



A Wikidata és a Nemzeti Névtér kapcsolódási lehetőségei

MOLNÁR Bence

Bevezetés

Hogyan tegyük népszerűbbé a *Nemzeti Névtér*et? Miképp érje el a Nemzeti Névtér azokat, akik még sosem hallottak róla? A következőkben olyan lehetőségeket tekintek át, amelyekkel a Nemzeti Névtér (vagy más állami közadatbázis) egyszerre növelheti ismertségét, illetve látogatottságát. Az Országos Széchényi Könyvtár (OSZK) fejlesztésében indult, majd 2020-tól a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központja (PIM–DBK) által gondozott Nemzeti Névtér (<https://nevter.hu/>, illetve <https://abcd.hu/>) projekt legfontosabb célkitűzése egy olyan felület megteremtése, ahol az érdeklődők különböző adatbázisok információihoz egy helyen, összekapcsolt és átlátható módon férhetnek hozzá – a használó így egyetlen oldal felkeresésével érheti el az általa keresett, releváns tudást. A jelenleg (2021 februárjában) még béta állapotban lévő, de a közeljövőben a pub-

likum számára véglegesülő projekt számára adatforrásként az átfogó PIM Névtértől kezdve a Központi Statisztikai Hivatal helységnévtáráig potenciálisan számos és sokféle adatbázis szolgálhat, miközben ezek gondozását a projektben kompetenciák mentén elsősorban szakemberek végzik. A kényelmes felhasználói élmény megteremtésében és az adatbázis bővítésében hasznos partner lehet a *Wikidata*, amint a Nemzeti Névtér kilép a béta állapotból, és azonosítói véglegesednek (jelenleg ezek csak ideiglenes ID-k). Az egyes adatlapok ugyanis jelenleg is tartalmaznak linkeket a magyar Wikipédiának az adatlap entitásához kapcsolódó szócikkeire; lévén, hogy a Wikipédia és a Wikidata egyazon mozgalomhoz tartozó társprojektek, a köztük való átjárhatóság biztosított. Írásomban egy, az egyszerű adatbázis-integrációnál több, kölcsönös előnyökkel járó Nemzeti Névtér–Wikidata integrációt kívánok megvizsgálni

nemzetközi és hazai példákon keresztül, reflektálva a Nemzeti Névtér és a Wikipédia között már meglévő linkelésre. Ennek megértéséhez azonban tekintsük át először, mit is takar pontosan a Wikidata.

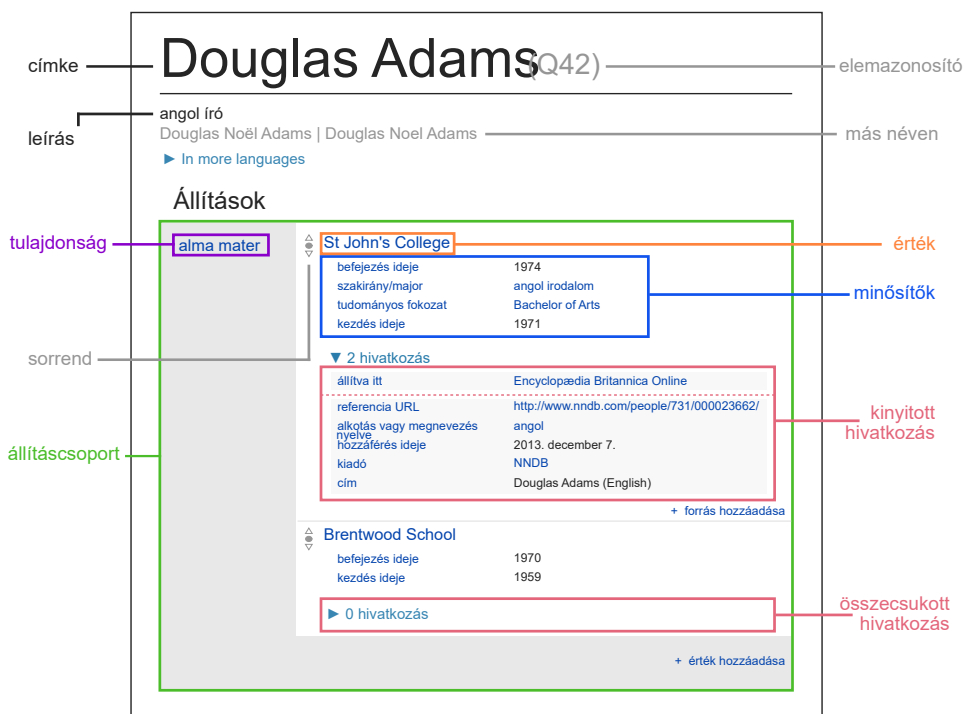
A Wikidata egy wiki-alapú (közösség által szerkesztett) tudásbázis, amelynek adatait bárki szabadon, megkötések nélkül felhasználhatja. A weboldal fenntartója a Wikimedia Foundation, amely – további projektek mellett – a Wikipédia üzemeltetését is végzi. A Wikidata 2012-ben indult, e tanulmány lezárásának időpontjáig (2021. február 5-éig) pedig több mint 1,35 milliárd szerkesztés és 92 millió elem létrehozása történt meg.¹ Kevesen tudják, hogy a Google-keresések azonnali válaszai és „tudáspaneljei” (*Knowledge Graph*), a virtuális asszisztensek és tengernyi egyéb általunk használt weboldal mögött „szürke eminenciásként” a Wikidata húzódik meg mint adatszolgáltató.

