

## Azonos személyekre vonatkozó név besorolási rekordok automatikus felderítése a PIM adatbázisában

*Cikkünkben áttekintést adunk a Petőfi Irodalmi Múzeum névterébe beépített legfontosabb adatbázisokról, a személynév besorolási rekord modelljéről. Ezt követően részletes ismeretével szolgálunk a névrekordok egységesítésének folyamatáról, s az egységesítést követő feladatokról. Végezetül röviden kitérünk a PIM névtér hasznosulásának jövőbeli lehetőségeire is.*

Tárgyszavak: névtér; adatbázis; adatmodellezés

### A Petőfi Irodalmi Múzeum szakmai tevékenységének bemutatása

A *Petőfi Irodalmi Múzeum* (PIM) a magyar irodalmi muzeológia országos múzeuma, egyike Magyarországnak legjelentősebb múzeumainak. A magyar irodalom országos múzeumaként minden típusú dokumentumot, tárgyat és alkotást gyűjt, ami valamelyik magyar íróval, életművével vagy egy konkrét művével kapcsolatos. Így könyv, kézirat, fotó és film, képzőművészeti alkotás és személyes használati tárgy, hangzóanyag és elektronikus dokumentum egyaránt megtalálható a gyűjteményben.

Múzeumunk szellemisége nyitott és befogadó a korszerű muzeológiai gondolatok, trendek és szakmai megoldások számára, ezért tudományos munkánkban, kiállításainkban, rendezvényeinken folyamatosan alkalmazzuk az irodalmi muzeológia legfrissebb eredményeit. Ennek megfelelően a gyűjtemények feldolgozásában és publikálásában alkalmazott IT eszközök jelen vannak tevékenységünkben már az 1980-as évek óta. Hosszas fejlődés folyamán végigjártuk a múzeumi gyűjteményi adatbázisok fejlesztésének útját, tudományos kutatói rendszereket építettünk, és mára korszerű integrált gyűjteménykezelő rendszert (Qulto ICMS) tudhatunk magunkénak, amelyben műtárgyaink feldolgozása mellett jelentős névteret is alkalmazunk, illetve építünk. Ennek a névtérnek kiemelkedő eleme a személynevek besorolási állománya, amely nagyságrendileg 600 000 besorolási rekordot tartalmaz, és

mint ilyen a legnagyobb magyar közgyűjteményi személynévtér. Az állomány döntő többsége hitelesített, kontrolált, döntően nyomtatott forrásból származó megbízható adatokon alapul, és publikálva van a múzeum online katalógusán (*opac.pim.hu*) keresztül. A személynévtér a magyar könyvtári és múzeumi rendszer elsődleges névazonosításra szolgáló forrása.

Névterünkre elsősorban műfaji gazdagság a jellemző: életrajz, adattár, genealógia egyaránt képviselve van. Több olyan, kutatás eredményeképpen létrejött szakértői rendszert építünk, amely tematikáját tekintve egyedülállóan számít Magyarországon. A munkatársakat mindvégig az a cél vezette, hogy az adatokat hiteles forrásokból szolgáltatassák, és minden lényeges eltérést, különbséget jelezzenek, valamint, hogy az adattartalmak alkalmasak legyenek tudományos igényű hivatkozásokra.

A lexikográfiai alapú besorolási adatokból álló adatbázisok nagy része életrajzi típusú adatbázis, melyekben a vizsgált személyhez az életére, munkásságára vonatkozó adattartalmak jelennek meg.

Az adatgyűjtés jellegénél fogva a névtér jelentős számú eleméhez nem kapcsolódik a múzeum rendszerében bibliográfiai rekord, de a műtárgyak feldolgozásához alapot szolgáltat, vagyis folyamatosan növekszik a bibliográfiai hivatkozással ellátott névtérellemek száma.

