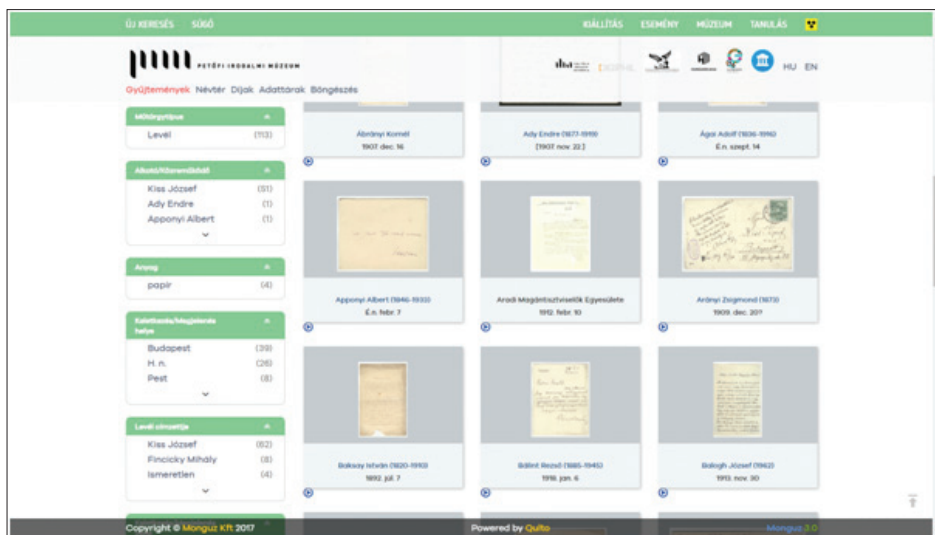
Publikációs, szűrési, keresési, rendezési lehetőségek

A kétrétegű PDF-ek publikálásával viszonylag gyors munkafolyamat valósult meg, a PIM OPAC-jában kapcsolt médiarekordként való közzététel szinte semmilyen fejlesztést nem igényelt. (A PDF-ek megjelenítésével és a találati listákkal kapcsolatban merültek fel kisebb testreszabás-jellegű feladatok.) Bebizonyosodott tehát, hogy ebben a formátumban jól megoldható a szöveggözlés integrációja egy meglévő közgyűjteményi rendszerbe. A PIM OPAC-jában külön URL-en elérhető a teljes Kiss József-levelezés,¹⁴ és a Kiss József-levelezés azon része, amely már PDF formátumban hozzáférhető.¹⁵

¹⁴ [https://resolver.pim.hu/gyujtemeny/levelek/"Kiss József 1843-1921"](https://resolver.pim.hu/gyujtemeny/levelek/) (2022. január 7.)

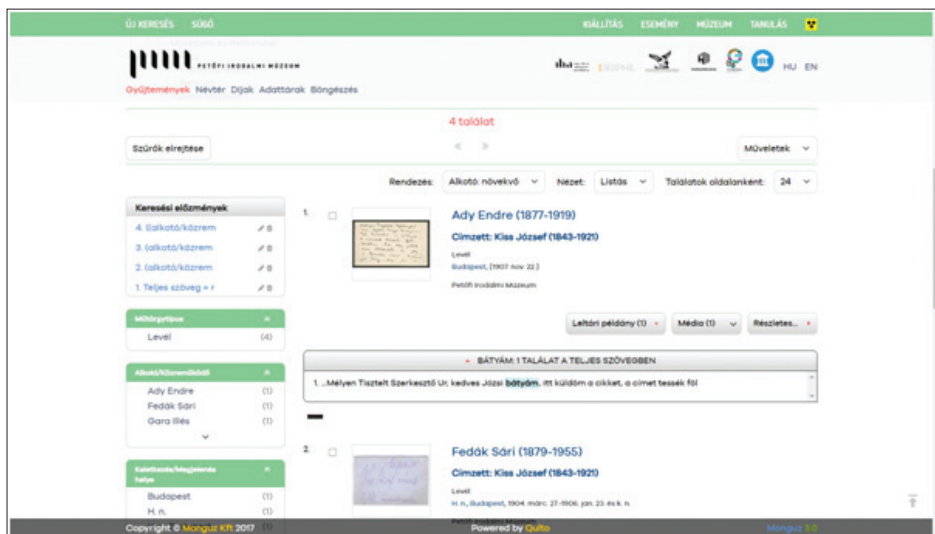
¹⁵ [https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/"Kiss József 1843-1921"](https://resolver.pim.hu/gyujtemeny/levelek/media/csatolt/) (2022. január 7.)

