
Szerényi Ildikó

A csendőrkarton-projekt

Bevezetés

A Magyar Nemzeti Levéltár (MNL) 2021-ben publikálta az 1828. évi országos összeírás adatbázisát, amelyet mesterséges intelligencia¹ alkalmazásával és önkéntesek bevonásával valósított meg a European Digital Treasures nemzetközi projekt² keretein belül. Az adatbázis közzétételén túlmenően sikertörténetnek bizonyult az a társadalmi összefogás, amely lehetővé tette az iratanyag feldolgozását és a névre kereshető adatbázis publikálását az Adatbázisok Online felületén.³ Ez a pozitív tapasztalat ösztönözte a levéltárat arra, hogy folytassa a közösségi segédletkészítőkkel a közös munkát, és egy újabb, nagy érdeklődésre számottartó adatbázis létrehozását tűzze ki célul. A második önkéntes projektbe partnerként a Magyar Családtörténet-kutató Egyesület (MACSE) is bekapcsolódott, elmélyítve ezáltal a levéltár és az egyesület között hosszú évek óta fennálló eredményes együttműködést.

A felhasznált iratanyag

A feldolgozandó iratanyag kiválasztása során az MNL szempontrendszerének élén ismét a levéltári irategyüttes társadalmi hasznosítása, a családtörténet-kutatással foglalkozó állampolgárok igényeinek kiszolgálása játszott a legfontosabb szerepet, hiszen a társadalmi összefogás ennek a célnak a mentén biztosítható leginkább. Mindazonáltal az iratanyag tudományos szerepének hangsúlyozása is kiemelt célként fogalmazódott meg.

Az újabb közösségi program keretében feldolgozott iratanyagot, a *Magyar Királyi Belügyminisztérium államrendészeti kartotékrendszerét* az MNL jogelőd intézménye, a Magyar Országos Levéltár digitális másolatban vásárolta meg 2011-ben az orosz társintézménytől. A magyar belügyminisztérium által keletkeztetett 40 doboznyi eredeti, papíralapú irat ugyanis ma az Oroszországi Föderáció Katonai Levéltárának (Российский государственный военный архив – RGVA) őrizetében található, előtte pedig a szovjet titkosszolgálat, a KGB (Комитет государственной безопасности) birtokában volt. 1945-ben a Magyarországot megszálló szovjet haderő számos más, a Horthy-korszak magyar állami szervei által keletkeztetett irattal együtt ezt az irategyüttest is magával vitte a Szovjetunióba. Az 1920 és 1944 között a magyar rendőrség és csendőrség működése során keletkezett, zsákmány-anyagként kezelt teljes kartotékrendszer Budapest ostro-

¹ Hegedűs, 2020: 57–64.; Hegedűs, 2021: 90.


² Lásd erről bővebben: Szabó, 2022: 57–72.

³ Lásd erről bővebben: Szerényi, 2022a: 73–86.; Szerényi, 2022b: 57–64.

0.	1.	2.	3.	4.	5.	6.	7.	8.	9.
Név: <i>Kiss Mihály</i>									
Áll.év:									
Apa:									
Anyja:									
Házastársa: <i>Szalai Mária</i>									
Születési év: <i>1877</i> állam- juni hó <i>23</i> nap. n.év.									
Születési hely: <i>Köny</i>									
Államnév lez hely:									
Vallás: <i>rk.</i> Csal. áll. <i>nő</i>									
Foglalkozás: <i>kisbirtokos</i> Műveltség:									
Különös ismertető jelei:									
Terméte:									
Arca:									
Haja:									
Szeme:									
Nyelv:									
Széjség:									
Függelék:									
Székfoglalás:									
Fajta:									

Nyilvántartásba vétel okának részletes megjelölése: *Köny alatt Köny községben a paraszti lakás elölke volt.*
Köny névvel Családoktatásban kéresem kerélt rendőri felügyelet alá helyezés.
Vizsgálataim folyamatos. Ha levelem lett.