## Névterünkben épített legfontosabb adatbázisaink

Magyar életrajzi index  
Emigráns írók lexikona  
Díjak és díjazottak  
Magyar főnemesség genealógiája  
Magyar családtörténeti adattár  
Magyar írók sírjai  
Francia becsületrend magyar kitüntetettjei  
Diplomások adattára  
Kortárs magyar írók adattára

Ezek közül a három legjelentősebb, legtöbb rekordot tartalmazó adatbázis:

### Magyar Életrajzi Index

A múzeum legnagyobb életrajzi adatbázisa, magyar viszonylatban a faktográfiai tájékoztatás egyik legfontosabb internetes forrása, amely segíti a könyvtárosok, bibliográfusok, biobibliográfiát készítő kutatók munkáját. A rekordok nem lexikon-szócikkek, csupán olyan adatsorból épülnek fel, amelyek segítségével egy személy önmagával azonosítható: születés, halálozás helye és ideje, foglalkozás, névvariánsok. A Magyar Életrajzi Index további funkciója, hogy kijavítja a feltárt nyomtatott források hibáit. E korrekciók – kizárólag a születési és halálozási adatsorok esetében – jelzik, hogy egy-egy forrás adatai mennyiben tér el az általunk hitelesnek minősített adattól.

### Magyar Emigráns Írók és Műveik

Az emigráns kifejezés jelen esetben kivándorlót jelent, s vonatkozik olyan szépírókra, közírókra, akik valamilyen okból hosszabb-rövidebb ideig külföldön éltek és alkottak. A szócikkek életútjuknak ezt az időszakát tárják fel.

Feltünteteti a keresett személy felsőfokú tanulmányainak adatait, az ország elhagyásának dátumát, a befogadó ország(ok) nevét, a tudományos vagy közéleti karrier állomásait, az elnyert díjakat és kitüntetések, valamint mindazon szervezeteket, melyek az adott személyt tagjai közé választották, végül pedig a publikációs forrásokat és az azokban használt álneveket is.

### Díjak és Díjazottak

A falerisztika egyetlen magyar nyelvű adatbázisa, mely az 1945 után kiosztott magyar vonatkozású, illetve magyar alapítású díjakat és díjazottak ne-

veit tartalmazza. A gyűjtés nemcsak az irodalmi és kulturális díjakra terjed ki, hanem más tudományterületekéire is. A szócikkekben: az alapító személyét, az alapítás helyét és évét, valamint a díj történetét találhatjuk.

Személynév besorolási rekordjaink nemcsak az általánosan elvárt egységesített adatalemekkel rendelkeznek, hanem olyan metaadatokkal is, mint foglalkozás, pontos halálozási, születési dátum és hely. A rekordok gazdagságát mutatja az adatforrások feltüntetése és az elnyert díjak/kitüntetések mező is. Feltüntetjük a rokonsági kapcsolatokat is, amelyek szintén alkalmasak arra, hogy szemantikus webben megjelenjenek.

### A PIM személynév besorolási rekordjának modellje

Egységesített név	(100\$a\$j)
Nemesi előnév	(100\$g)
Főnemesi rang	(100\$c)
Születés / Halálozás	(100\$d)
Rokonsági kiegészítő	(100\$m)
Névvariáns	(400)
Neme	(922)
Vallása	(923)
Foglalkozás	(667)
Születési hely	(902)
Születési megye	(904)
Pontos születési dátum	(900\$a)
Keresztelési dátum	(901\$a)
Halálozási hely	(906)
Halálozási ország	(908)
Pontos halálozási dátum	(905)
Halálozási életkor	(909\$a)
Temetés napja / helye	(910\$a)
Temető	(911\$a)
Parcella	(912\$a)
Életút időpontjai	(924\$a\$o)
Publikáció	(930)
Fordítás	(931)
Lakhely	(921)
Egyetem, kar	(925, 926)
Tudományos cím	(927)
Tagság	(928)
Díj / Kitüntetés / Fokozat	(960\$a)
Díj / Kitüntetés / Fokozat éve	(960\$d)
Díj / Kitüntetés / Fokozat típusa	(960\$4)
Díj / Kitüntetés / Fokozat – eltérés	(960\$x)
Megjegyzés a díjról	(929)
Intézmény neve	(610)
Szöveg / Indoklás	(680)
Önéletrajz	(941)