Az online katalógusban eddig is volt lehetőség teljes szövegű keresésre, illetve metaadatok szerinti szűrésre. Az ún. facetek segítségével (l. a következő képen bal oldalt) szűkíthetjük a találatokat az egyes adatmezők szerint, és a találati listánkat különböző szempontok szerint rendezhetjük is.



4. ábra

A teljes szövegű keresés esetén láthatjuk a rekordokon belüli találatokat (l. 5. ábra).



5. ábra

Egy adott rekordra kattintva csak az azon belüli találatokat látjuk (vö. 6. ábra).

ÚJ KERESÉS SÚGÓ

KIÁLLÍTÁS ESEMÉNY MÚZEUM TANULÁS

PETŐFI IRODALMI MÚZEUM

Olyajtemények Névtér Díjak Adattárak Bőngészés

Címzés MARCXML LIDO

https://resolver.pim.hu/bib/PIM/225902

Letöltés

Megnyitás

Levélíró: Kiss József (1843-1920)
Címzett: Bánóczy József (1849-1926)
Tervelem: 1 f.
Megjegyzés: Az eredeti levél ceruzajelű másolata?
Nyelv: Magyar
Dokumentumtípus: Levél

Művelem	Gyűjtemény	Hagyaték	Rövidítés	Leírás
PIM	PIM - Kéziratanyag	Kiss József-hagyaték	V. 5239/1	V. 5239/1

Találatok a teljes szövegben (3 találat):

1. oldal - állíts rabszolgát magad helyett. Kit ajánlanád? **Ölel híved Kiss A Hét**

6. ábra

A PDF megnyitásakor pedig magán a kéziratfotón is ki vannak emelve a találatok (l. 7. ábra).

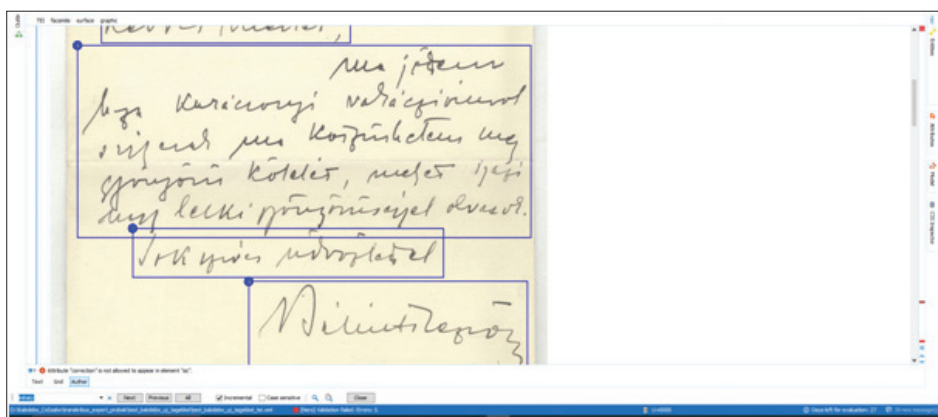


7. ábra

Digitális forráskiadás TEI XML formájában

Szerkesztőprogram

A Transzkribusból exportált TEI XML-ek szerkesztéséhez az Oxygen XML Editor¹⁸ programot használjuk. Ezen belül ki lehet alakítani egy ún. frameworköt, amely az adott projekt(ek) igényeire van szabva, annak minden szempontját figyelembe véve. A szerkesztők ebben a nézetben gombok használatával tudnak annotálni, validálni, különféle automatizált műveleteket elvégezni. A framework kiterjedt textológiai-filológiai jelölésrendszer alkalmazására nyújt lehetőséget anélkül, hogy a szerkesztőnek a technológiai hátteret ismernie kellene. Az adatgazdagítás, a metaadatok részletes kitöltése, a külső adatbázisok bekapcsolása, a legkülönbélebb annotációk alkalmazása és a text-image linking (kép és szöveg összekötése) mind felhasználóbarát módon megoldható ebben a környezetben.