Magyar Országos Levéltár
K 150 - VII - 9/b
14/346/2011



Kiss Mihály (szül. 1877. június 23.) kisbirtokos államrendészeti nyilvántartólapja (HU-MNL-OL-K 150-VII-9/b)

mát követően került a szovjet csapatok birtokába. A 69 208 darab fájlból álló digitalizált irategyüttest a magyar levéltárosok utólag illesztették be a Belügyminisztérium levéltárának rendszerébe, provenienciája alapján pedig a HU-MNL-OL-K 150-VII-9/b jelzetet kapta. Tartalmát tekintve a titkosszolgálati tematikájú állomány a két világháború közötti magyar politikai rendészeti szervek nyilvántartása az államra veszélyesnek ítélt, elsősorban baloldali mozgalmakhoz kötődő személyekről. Ez a rendőrségi háttéranyag körülbelül 60 000 személy, túlnyomó többségében magyar nyelvű, változó részletességgel kitöltött nyilvántartólapját, illetve adatait foglalja magában. A személyi kártyákhoz mintegy 9000 kiegészítő irat is tartozik. A kézzel vagy írógéppel kitöltött formanyomtatványok a megfigyelt személy adatait és gyakran az ügyben tett intézkedésekre vonatkozó megjegyzéseket is tartalmazzák. A személyekre vonatkozó adattartalom gazdagsága kiemelkedő: név, születési név, apa és anyja neve, pontos születési idő vagy életkorra vonatkozó adat, amelyből a születési év kiszámítható, születési hely és lakóhely, vallás, foglalkozás, személyleírás, különös ismertetőjel, a nyilvántartásba vétel okának részletes, akár több bekezdésnyi leírása. Érdekes, hogy elenyésző mennyiségben ugyan, de találhatunk ujjlenyomatot vagy fényképet is egyes megfigyelt személyekről.

A kártyákon gyakran látható egy „Zsebkönyvben feldolgozva” szöveggel ellátott pecsét, amely az Államrendészeti Zsebkönyvre, a nyilvántartás kisebb adattartalommal rendelkező, nyomtatott változatára utal. A belügyminiszter 4400/1931./VII.res. számú rendelete 1931. július 31-én előírta a központi államrendészeti nyilvántartás megszervezését és az Államrendészeti Zsebkönyv kiadását. A zsebkönyv, az úgynevezett „fekete könyv”

az államrendészeti nyilvántartás kivonatolt változata, 1931 és 1944 között a belügyminisztérium időszaki kiadványa volt. A bizalmas iratnak minősülő kiadványba a politikai nyomozók is csak a parancsnokuk engedélyével tekinthettek bele. A dokumentum eleve kis példányszámban jelent meg, a II. világháború nagy iratpusztulásának köszönhetően nagyon kevés fennmaradt példányról tudunk, az Állambiztonsági Szolgálatok Történeti Levéltára (ÁBTL) büszkélkedhet néhány darabbal.⁴ A zsebkönyv egy példányát Cseh Gergő Bendegúz és Papp István, az ÁBTL főigazgatója és főosztályvezetője bocsátotta rendelkezésünkre a projekthez. A példányt az MNL digitalizáló műhelyében digitalizáltuk, és digitális másolatban az adatbázis mellékleteként, kétrétegű PDF-formátumban közzétettük. A dokumentumot OCR technikával tettük szövegesen kereshetővé és az adatbázis információs oldaláról tölthető le.

A projekt lebonyolítása

A nagy terjedelmű iratanyag feldolgozásának megvalósítását a levéltár a közösségi segédletkészítő csoport ismételt bevonásával tervezte, így ezúttal nem került sor nyilvános toborzásra. A felhívást és az iratanyag mintaoldalait a korábbi projektben részt vevő önkéntesek kapták meg, és egy héten belül a korábbi csapat szinte teljes egészében jelentkezett az újabb feladatra. A MACSE részéről nyolc, sokéves anyakönyvi indexelési tapasztalattal rendelkező családfakutató csatlakozott a már korábban összeszokott társasághoz, így összességében 70 fő kezdte meg a feldolgozó munkát 2023. október 1-én.