Születési adat forrása	(942)
Születési adat forrása	(943)
Halálozási adatok	(944)
Rokoni kapcsolatok	(500)
Adatforrások	(940)

## A névegységesítés folyamata

### Névegységesítés – első fázis

A korábban külön-külön épülő adatbázisokról a MARC alapú Qulto ICMS integrált rendszerre való áttéréskor a névtér építői meglehetősen sok eltérő adattartalmú/adatgazdagságú rekordisméltóddal találták szemben magukat. A névtér 620 000-nél több névrekorddal rendelkezett, de mindez lényegesen kevesebb személyt jelentett a valóságban. Például egy személy „alpnévrekordja” mellett szerepelt egy-egy rekord a nekrológot, a gyászjelentést és az adott személy sírhelyét feltáró adatbázisban is. Ezért 2008-ban elkezdődött a névegységesítés első fázisának megvalósítása.

### Az *eminens rekord fogalma*

Az „alpnévrekordot” *eminens rekordnak* neveztük el, mely az ugyanazon természetes személyt leíró rekordok közül a hitelesnek elfogadottat jelenti. A rekordegységesítés során alapmezőknek a nevet és a születési dátumot (100\$a, 100\$j, 100\$d) tekintettük, járulékos mezőknek pedig a további besorolási és kapcsolódó biográfiai adatokat (6xx, 9xx). Mivel a különböző forrásokból érkező, eltérő adattartalmú rekordok egyaránt megtartandó adatelemeket hordoztak, fő célunk az volt, hogy az egyértelműen azonos entitások egy névrekord alá kerüljenek, adatvesztés nélkül gazdagítva az *eminens rekord* adatait, összekapcsolva a múzeum gyűjteményébe tartozó bibliográfiai rekordokkal. Szándékunk szerint a névrekord-egyesítés első körét teljesen automatizáltan hajtottuk végre, emberi intelligenciát csak az egységesítési algoritmus elkészítéséhez alkalmaztunk. Ennek érdekében a gépi egységesítés során olyan szigorú feltételeket szabtuk meg két névrekord azonoságát és összehasonlíthatóságát illetően, hogy csak a minden kétséget kizáróan azonos személyek rekordjai kerüljenek az *eminens rekord* mezőibe.

Ennek eredményeként a névtérben már az első névegységesítés után az *eminens névrekord* számottevő adatgazdagodása jött létre. Az áttöltött, immár feleslegessé vált névrekordokat törölve lényegesen csökken névterünk redundanciája

is, ezzel újabb lehetőséget adva egy magyar nemzeti névtér alpnévállományának megalapozásához szükséges, minél teljesebb, pontosabb és a munka közben formálódó „egy entitás-egy rekord” elvének.

Az első ilyen összevonásnak köszönhetően mintegy 80 000 rekorddal csökkent a személynév besorolási állomány, ami legalább 140 000 név azonosítását jelentette.

### Névegységesítés – második fázis

A névrekord-egységesítés első köre után világossá vált, hogy újabb, az előzőnél is finomabb szempontrendszer alapján végzett rekordegységesítésre van szükség, ezúttal azonban már az emberi intelligencia bevonásával. Az egységesítés első köre után újabb egységesítésre váró párok keletkeztek és fény derült egyes adatszoportokkal kapcsolatos anomáliákra is. Célként azt jelöltük meg, hogy eljussunk addig a határig, amíg gépi algoritmussal eldönthető két entitás azonosága, ahol pedig nem, arról automatizált lekérdezések, listák segítségével, intellektuális azonosítás révén jussunk el két rekord egységesíthetőnek nyilvánításáig. Azon, önmagukkal azonos névintitások rekordjainak egységesítése vált szükségessé, melyek nem kerültek bele az első fázisba, mert az akkor azonosításra felhasznált evidens adatelemek kevésnek bizonyultak az egyértelmű azonosításra.