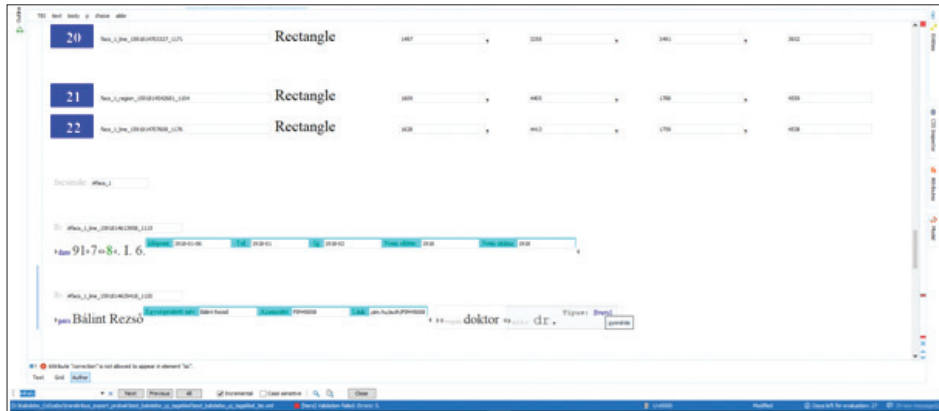


9. ábra

A szerkesztőfelületen a szöveg és a benne lévő annotációk megjelenítését saját igényeink szerint alakíthatjuk ki, akár úgy is, hogy a szerkesztő a majdan publikálható formát (is) láthassa. Meghatározhatjuk a template-fájlokat, amelyek az új dokumentum készítéséhez szükséges alapvető elemeket tartalmazzák, emellett ún. sémafájlokat (DTD, RELAX NG, Schematron stb.) alkalmazhatunk, amelyek a nemzetközi szabvány mellett saját szabályrendszerünk szerint is ellenőrzik a dokumentumokat, beépíthetünk automatikus műveleteket (pl. azonosítókiosztás, metaadatok beemelése, illetve továbbítása).

Az Oxygen rendelkezik ún. Git-klienssel is, amelynek segítségével a szerkesztők közvetlenül a verziókezelő rendszerbe dolgozhatnak, illetve megoldható a szerkesztőségi környezetből való közvetlen publikáció is.

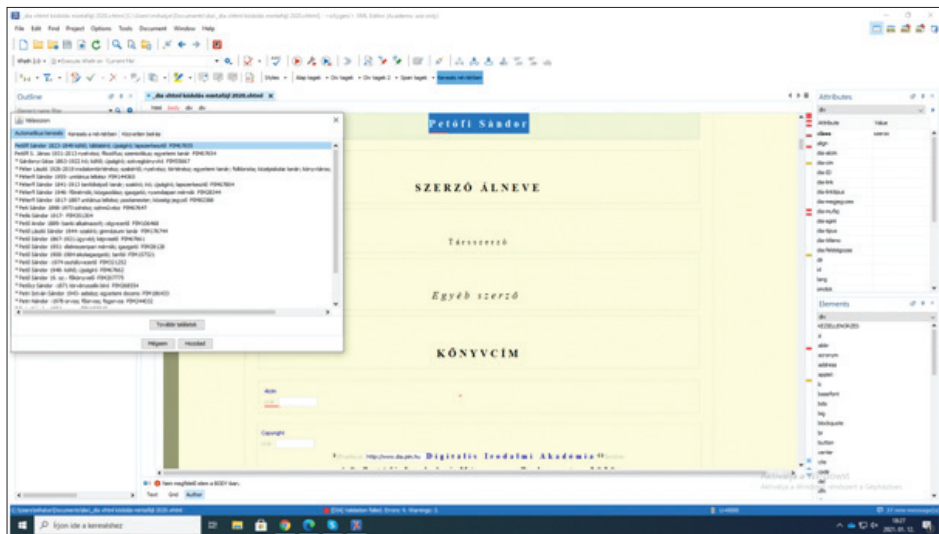
¹⁸ https://www.oxygenxml.com/xml_editor.html (2022. január 7.)



10. ábra

Eszközök integrációja

Ezen a területen is összeköttöttük a közgyűteményi infrastruktúra egyes elemeit a szövegfeldolgozás szerkesztőségi környezetével. Beépítettünk az Oxygen-frameworkbe egy funkciót, melynek segítségével a szerkesztők közvetlenül a PIM Névterében kereshetnek, illetve azon keresztül azonosíthatják a különböző entitásokat. Az Oxygen-frameworkbe integrált névtérfunkció a DIA (Digitális Irodalmi Akadémia¹⁹) frameworkjében valósult meg elsőként (l. 11. ábra).



11. ábra

¹⁹ <https://pim.hu/hu/dia> (2022. január 7.)