A projektindító rendezvény egyúttal az előző önkéntes projekt lezárását, azaz az 1828. évi összeírás adatbázisa ünnepélyes átadásának szerepét is betöltötte. Németh István osztályvezető tudományos szempontból ismertette az új, feldolgozandó iratanyagot, majd Szerényi Ildikó projektvezető tájékoztatta a közösségi segédletkészítőket az előttük álló feladat részleteiről. Azok az érdeklődők, akiknek nem volt lehetőségük személyesen megjelenni a rendezvényen, Zoom-rendszeren keresztül élőben követhették az eseményt, amelyet később felvételről is elérhetővé tett a levéltár.

Az önkéntesek előtt álló feladat kilenc alapadat rögzítése volt főkartononként, ez mintegy 450-500 ezer adat átírását jelentette összességében. A levéltár az adatbázisban a digitalizált iratok képét is közzéteszi, így minden további adat a dokumentumról leolvasható, illetve a rögzített adat ellenőrizhető. Az MNL a közösségi levéltári adatrögzítő munka során olyan személyeket foglalkoztatott, akik érdeklődtek a magyar történelem és a családtörténet-kutatás iránt, rendelkeztek bizonyos fokú paleográfiai ismeretekkel, jártasak voltak a kézírással készült iratok, valamint a német, zsidó és szláv család- és keresztnevek kiolvasásában, illetve az átlagosnál valamivel magasabb szintű földrajzi ismeretekkel rendelkeztek, ami segítette őket a településnevek beazonosításában. A jelentkezők mintaoldalal segítségével tesztelheték tudásukat és paleográfiai ismereteiket, és a projekt megkezdése előtt megismerkedhettek az adatbeviteli szoftver kezelőfelületével is.

⁴ Kántás Balázs: Magyar Királyi Belügyminisztérium államrendészeti kartotékkendzszere – Csendőrkartonok információs oldala – Adatbázisok Online

A rögzítendő adatok többnyire az előre nyomtatott kartonok bal felső sarkában szerepeltek (lásd a 16. oldali kartonon a bekeretezett részt).

A telepítést nem igénylő, online hozzáférhető Oracle alapú adatbázis adatbeviteli felületét a Helion Kft. fejlesztette az MNL megrendelésére. Az adatbázisban egymás mellett helyezték el az adatmezőket és a digitalizált képállományt. A felhasználóbarát felületre egyedi jelszóval lehetett belépni.

A kilenc rögzítendő személyi alapadat mezői túlnyomórészt szabad szöveges mezőként jelentek meg az adatbázisban. A *Vallás* mezőt értékkészletből történő választással lehetett kitölteni. A *Születési hely* és *Lakhely* mezők bekapcsolása az MNL Földrajzi Névtérbe érdekes újdonság volt a feldolgozó munkát végzők számára. Ha a települést tartalmazó mezőkbe, jelen esetben a *Születési hely* vagy *Lakhely* adatmezőbe beírunk egy településnevet, akkor a Névtér szóra kattintva egy listát kapunk, ahonnan kiválaszthatjuk az adott településnévhez kapcsolódó földrajzi helyek közül az adott mező tartalmához legjobban illeszkedőt. Ezzel a módszerrel egyértelműen azonosítjuk egy térképen a települést. E funkció a településnévre vonatkozó keresések lefuttatásakor bizonyul hasznosnak.

Csendőrkarton adatai	
<p>Adja meg a szükséges adatokat, majd kattintson a M... folyamatok hajtódnak végre: Feldolgozva + Mentés + mező: Enter Előző szöveges mező: Shift+Enter</p>	
Azonosító	30786010
Név	Gubis András <small>Kartonon szereplő név</small>
Születési neve	
Apja neve	Gubis János
Anyja neve	Kohut Zsuzsanna
Házastárs neve	Séfer Julianna
Születési idő	1898 . 8 . 27 <small>Amennyiben nem ismert a pontos dátum, a hónapot is meg kell adni</small>
Születési hely	Vésztő 67843 Névtér
Lakhely	Mezőberény 40479 Névtér
Vallás	evangélikus (lutheránus)
Foglalkozás	napszámos