Az egységesítés e második fordulójához elsőként a különböző adatbázisok adatgazdagsága alapján prioritási sorrendet állapítottunk meg, mely az *eminens rekord* kijelölésének is alapjává vált. Így került első helyre az *eminensség* szempontjából a *Magyar életrajzi index / Név* és utolsó helyre a *Becsületrend magyar kitüntetettjei* adatbázisából származó nevek csoportja.

### **Prioritás megállapítása, *eminensség* sorrendje:**

1. Magyar életrajzi index / Név
2. Magyar életrajzi index / Partecédula
3. Magyar életrajzi index / Nekrológ
4. Magyar emigráns írók és műveik
5. Díjazottak
6. Magyar családtörténeti adattár
7. Magyar főnemesség genealógiája
8. Budapest topográfia / Lakók
9. Magyar írók sírjai
10. Diplomások adattára
11. Kortárs írók bibliográfiája
12. Gyűjteményi név

13. Életrajzi index / Település
14. Ki kicsoda?
15. Magyar életrajzi index / Gyászjelentés
16. Emigráns írók és műveik / Intézmény
17. Becsületrend magyar kitüntetettjei

Az adatgazdagság és relevancia alapján pontrendszert állítottunk fel, melynek alapján párokat képeztünk. Párok (ezúttal is) azok lettek, ahol a teljes, egységesnek elfogadott név megegyezik, és a születési vagy halálozási év sem különbözött. A szigorú pontrendszer alapján 11 784 olyan eminens találtunk, melynek csak egy nem eminens párja keletkezett. Mivel ezt kevésnek találtuk, ezután lista készült azokról a párokról, melyek nem biztosan, de nagy valószínűséggel egy eminenshez köthetők. A pároknál az adatkitöltöttség plusz pontokat, az anomáliák mínusz pontokat jelentettek, ezek alapján a párok pontértéket kaptak.

#### **A „párkapcsolat” értékelése:**

- Ha a születési év a pár minden tagjánál ki volt töltve és azonos volt, 1 pontot kapott.
- Ha a halálozási év a pár minden tagjánál ki volt töltve és azonos volt, 1 pontot kapott.
- Ha a születési és halálozási év egyaránt mindkét esetben ki volt töltve és meg is egyezett, újabb 2 pontot kapott.
- Ha a900 (születési dátum) megegyezett 2 pontot kapott.
- Ha születési település megegyezett 2 pontot kapott.
- Ha születési megye megegyezett 1 pontot kapott.
- Ha halálozási dátum megegyezett 2 pontot kapott.
- Ha halálozási település megegyezett 2 pontot kapott.
- Ha a900 (születési dátum) különbözött -2 pontot kapott.
- Ha születési település különbözött -2 pontot kapott.
- Ha születési megye különbözött -2 pontot kapott.
- Ha halálozási dátum különbözött -2 pontot kapott.
- Ha halálozási település különbözött -2 pontot kapott.
- Ha a páros egyik vagy mindkét tagjának volt olyan párja mely a 900-as csoportban valamely értékpárban eltért anomáliának lett minősítve, akkor az illető pár gyanúsnak lett minősítve.

- Minden gyanús 900-as elemért egy pontlevonás járt.
- Ha a halálozási vagy születési év hiányzott valamelyiknél, akkor egy pontlevonás járt.
- Ha a keresztnévet valamelyiknél bontott és a kapcsolat úgy jött létre, akkor egy pontlevonás járt.