A szerkesztő a szövegben szereplő névalakot kijelölve egy kattintással azonosíthatja a nevet a PIM Névterében, illetve egy újabb kattintással beemelheti a nevet azonosító adatokat annotációként. Folyamatban van a funkció implementálása a dHUpa Oxygen-frameworkjeibe, illetve bővítése az egyéb entitás-típusokra és más külső névterekre, adattárakra.

Publikációs, szűrési, keresési, rendezési lehetőségek

A TEI XML megjelenítési lehetőségei korlátlanok. A pilot projekt során arra koncentráltunk, hogy a felhasználók, kutatók számára leginkább hasznos vizualizációs módokat hozzuk létre. Ezért fejlesztettünk egy új eszközt a digitális fakszimile és az átírt szöveg együttes megjelenítésére, a text-image linking szövegfeldolgozó módszert elsőként alkalmazva Magyarországon. Ennek lényege, hogy a kép zónái össze vannak kapcsolva az átírt szöveggel, tehát a felhasználó együttesen látja a kettőt.

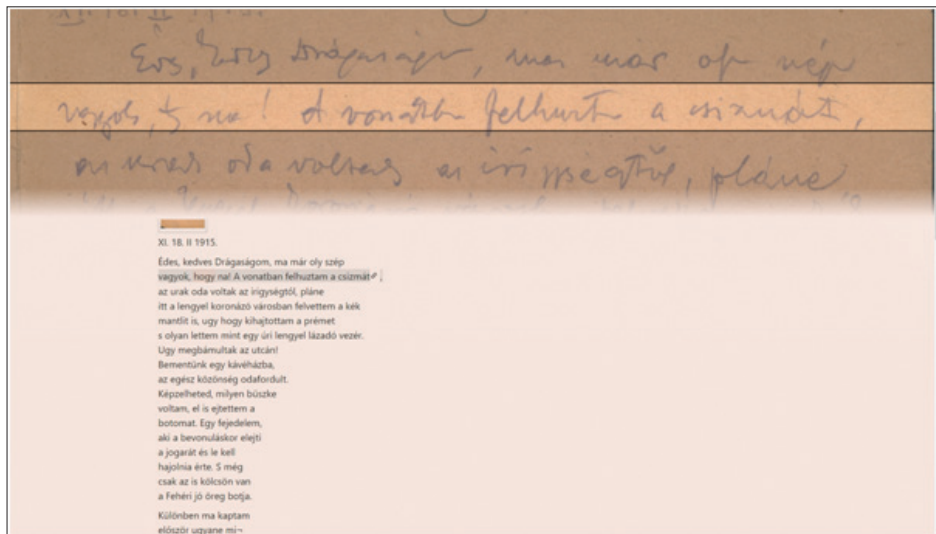
Háromféle nézetet is kialakítottunk:

1. A kurzorral az átírt szöveg adott sora fölé menve megjelenik a fakszimile megfelelő sora:



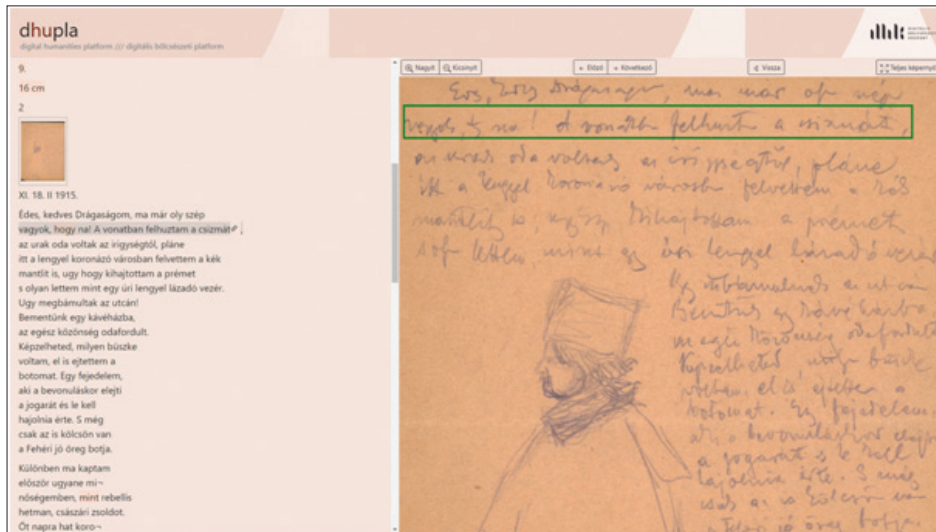
12. ábra

2. A kurzorral az átírt szöveg adott sora fölé menve láthatjuk a fakszimile megfelelő részletét kisebb kontextusban:



13. ábra

3. Osztott képernyőn láthatjuk a fakszimilét és az átírt szöveget (mindkettőn az éppen olvasott szövegrész van kijelölve):