0	1	2	3	4	5	6	7	8	9
Név: <i>Gubis András</i>									
Álnév:									
Apa: <i>Gubis János</i>									
Anyja: <i>Kohut Zsuzsanna</i>									
Házastársának neve: <i>Séfer Julianna</i>									
Szül. <small>év</small> <i>1898</i> <small>hó</small> <i>aug.</i> <small>nap</small> <i>27.</i> Állampolg.:									
Születési hely: <i>Vésztő</i>									
Állandó lakhely: <i>Mezőberény</i>									
Vallás: <i>égy</i>					Családi áll.: <i>nő</i>				
Foglalkozás: <i>napsz.</i>					Ujinyomat:				
Termete:					Fénykép:				
Különös ismertető jelei:									

Gubis András (szül. 1898. augusztus 27.) adataival kitöltött adatlap, amely a földrajzi névtérbe kapcsolást is megjeleníti.

A feldolgozás során alap statisztikai tájékoztatást a munkafolyamat állásáról a Listanézet nyújtott. A (narancssárga színnel jelölt) információs sorban vált láthatóvá a következő, adataiban folyamatosan frissülő szöveg: Feldolgozva 12 904 karton (18,6%), feldolgozásra váró: 56 304 karton (81,3%). A közösségi segédletkészítőknek az adatbázis véletlenszerűen kínálta fel a kartonokat. Három lehetőség állt a feldolgozó munkát végzők

előtt a karton megjelenésekor: megoldják a kartont (zöld), nem dolgozzák fel, „passzolják” (fekete) vagy részben átírják és félkésznek minősítik (kék színkód). A fekete és kék színkódos kartonok később ismét megjelentek valamelyik másik feldolgozó számára, egészen addig, amíg valaki nem véglegesítette a megoldást, azaz zöld színkódot nem adott.

Az egyéni statisztikai adatok csak a levéltárosok számára voltak elérhetőek az adminisztrátori felületen. Az egyéni statisztikát körlevélben küldtük ki heti rendszerességgel. A közösségi feldolgozók nevét személyiségi okokból stílusosan titkosszolgálati fedőnevekkel helyettesítettük, így minden résztvevő csak a saját magára vonatkozó adatokat ismerhette meg.

FEDŐNÉV	MEGTEKINTETT KARTONOK	FELDOLGOZOTT KARTONOK
Angyalföld	3229	2015
Arad	172	156
Baktalórántháza	5432	3800
Balatonfenyves	905	613
Bánság	826	328
Békásmegyer	2157	1491
Berettyó	80	65
Berettyóújfalú	1302	606
Biharkeresztes	109	46
Budafok	520	429
Burgenland	1205	940
Debrecen	137	87
Diósgyőr	421	355
Dunaföldvár	4878	3082
Eger	641	511
Egerszeg	1369	1026

A feldolgozó munka megkezdésekor az indexelési feladat elvégzését megközelítőleg öt hónap időtartamra becsültük, a valóságban azonban ennél lényegesen hamarabb befejeződött az adatok rögzítése; a projekt ezen szakasza 2022. október 1-jén indult és november 15-én nagy meglepetésünkre már véget is ért. Az átírási szakaszban Kántás Balázs, az MNL főlevéltárosa tudományos-ismeretterjesztő Facebook-posztokkal segítette elő a korszak mélyebb megértését, betekintést nyújtva a politikai nyomozók munkájába, a két világháború között működő titkosszolgálatok tevékenységébe.

Az előző projekthez hasonlóan a program zárásaként díjakat sorsoltunk ki a legalább 300 kartont feldolgozó segédletkészítők között. Az 1. díj egy 3 hónapos Ancestry előfizetés, a 2. díj egy lakoma két fő részére a Sir Lancelot középkori étteremben, a 3. díj pedig egy 1 éves online Rubicon előfizetés volt. Ezen túlmenően az első három legszorgalmasabb feldolgozónak színházjegyeket nyújtottunk át különdíjként. A program zárásaként minden résztvevő egy, az MNL grafikusától tervezett elismerő oklevelet kapott. Az emléklap háttérben a két szembenálló politikai erőt reprezentáló szimbólumok: egy csendőrkalap és egy Lenin-sapka kontúrja látható.