36 000 párosítást néztünk át és valamivel több, mint 30 000 pár azonosságát állapítottuk meg, mely 58 000 névrekordot érintett. Az ellenőrzés során nagy gondot jelentettek a gyakori nevek (Tóth, Kovács, Kiss stb.), mivel ezek viselői nem pusztán nagy számban voltak jelen az adatbázisban, de sokszor más írásmóddal voltak jelen a különböző adatbázisokból származó névrekordokban (pl. Kovács, Kovách, Kováts). Ugyancsak sok gondot jelentett a többtagú keresztnévű személyek azonosítása, különösen, amikor nehezen eldönthető valakiről, hogy például kettős nevéből a második tag a vezetéknevhez vagy a keresztnévhez tartozik (pl. Kiss András Béla). Ugyancsak gondot jelentett, ha az azonosnak tekinthető névrekordok egyikében szerepelt nemesi előnév, a másikban nem, illetve ha a nem eminensnek tekinthető rekord szerint névváltoztatást hajtott végre az eminens rekorddal minden bizonnyal megegyező személy.

További problémát jelentett, hogy az egyes nem eminens rekordokban található (leggyakrabban a legfontosabb születési/halálozási) adatok eltérnek az eminensnek tekintett rekordban találhatóéhoz képest. Ezért több, ebben érintett mezőt ismételtetővé tettünk, mert adatot nem akartunk eldobni: lényegében nem duplumszűrést, hanem egységesítést végeztünk. A gondos előkészítő munkálatokkal és ellenőrzésekkel együtt is természetesen, hogy a névrekordok egységesítése közben és után is végeztünk ellenőrzéseket. A köztes ellenőrzés elsősorban a „próbaegyesítés” során tapasztalt hibák ellenőrzését jelentette (pl. az áttöltés során az eminens rekordba a díj neve átkerült, de a díjazás éve nem; az elvégzett egyetem neve megjelent, de az elvégzett kar neve nem stb.), a végső ellenőrzés pedig azt, hogy valóban megjelennek-e az egységesített adatok az eminens rekordban.

#### **Az egységesítést követő feladatok**

Az egységesítés több újabb feladatot is magával hoz. Támogató, lekérdezéseket és ellenőrző listákat kérünk azokról az eminens rekordokról, ahol változott vagy nőtt a rokoni relációk száma

(több apa, több anya) Ugyancsak listatámogatást kérünk olyan eminensekről, ahol ismétlődik a születési/halálozási dátum, mert több eltérő dátum képződött.

Az egységesítés második körének több egyéb hozadéka is volt.

- Kiderült, hogy azonos **földrajzi nevek (települések)** eltérő írásmóddal szerepeltek az adatbázisban illetve az idők során különböző módon szerepeltették a születési/halálozási dátumokat.
- Kiemelt feladatnak mutatkozik a **foglalkozásnevek** egységesítése, ami ismét csak tisztábbá tenné az egyre több eminens, és egyre kevesebb nem eminens rekordot tartalmazó névtérünket.
- Ugyancsak e második egységesítési kör eredménye, hogy körvonalazódnak olyan algoritmusok, melyek keresés formájában **évente lefutathatók**, így az eddigi gyakorlat alapján lehetőség nyílik a névtér évenkénti tisztogatására, és ezáltal a további redundáns rekordok újratermelésének elkerülésére.
- Az egységesítés érdekében új szempontú megközelítésként lekérdeztetjük azokat a rekordokat, ahol a **név figyelembe vétele nélkül a születési/halálozási hely és idő azonossága alapján** próbálunk újabb névpárokat találni. Azokat a rekordokat listázzuk, melyeknél a négy szempontból legalább három megegyezik.
- Végül nem túl hálás feladatnak mutatkozik az **adatszegény néventítések** azonosítása gyűjteményeink bibliográfiai rekordjai alapján.

Elvi célunk a „teljes névrekord” elérése, ahol a születési, halálozási adatok mellett az adott személy foglalkozása, díjai, elvégzett iskolái, családtagjai, különböző társaságokban betöltött tagsága, halálához köthető gyászjelentése és nekrológadatai is az eminens névrekord alatt szerepelnek.