Miközben egy Wikipédia-szócikket olvasunk, az adatok egy része valójában nem is magában a szócikkben található, hanem közvetlenül a Wikidatából, úgynevezett „sablonok” segítségével kerül megjelenítésre. A személyekről szóló cikkek jellemzően két ilyen sablont tartalmaznak: egy életrajzi adatokat tartalmazó

adatpanelt- vagy dobozt (rögtön a bevezető mellett, jobb oldalon), illetve egy nemzetközi katalógusokat gyűjtő szövegdobozt (a szócikk-oldal legalján, nemzetközi katalógusok címmel).

A Wikidata felépítése

A Wikidata egy milliányi elemből álló szemantikai adatbázis, ahol világunk entitásait egy-egy elem (*item*) írja le.² Minden elem egy „Q” prefixet követő egyedi azonosítóval rendelkezik, így például *Douglas Adams* író a „Q42” (<https://www.wikidata.org/wiki/Q42>), míg a *Könyvtári Figyelő* a „Q1125403” (<https://www.wikidata.org/wiki/Q1125403>) azonosító alatt található. Az entitást adott nyelven egy címke (*label*) és egy leírás (*description*), valamint amennyiben vannak, egyéb nevek (*alias*) azonosítják. A Wikidata egy többnyelvű projekt, ezért egy elem több (eltérő nyelvű) címkével is rendelkezhet. Az entitásra vonatkozó információk állításokként (*statements*), tulajdonságok (*property*) értékeiként (*value*) tárolódnak: személyek elemeinél például a születési idő, a vezeték- és keresztnév vagy a foglalkozás a jellemző tulajdonságok.



1. ábra

A Wikidata adatmodelljének grafikai reprezentálása Douglas Adams példáján keresztül

Szerző: Charlie Kritschmar (WMDE), Samat, CC0 1.0 licenc alatt

Forrás: https://commons.wikimedia.org/wiki/File:Datamodel_in_Wikidata_hu.svg

Az állítások különleges csoportját képezik az azonosítók (*identifier*), melyek jellemzően egy külső adatforrásra, adatbázisra mutatnak, ahol szintén megtalálható az adott entitás: így például *Szentágothai János*³ a Petőfi Irodalmi Múzeum személynévterében a „PIM71401”⁴, míg parlamenti munkássága az Országgyűlés honlapján (<https://parlament.hu/>) „s014” azonosító alatt található. Ezen külső (Wikidatan kívüli) azonosítók segítségével többek között az is kivitelezhető, hogy egymástól teljesen eltérő célú adatbázisok elemeit kapcsoljuk, linkeljük össze (nyelvi korlátok nélkül), ahol a Wikidata egy központi csomópontként (*hub*ként) működik, amely az elemeket és azok azonosítóit szolgáltatja.

A legtöbb adatbázisnál fontos az adatok visszakövetethetősége és ellenőrizhetősége, a Wikidata esetében – hasonlóképp a Wikipédiához – azonban különösen fontos az adatbázisba kerülő állítások forrása, eredete és ellenőrizhetősége. Ne felejtjük el, hogy bizonyos állítások közvetlenül a Wikipédia-szócikkekben is megjelennek, így az ezeken végzett változtatások egyben az olvasók számára is azonnal láthatóak lesznek. A Wikidata – többek között – milliányi személy adatát (születési idő, hely stb.) tárolja, ezek hitelességének ellenőrizhetőségéhez pedig forrásokra van szükség. Erre szolgálnak a hivatkozások (*reference*), melyek segítségével megadható, honnan származik az adott információ, vagy éppen honnan importáltuk azt. Egy Wikidata-elem több, az entitás típusára jellemző állítást is tartalmazhat, attól függően, hogy épp egy ember vagy egy égitest eleméről van szó (előbbi esetében állítás lehet például a személy születési ideje vagy neve, míg utóbbinál az égitest sugara vagy épp fényessége).