Utómunkálatok

Az átírási szakasz lezárulásával megkezdődhetnek az adatbázis utómunkálatai, amelyek mind az adattartalmat, mind pedig annak struktúráját nagy mértékben érintették. A szovjet titkosszolgálat minden kartonról orosz nyelvű fordítást készített (amelyeket az iratok mellett helyezett el), hogy a lehető legpontosabban feltérképezze a kommunista érzelmű magyar lakosságot. Az orosz társintézmény az eredeti, magyar nyelvű kartonok digitalizálására kapott megbízást, véletlen folytán azonban mintegy 300, oroszra fordított irat digitalizálása is megtörtént, amelyeket az adatbázisban szintén közzéteszünk. Az oroszra átültetett, cirill betűs kartonok transliterálását a Hajdú-Bihar Vármegyei Levéltár főlevéltárosai, Szikla Gergő és Husvéth András végezték.

Az adattartalom ellenőrzése, esetleges javítása manuális és informatikai módszerek felhasználásával történt. A listanézetből kiszűrhető hibák javításában 5-6 közösségi segédletkészítő is részt vett. A listanézet oszlopainak rendezése során a hasonló adatok egymás alá kerültek, ami megkönnyítette a rosszul rögzített információk, nevek, egyéb hibák felderítését. Összességében elmondható, hogy az önkéntesek remek munkát végeztek, rendkívül kevés listanézetből kiszűrhető hibára derült fény.

A munkafolyamat következő szakaszában következtek az informatikai SQL-lekérdezések, amelyek már lényegesen nagyobb terjedelmű megoldásra váró listákat eredményeztek. Nyolcezer esetben maradt el valamilyen oknál fogva a névtérhez kötés, ezek egy részét automatizáltan pótolni lehetett, ám akadtak manuális feladatok is ezen a téren, pl. települések azonosítása. A közösségi segédletkészítők alaposan és pontosan dolgoztak, számos esetben rögzítettek adatot a kilenc alapadaton túlmenően: 18 000 esetben töltötték ki a *Megjegyzés* mezőt. Ennek az egyébként örvendetes szorgalmi munkának köszönhetően azonban e mező túlszűfoltta, áttekinthetetlenné vált. Megoldásként a *Megjegyzés* mező adattartalmának strukturálása, szétbontása kínálkozott. A prefix (előtag) az adatbázisokban a mezők azonosítására szolgáló rövid név vagy betűcsoport, amelyet kettőspont követ, utána következik maga az adat, pl. Állampolgárság: francia. Jelen esetben az adatok szakaszolásához 27 különböző előtagra volt szükség. Néhány a létrehozott prefixek közül: Adatpontosítás, Állampolgárság, Beosztási hely, Családi állapot, Életkor, Irat nyelve, Különös ismertetőjel, Külső hivatkozás, Lakcím, Munkahely, Ügyleírás. A prefixek használata lehetővé tette, hogy ezek az információk külön sorban jelenjenek meg az adatbázisban és szűrni is lehessen ezekre az adatokra. Mivel a *Megjegyzés* mező használata minden önkéntes számára egyedi döntésen alapuló rögzítési lehetőséget biztosított, az ily módon felvitt adatok esetlegesek, nem terjednek ki minden kartonra, bizonyos extra információk azonban az alapadatokon túl is rendelkezésre állnak egyes kartonoknál.

Az utólagos adattisztítás, rendszerezés, olvasószerkesztés időigényes feladatot jelentett a Digitális Szolgáltatás-fejlesztési Osztály munkatársainak. Ezt a helyzetet a jövőben egy minden részletre kiterjedő űrlap létrehozásával tartjuk elkerülhetőnek.

A feldolgozó munka során vált ismertté az iratanyagban szereplő, mintegy 550 többé-kevésbé ismert prominens személy kartonja. Prominens személynek azokat tekintettük, akik önálló magyar vagy idegen nyelvű Wikipédia-oldallal rendelkeznek vagy szerepelnek a Petőfi Irodalmi Múzeum személynévtérben. Később ismertté vált kommunista politikusokra, írókra, művészekre, orvosokra és egyéb értelmiségiekre gondolkunk itt, például Apró Antal, Kádár János, Déry Tibor, Weöres Sándor, Klaus Mann, John Dos Passos vagy épp Hanuszik Elemér, a genealógus. Foglalkozás alapján történő szűréssel kevésbé ismert személyeket is sikerült felfedezni az adatbázisban. A prominens személyeket utólag kötöttük be az MNL személynévtérébe; mintegy 200 esetben új személynévterek rekord létrehozása is szükségesnek bizonyult. Nagy élményt jelentett mind a feldolgozók, mind a levéltárosok számára egy-egy méltán híressé vált személy felbukkanása a kartonok között, ami szintén növelte a projekt vonzerejét, érdekességét.