### Technikai kivitelezés

Az egységesítő algoritmus célja, hogy a rendszerben különálló besorolási rekordok esetében, ha azonos személyt írnak le, az azonosság feltételeinek teljesülése esetén párok jöjjenek létre. Mivel a személynév besorolási rekordok eredetileg a korábbi elszigetelten épülő kis adatbázisok konverziójából, illetve a folyamatos név besorolási állomány építése, valamint a gyűjteményi feldolgozó munka során jöttek létre, sok esetben redundánsnak tekinthetők, mert egy azonos ter-

mészetes személyre több besorolási rekord is vonatkozott. A pár egyik tagját a fentebb ismertett módon eminensnek kell tekinteni, ennek kell megmaradnia, a másikat, a törlendőt pedig nem eminensnek. Természetesen a pár eminens tagja része lehet egy másik párnak, ahol az illető rekord nem eminens.

A pár tagjainak minősítésénél ügyelni kellett, hogy a szempontrendszer annyira pontos legyen, hogy az eminens nem eminens párok rácsozatában ne képződjék kör. A rácsozat csúcsán csak szuper eminens rekordok állhattak, melyek már nem szerepelnek egy párban eminensként. Ezt olyan módon értük el, hogy első szempontként a személynév besorolási rekordokat tartalmazó adatbázisokat állítottuk rangsorba és a párból mindig a magasabb rangú adatbázishoz tartozó volt az eminens. Azonos adatbázisba tartozóknál mindig a kisebb rekordazonosítóval rendelkező rekord lett az eminens. Az adatok kitöltöttségét, minőségét nem kellett figyelni, mert úgyminden adattartalom egységesítve lett az eminens rekordhoz. A névalak helyes vagy pontatlan voltát gépi úton nem lehet eldönteni, ez a feladat emberi szakmai döntésként a katalógizálókra hárult és kézzel kellett elvégezni. A párok így minden esetben kiegészülhettek egy-egy szuper eminens taggal, akire végül minden információ át lett kötve.

A párosítás másik kritikus pontja a villák képződése lehetett. A villa azt jelenti, hogy egy eminens nem eminens rekordpárokból álló hálózatnak több csúcsa is van, több szuper eminens rekord is tartozik hozzá. A villák felderítése (mely minden esetben hibát sejtet) során fény derült a párosítás hibáira, és ez alapján bizonyos párokat tipikusan kiiktattunk, például az azonos évben született hasonló nevű személyeket, amikor a halálozási év nem volt kitöltve. Két 1942-ben született *Horváth Józsefre* vonatkozó rekord nem feltétlenül jelent azonos személyt, míg két *Keresztury-Köpcsényi Bertalan* nyilván igen. Ennek eldöntését a kereszt- és személynevek gyakoriságára utaló vizsgálatokkal próbáltuk meg automatizálni, de a számítások olyan szórt eredményeket hoztak, hogy a határt a gyakori és ritka kereszt- és vezetéknév között nem lehetett meghúzni, ezért az ilyen információhiányos párokat végül elvetettük. A párosítás feltételrendszerének szigorítása után megmaradt kevés villát egyenként kellett megnézni, és kézzel beállítani, azonosnak tekinthetők-e vagy sem.

## A párosítás technikai kivitelezésének lépcsőfokai

Az érintett besorolási rekordokat összeválogattuk. A nem törölt és megfelelően kitöltött rekordok jöttek számításba. Például, a keresztnév vagy a születési dátum teljes hiánya kizáró ok volt.

A kiválogatott neveket normalizáltuk, kis- és nagybetű, bizonyos írásmódbeli különbségek, mint Y és i, cz és c, ts és cs egységes formára lettek hozva. Minden központosítási jel ki lett iktatva, és a több tagú keresztnveket szétbontva is külön-külön név kulcsstringbe lettek beiktatva.

Például: *Mezőhegyesi Szilveszter Aladár*ból három rekord jött létre:

Mezőhegyesi Szilveszter Aladár  
Mezőhegyesi Szilveszter  
Mezőhegyesi Aladár

A vezeték- és keresztnévből, illetve a születési és halálozási évből egy kulcsstring lett képezve, és ez alapján a string alapján lettek a párok összeállítva.