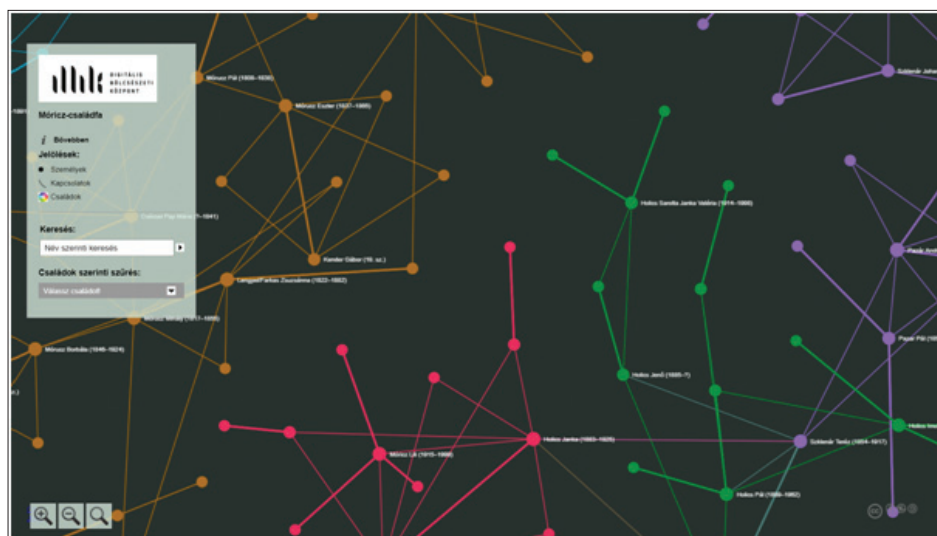
14. ábra

A szövegben szereplő annotációk kurzorral a kiemelt szövegrészek fölé menve jelennek meg:

(Natural Language Processing²⁰) eszközökkel. A modul beépítésével lehetséges lesz a szövegekben szereplő nevek automatikus megjelölése, még hozzá névtípusonként. Ezután következhet a fentiekben vázolt azonosítási procedúra. Amennyiben nincs kapacitás a nevek manuális azonosítására, a technológia mindenképpen alkalmazható a publikációs felület keresőjébe beépítve, így a felhasználó a program által névként felismert szóalakok között tud böngészni.

Adatvizualizációs lehetőségek

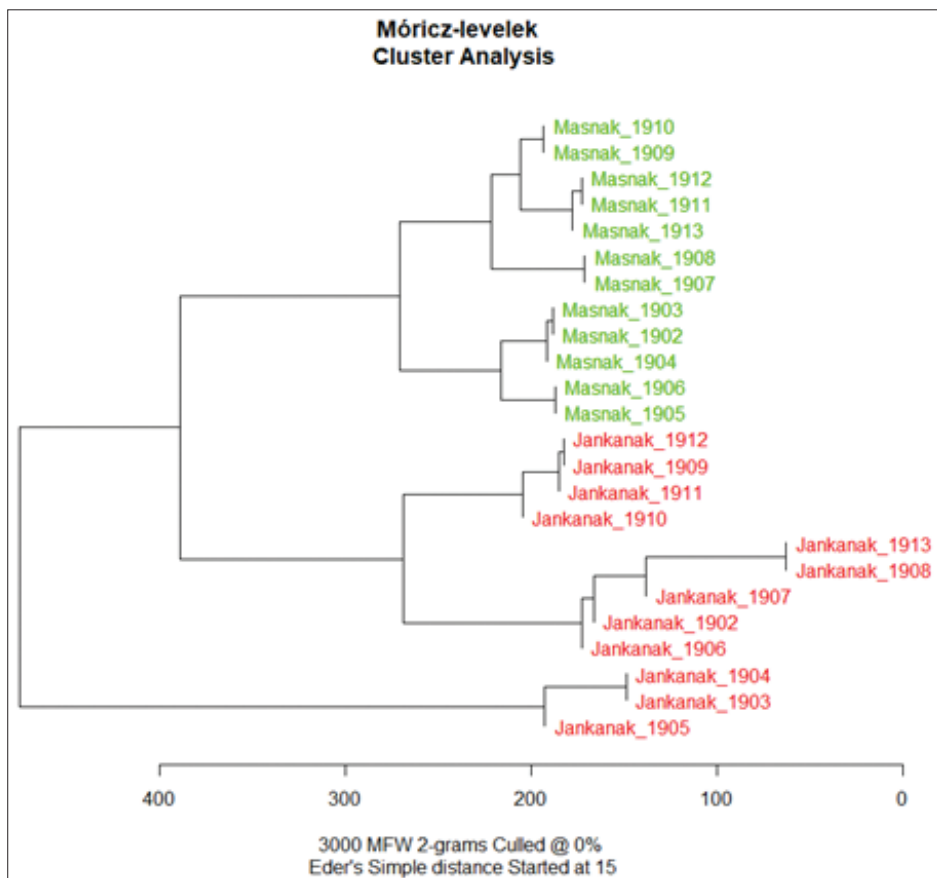
A metaadatok, illetve a szövegekben elhelyezett annotációk felhasználásával a legkülönbözőbb statisztikák, vizualizációk állíthatók elő az egyes szövegtörzsekről. Ilyen például a Móricz Zsigmond levelezésének adataiból és új kutatási eredményekből előállított Móricz-családfa (l. 17. ábra).