Köblös Artúr a szolnoki kir. ügyészség 13679/1920 k.ú. számú vádiratával vádolva van a btk. 278 §-sába útk. 1. rendbeli gyilkosság büntette címén, melyet azáltal követett el mint vöröstengerész századparancsnok 1919. jun. 3-án Mezőtúron Szabó Istvánt halálra ítélte és az ítéletet a terroristaival végre is hajtatta. Nevezett a kommun bukása után ismeretlen helyre szökött amiért a szolnoki kir. ügyészség 13679/1920 k.ú. szám alatt a vádhatár kézre kerítése végett gyilkosság bünt. címén nyomozólevelet bocsátott ki, mely nyomozólevél hatályát veszíti 1940. jun. 7-én.

Köblös Artúrra vonatkozólag bővebb személyi adat sem az ügyészségi sem a B. 4645/1921 számú iratoknál nincsenek sőt a személyleírása sem lett annak idején beszerezve.

Maglód 1933 dec. 29.
Öklény
 r. fogalmazó.

0 1 2 3 4 5 6 7 8 9										Nyilvántartásba vétel okának részletes megjelölése:
Név: <i>Déri Tibor</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Álnév:										Nemzetiségi nyilatkozat: <i>Magyar</i>
Apa: <i>Nápoly</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Anyja: <i>Rosenberg Erzsébet</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Házastársának neve:										Nemzetiségi nyilatkozat: <i>Magyar</i>
Szül. hely: <i>Budaörs</i>		Szül. idő: <i>1894. X. 18</i>			Állampolg.: <i>Magyar</i>					Nemzetiségi nyilatkozat: <i>Magyar</i>
Állandó lakhely: <i>---</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Vallás: <i>izl.</i>		Családi áll.: <i>nős</i>			Ujnyomár: <i>---</i>					Nemzetiségi nyilatkozat: <i>Magyar</i>
Foglalkozás: <i>író, költő</i>		Fénykép: <i>---</i>			Különös ismertető jelei: <i>---</i>					Nemzetiségi nyilatkozat: <i>Magyar</i>
Termete: <i>körép</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Arca: <i>karcsú</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Haja: <i>fekete</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Szeme: <i>karcsú</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Orra: <i>puha</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Szája: <i>---</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Fogai: <i>---</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Szakálla: <i>---</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>
Bajusza: <i>---</i>										Nemzetiségi nyilatkozat: <i>Magyar</i>

Déry Tibor Kossuth-díjas író, költő államrendészeti kartonja (HU-MNL-OL-K-150-VII-9/b)

Az adatbázis szerkezetének, felépítésének kialakítása jelentette a legnagyobb kihívást a projekt során. Kiinduló állapotként 69 208 digitális képfájl állt rendelkezésünkre jpg formátumban. Az orosz társintézmény a digitalizálás során nem alkalmazta a fájlnevekben a recto-verso megjelölést, emellett az egyéb módon az egy adott személyhez szorosan kapcsolódó oldalak összerendelését is meg kellett oldanunk, hiszen a kezdetektől nyilvánvaló volt számunkra, hogy személynéven alapuló adatbázist építünk. A fájl keletkezési dátuma alapján kialakult digitalizálási sorrendben az egy személyhez kapcsolódó iratok többnyire egymás után következtek, legalábbis doboz szinten. Az összetartozó képek megjelöléséhez Záros Zsolt informatikus kollégánk készített képösszekapcsoló programot, amelynek segítségével a digitalizált iratokat áttekintő nézetben jelenítettük meg, így váltak láthatóvá fizikai jellemzőik alapján az összetartozó oldalak. Egyes esetekben nem lehetett elkerülni az iratok szövegébe történő beleolvasást, mivel csak így lehetett a kapcsolatot teljes bizonyossággal megállapítani.