Megtörtént a párosításban részt vevő mezők leválogatása, melyben a pontos dátummezőket a születési és halálozási hely, foglalkozás, illetve a névkiegészítő mezők egészítették ki. Ezzel a Qulto ICMS szegmentált MARC adatszerkezetből adatfolyam-összeépítés történt meg.

Az első párképzés után – amint azt korábban említettük – az ellenőrzés anomáliákat (dátumok és helyek eltérései) tárt fel. A második párképzésben immár „pontozott” párokat, a pontértékek szerint listáztuk, és a listák kézi ellenőrzése során megállapítottuk azt a ponthatárt, mely felett külön emberi ellenőrzés kellett az azonosság kimondásához, illetve azt, amely felett a listát már át sem kellett nézni, mert az azonosság kételyen felül állt.

Ezt követően az egységesítés több lépésben történt meg:

- elsőként a kézenfekvő azonosságú tételek összevonása,
- majd a bizonytalan azonosságúaknál is még egy halmaz ismét kézi ellenőrzés nélkül is azonosnak lett tekintve.

A munkafolyamat minden lépésnél azonos elemekből épült fel:

1. behasonlítás kulcsképzés,

2. normalizálás,
3. párok válogatása,
4. párokból többszörös kapcsolatok kapcsán gráfok képzése,
5. körök és villák kiiktatása,
6. végleges párok összeállítása,
7. azonos kapcsolatok kiiktatása,
8. szuper eminensek összeválogatása,
9. párok ellenőrzése, törölt tagból álló párok figyelmen kívül hagyása (a folyamat hosszú ideig zajlik, egy élő, működő folyton alakuló adatbázison),
10. párosítás után ellenőrzés,
11. végleges párosítási lista összeállítása,
12. adategységesítés, először ellenőrző nézet kialakítása, majd tényleges adat-összeolvasztás, azonos tartalmú, azonos mezőkre vonatkozó és új mezők egymástól eltérő kezelése,
13. a felesleges adattartalmak és rekordok törlése.

## A PIM névtér hasznosulása – a magyar Nemzeti Névtér koncepciója

2012-ben a *Magyar Nemzeti Digitális Archívum*, az *Országos Széchényi Könyvtár*, a *Magyar Országos Levéltár* és a Petőfi Irodalmi Múzeum szakemberei elkezdtek kidolgozni egy kollaboratív elven épülő Nemzeti Névtér koncepciót, felismerve a már működő jelentős névtérek (Getty, VIAF, ICONCLASS) szakmai hasznosulását. Egyértelműnek tűnt számukra, hogy a jövő a szemantikus webben publikált névtérhasználaté, amely egyaránt kiszolgálja az archívumi és a legszélesebb felhasználói igényeket.

A szakemberek között abban is egyetértés alakult ki, hogy a közös névtér alapjait az egyes intézményi névtéreknek kell képezniük, és egyes kitüntetett minőségű archívumi névállományok képezhetnék a kiindulópontot.

Bár a magyar Nemzeti Névtér még nem valósult meg, de az elméleti tervezés java része már elkészült. A PIM névegységesítési projektje összhangban áll a Nemzeti Névtér megvalósítását megalapozó – *Szakadát István* és *Ungváry Rudolf* által jegyzett – tanulmánnyal. A dolgozat a következőket mondja ki:

„A Nemzeti Névtér létrehozásának és működtetésének igazi értelme abban van, hogy a névtérek közös archívumi használata révén átjárhatóvá tegyük a kulturális gyűjteményi adatbázisokat. Csak azáltal tudjuk megmutatni a magyar kultúra

teljességét, kulturális értékeink összességét a maguk egybefüggő egymáshoz kapcsolódásukban, teljes összefüggésrendszerükben, ha a közös névterek működtetése révén összekapcsoljuk a kulturális gyűjteményeinket. Amíg a kulturális adatokat a közgyűjteményi logika szerint intézményes elkülönültségben tároljuk, addig szűkszerűen csak szigetszerű hozzáférést nyújthatunk az érdeklődők számára. A cél az, hogy a látogatók, ahelyett, hogy archívumról archívumra vándorolniuk kelljen ahhoz, hogy az őket érdeklő adatokat összegyűjthessék, egyetlen integrált rendszerként láthassák és használhassák a magyar kulturális adatbázisok összességét.