17. ábra

Izgalmas lehetőség az egyes szövegtörzsek vagy azok alegységeinek stilometriai vizsgálata is. Az alábbi ábrán Móricz Zsigmond feleségének, Holics Jankának, illetve másoknak írt levelei stílusának vizsgálata látható évenkénti bontásban. Jól láthatóan elkülönül a két csoport, de a Holics Jankának írt levelekben belül is a házasságkötésük (1905) előtti és utáni szakasz is (l. 18. ábra).

²⁰ https://en.wikipedia.org/wiki/Natural_language_processing (2022. január 7.)



18. ábra

A vizualizációs lehetőségek száma végtelen,²¹ ezeknek csak a képzeletünk szabhat határokat. Fontos azonban mindig szem előtt tartani, hogy a vizualizációk az adatok olyan aspektusait világítsák meg, amelyek enélkül nem, vagy nagyon nehezen lennének értelmezhetők, hiszen a végső cél minden esetben a felhasználók számára hasznos szolgáltatások nyújtása.

A Digitális Bölcsészeti Központ 2021. december 1-jétől az Országos Széchényi Könyvtár szervezeti egységként működik tovább.

²¹ L. <https://dhupla.hu/page/kreativ> (2022. január 7.)

Rezümé

A kéziratos forrásokat őrző közgyűjtemények hagyományos feladatait a digitális technológiák fejlődése és rohamos elterjedése miatt egyre sürgetőbb újragondolni, átformálni, kibővíteni. A hagyományos eljárások, mint a források leírása, metaadatolása, gyűjteményezése már nem látszik elegendőnek, illetve ezek módszertana is korszerűsítésre szorul. Az archiválás, hosszú távú megőrzés, illetve a kereshetővé és hozzáférhetővé tétel szintén mást jelent digitális közegben, mint az analóg világban.

Mindenesetre ma már alapvető feladatnak tekinthetjük, hogy a közintézményekben őrzött kéziratos források valamilyen módon hozzáférhetőek legyenek anélkül, hogy a fizikai példányt a kezünkbe kéne vennünk. Ez viszont azt jelenti, hogy a szövegkiadás legalábbis bizonyos szintjei a közgyűjtemények hatáskörébe kerültek. A GLAM szektor kéziratos forrásokat őrző szereplőinek ezen új feladatkör kihívásaival kell szembesülniük és válaszokat adniuk. A tanulmány a Digitális Bölcsészeti Központ (PIM) e területen végzett fejlesztéseit, eredményeit mutatja be.

Digital processing of manuscript resources in the public collections sector

Due to the development and rapid spread of digital technologies, it is increasingly urgent to rethink, reshape, and expand the traditional tasks of public collections that preserve manuscript sources. Traditional procedures, such as the metadata description and collecting sources, no longer seem to be sufficient, and their methodology also needs to be modernized. Archiving, long-term preservation, and making sources searchable and accessible also mean something different in the digital environment than in the analog world.

In any case, we can now consider it a fundamental task that the manuscript sources kept in public institutions are accessible in some way without having to get hold of the physical copy. This, in turn, means that at least certain levels of text publishing have come under the purview of public collections. The actors who preserve manuscript sources in the GLAM sector must face the challenges of this new task and provide answers. The study presents the developments and results of the Digital Humanities Center (PIM) in this area.

MIHÁLY ESZTER
csoportvezető

Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