Legtöbb esetben azonban az irat egy oldalból állt, a hátlap pedig nem tartalmazott adatot, ezért nem is digitalizálták. Amennyiben a főkarton hátoldalán is szerepelt adat, akkor azt is felvételezték, ezek alapján jöttek létre a kettes kapcsolatok. Ha az adott főkartonnak nevet nem tartalmazó melléklete volt, hármas, illetve négyes kapcsolatok kialakítására is szükség volt, erre azonban csupán mintegy 150 esetben került sor. Az „összekattintással” létrejött kapcsolatokat változó, a számítógép által generált színjelölés különböztette meg egymástól. Annak érdekében, hogy a kapcsolatok egyedileg megkülönböztethetők legyenek a közel 70 ezer képfájltra vonatkozóan is, a szoftver oly sokféle színárnyalatot (és a háttérben ezekhez tartozó húsz karakterből álló betű- és számkombinációt) használt, hogy azokat egymás mellé helyezve az emberi szem nem is lenne képes megkülönböztetni őket. Szövegolvasással rendkívül időigényes lett volna az összerendelési döntést meghozni, ezért választottuk a külső megjelenésre alapozott döntést, mivel így a bélyegkép nézet is elegendő volt az összekapcsoláshoz. Az emberi szem és intelligencia rövid idő alatt megtanulta értelmezni a látottakat, és a vártnál is gyorsabban tudtunk haladni e feladattal, a módszer megbízhatónak és egyértelműen hatékonynak bizonyult.

A digitalizálási sorrend nem mindig bizonyult logikusnak, előfordult, hogy a hátlapot előbb digitalizálták, mint a főoldalt, vagy hogy a nevet nem tartalmazó melléklet a főkarton elé került. Az olvasási sorrendet többes kapcsolatoknál, bonyolultabb esetekben manuálisan állítottuk be.

Az adatbázist 2023 áprilisában ugyan átadtuk a felhasználóknak, ám ezzel az utómunkálatok korántsem értek véget, további fejlesztési lehetőségek előtt állunk. A 40 doboznyi iratanyag digitális változatát részben személyre rendezetten kaptuk meg az orosz társintézménytől. Egy dobozon belül egymás után voltak megtalálhatók az ugyanarra a személyre vonatkozó iratok, ám akadtak olyan személyek, akiket többször is nyilván tartásba vettek, hiszen a korszakban még nem léteztek online adatbázisok, amelyekből azonnal nyilvánvalóvá válhatott volna a kétszeres vagy akár háromszoros adatfelvétel. Ezek az ismétlődő személyi kartonok elszórtan találhatók az iratanyagban.



A kép a képösszekapcsoló program működését mutatja be, szaggatott keretben utólag kiemelve két kettős kapcsolatot láthatunk. Jól kivehetők a kapcsolatot igazoló fizikai jellemzők: a hátlapok az előlapok külső megjelenését tükrözik: szakadást, foltokat vagy gémkapocsnymokat láthatunk.

A Stratis Kft. az MNL megbízásából elkészítette a MIREL szoftvert, amely egy mesterséges intelligencián alapuló személynév entitás egységesítő program. A szoftver felhasználására elsőként a *Szovjet táborok magyar foglyai adatbázis* készítésénél került sor. A program a Név információ segítségével meghatározza azon személyek halmazát, amelyek duplikációk lehetnek a névazonosságot vagy nagyon magas névhasonlóságot véve figyelembe. Segdinformációként lehet többek között használni a személy foglalkozását, születési helyét, lakhelyét és vallását is. A hasonlóság mértékének definiálásához az egyes adatpontok között karakterhasonlósági (pl. jaro-winkler, levenshtein) és szóhasonlósági metrikák (pl. TF-IDF, jaccard) alkalmazhatóak. Gépi tanulási modell segítségével meghatározható, hogy milyen valószínűséggel lehet szó ugyanarról a személyről. Célnk az egy személyhez kapcsolódó összes irat felderítése, és bizonyított többes adategyezés esetén automatikus összerendelése. Kérdéses esetekben humán erőforráson alapuló validálás válik szükségessé. Nagy eltérés esetén pedig a szoftver egyszerűen elveti a rekordkapcsolás lehetőségét. A cirill betűs, oroszra fordított kartonok is ezzel a módszerrel kerülhetnek majd kapcsolatba az eredeti, magyar nyelven kiállított nyilvántartási kartonnal.