A Nemzeti Névtér projekt stratégiai célcsoportja a laikus érdeklődő nagyközönség. A stratégiai cél elérésére, a nagyközönség igényeinek kiszolgálására azonban akkor van nagyobb esélyünk, ha a Nemzeti Névtér felépítéséig és biztonságos működtetéséig a Névtér adminisztráló archívumok szempontjaira figyelünk elsősorban. Taktikai megfontolások miatt tehát kétlépcsős fejlesztési ütemezést érdemes követnünk, mely szerint a fejlesztés első ütemében az archívumok szempontjait és elvárásait vesszük figyelembe, és csak a második körben építjük ki azokat a funkcionalitásokat, szolgáltatásokat, amelyek a közönség érdeklődésére is számot tarthatnak.

Felmerült javaslatként, hogy a Nemzeti Névtér felvállalhatna egyfajta adathitelesítő szerepet is. Az elképzelés szerint akár érdeklődő laikus, akár adatgazdai szervezet keresné fel a Nemzeti Névtérrel, az a megkeresésekre hitelesített, ellenőrzött, egyértelműsített adatokat szolgáltatna. Ezt a stratégiai célt természetesen vállalhatóan, sőt, vállalandónak kell tartanunk...”

E koncepció értelmében a PIM személynévterének tisztítása, az entitások egyértelműsítése, azonosítása jelentős lépést jelent a magyar Nemzeti Személynévtér megvalósulása irányába. Úgy gondoljuk, hogy a múzeum személynév besorolási állománya megfelelő alap lehet a közös névtér számára.

## Irodalom


KÓMÁR Éva – LENGYEL Monika – SIMON András: A Petőfi Irodalmi Múzeum személynév állományának egységesítése és szűrése: Hozzájárulás az NDA névtér projektjéhez. = Networkshop 2008. Országos konferencia. IIF program. Dunaújváros, (2008)

SIMON András: A Petőfi Irodalmi Múzeum személynévterének publikálása a szemantikus weben. = Networkshop 2016. Országos konferencia. IIF program. Debrecen, (2016)


DANCS Szabolcs: A Bibliográfiai Átállás programja – az RDA franciaországi átültetése. = TMT 63. évfolyam (2016) 9. szám:

[http://tmt.omikk.bme.hu/show\\_news.html?id=6264&iss\\_ue\\_id=583](http://tmt.omikk.bme.hu/show_news.html?id=6264&iss_ue_id=583)


Beérkezett: 2016. XI. 15-én.




**Bánki Zsolt István**  
Petőfi Irodalmi Múzeum,  
Könyvtár és Informatika,  
főosztályvezető.  
E-mail: [banki@pim.hu](mailto:banki@pim.hu)



**Mészáros Tibor**  
Petőfi Irodalmi Múzeum,  
Könyvtár és Informatika,  
muzeológus, múzeumi könyvtáros.  
E-mail: [meszarost@pim.hu](mailto:meszarost@pim.hu)



**Németh Márton**  
MONGUZ Információtechnológiai Kft.  
közgyűjteményi szakértő,  
tartalommenedzser.  
E-mail: [mnemeth@gmail.com](mailto:mnemeth@gmail.com)



**Simon András**  
MONGUZ Információtechnológiai Kft.  
közgyűjteményi szakértő.  
E-mail: [asimon@monguz.hu](mailto:asimon@monguz.hu)