Az alábbi képen egy példát láthatunk az adatok összevetésére: Stiblar Istvánról két kartont is kiállítottak, a születési hely és a születési idő mindkettőn megegyezik, a foglalkozásnál molnár, illetve molnárségéd megjelölést olvashatunk. Az adategyezések alapján a két személyrekord összekapcsolható.

NÉV_A	NÉV_B	SZÜL_DÁT_A	SZÜL_DÁT_B	SZÜL_HELY_A	SZÜL_HELY_B	FOGLALKOZASA_A	FOGLALKOZASA_B
Stiblar István	Stiblar István	1889	1889	Bodoló	Bodoló	molnár	molnárségéd
Szekér András	Szekér András	1883.09.05.	1883.09.05.		Hosszúhetény	téglaégető	téglaipari munkás
Steinitz Johanna	Steinitz Johanna	1881.11.02.	1881.11.09.	Kaposvár	Kaposvár		

Az MNL a Nemzeti Emlékezet Bizottságával együttműködésben 2018-ban publikálta a Párt-Állam-Párt, a pártállam állampárti funkcionáriusainak adatbázisát, amelynek évköre 1948–1989 (www.paap.hu). Az adatbázisban összegyűjtve találhatóak meg a kommunista párt helyi képviselői, akik megyei, járási, városi szinten vezető tisztséget töltöttek be. Ezek a vezetők a Horthy-korszakban megfigyelt személyek lehetnek, ezért valószínűsíthető, hogy szerepelnek a „Csendőrkarton-adatbázisban” is. A két adatbázis egybevetése, az összetartozó rekordok összekapcsolása is szerepel terveink között.

A hetven közösségi segédletkészítő munkája révén megvalósult projekt sikeresen zárult, a kutatók egy 70 000 személyrekordot tartalmazó adatbázist vehettek birtokukba 2023 áprilisában. A közösségi segédletkészítőkkal és a Magyar Családtörténet-kutató Egyesülettel folytatott együttműködés elmélyítette a levéltár és az általa biztosított szolgáltatásokat igénybevevő felhasználók közötti kapcsolatot. A pozitív visszajelzések birtokában folytatni kívánjuk a közös, minden érintett fél számára hasznos munkát.

A projektben részt vevő MNL munkatársak: Szatucsek Zoltán (igazgató), Bánki Zsolt (osztályvezető), Szerényi Ildikó (projektvezető), Kántás Balázs (tudományos munkatárs), Németh István (tudományos munkatárs, osztályvezető) Záros Zsolt (informatikus) Debreczeni László (informatikus).

Szakirodalom

- Hegedűs, 2020 = Hegedűs, István: How Artificial Intelligence and Machine Learning Can Help Rethink Archives? *Atlanti+*, 30/2 Trieste–Maribor, 2020. 57–64.
- Hegedűs, 2021 = Hegedűs István: Hogyan segít újragondolni a levéltárak szerepét a mesterséges intelligencia és a gépi tanulás? *Levéltári Közlemények*, 2021. 89–98.
- Szabó, 2022 = Szabó Dorottya: European Digital Treasures – új perspektívák a digitális világban. *Levéltári Szemle*, 2022. 2. sz. 57–72.
- Szerényi, 2022a = Szerényi Ildikó: Az 1828. évi országos összeírás feldolgozása mesterséges intelligencia támogatásával és önkéntesek segítségével. *Levéltári Szemle*, 2022. 2. sz. 73–86.
- Szerényi, 2022b = Szerényi, Ildikó: Indexing Using Artificial Intelligence and Involving Volunteers at the National Archives of Hungary (NAH) *Atlanti+*, 32/2 Trieste–Maribor, 2022. 15–23.