A Wikidata a Wikipédiához hasonlóan döntően szabadon, bárki által szerkeszthető, a két társprojekt szerkesztőgárdájában pedig jelentős átfedés van (ennek egyik oka például az, hogy ha valaki létrehoz egy új Wikipédia-szócikket, azt a Wikidatan kell „összekapcsolni” a téma más nyelven már létező szócikkeivel). A korábban már említett 1,35 milliárd szerkesztés és 92 millió különböző elem azonban predesztinálja, hogy ezek többségét úgynevezett „bot” (automatizált) szerkesztők végezték; ezek a hús-vér emberek által programozott és felügyelt eszközök végzik a Wikidata és különféle adatbázisok (például a későbbiekben részletezett VIAF) közötti importálásokat, tartják karban a már meglévő elemeket (pl. végzik az átirányítások feloldását, forráshivatkozások beszúrását, a fölöslegessé vált elemeket törlését, Wikipédia-hivatkozások frissítését stb.), vagy épp hoznak létre

újakat Wikipédia-szócikkek és különböző külső adatforrások alapján. Mindazonáltal nem elhanyagolhatóak az emberi szerkesztők közreműködései sem, hiszen például az összekuszálódott, komplex entitások problémáinak disztingválásához és megoldásához nélkülözhetetlen az emberi szem (ld. például később, „Az integráció előnyei” szakaszban taglalt duplum VIAF-azonosítók kérdését), illetve az általunk végzett, entitások és elemek közti osztályozási tevékenység kiváltása jelenleg megoldhatatlan feladat. Összegezve tehát, a Wikidata döntően emberek által írt botok munkájának köszönhetően bővül, mindazonáltal a szerkesztőközösség meghatározó többségét alkotó, több tízezernyi aktív humán szerkesztő fáradozása sem hagyható figyelmen kívül (különösen a minőségellenőrzés és a destruktív szerkesztések, úgynevezett „vandalizmusok” elhárítása terén).

Virtual International Authority File – egy működő integráció

A VIAF dióhéjban

Az *Online Computer Library Center* (OCLC) *Virtual International Authority File* (VIAF) projektje nemzetközi szinten is ismert. A VIAF egy olyan nemzetközi authority adatbázis, amelynek bővítéséhez több mint 50 nemzeti könyvtár, múzeum és szervezet szolgáltat adatokat a világ számos országából, célja pedig ezek különböző névrekordjainak az összekapcsolása. Példának okáért *Szentágothai János* VIAF-azonosítója („12315483”)⁵ egy tucatnál is több nemzeti könyvtár – többek között az OSZK – azonosítóját foglalja magában; ez a kiterjedt, intézmények gyűjteményeinek összességéből álló gráf a 2. ábrán látható. A jobb oldalon elhelyezkedő interaktív gráf a különböző intézmények közötti oda-vissza linkelést is megjeleníti. Mivel írásomnak nem célja a VIAF részletes ismertetése, ezért álljanak itt csupán felsorolásszerűen a rekord legfontosabb, az alanti ábrán nem látható részei: alternatív névváltozatok (névvariánsok), kapcsolódó társszerzők és intézmények, a személy művei és azok orszámgegoszlása, valamint kiadók.

A VIAF bővüléséhez az adatforrások maguk a résztvevő partnerek; ahogy hónapról hónapra egyre több adattal járulnak a projekthez az intézmények, úgy jönnek létre, bővülnek, kerülnek szétválasztásra és törlésre a VIAF-rekordok, annak függvényében, milyen változtatások történtek a forráselemeken.

Szentágothai, János, 1912-1994

Szentágothai, János

Szentágothai János Kossuth-díjas magyar anatómus, egyetemi tanár, országgyűlési képviselő, a Magyar Tudományos Akadémia elnöke

VIAF ID: 12315483 (Personal)

Permalink: <http://viaf.org/viaf/12315483>

Preferred Forms

- 200 _1 _ta_ Szentágothai _tb_ Janos
- 100 1 _1 _ta_ Szentágothai_ János
- 200 _1 _|_ta_ Szentágothai _tb_ János _ff_ 1912-1994
- 100 1 _1 _ta_ Szentágothai_ János
- 100 1 _1 _ta_ Szentágothai_ János
- 100 1 _1 _ta_ Szentágothai_ János
- 100 1 _1 _ta_ Szentágothai_ János _td_ 1912-1994
- 100 1 _1 _ta_ Szentágothai_ János _td_ 1912-1994
- 100 1 _1 _ta_ Szentágothai_ János _td_ 1912-1994
- 100 0 _1 _ta_ Szentágothai_ János _tc_ Kossuth-díjas magyar anatómus, egyetemi tanár, országgyűlési képviselő, a Magyar Tudományos Akadémia elnöke
- 100 1 _1 _ta_ Szentágothai_ J_ János _td_ 1912-1994
- 100 1 _1 _ta_ Szentágothai_ János
- B 100 1 _1 _ta_ Szentágothai_ János
- NI 100 1 _1 _ta_ Szentágothai_ János
- 100 1 _1 _ta_ Szentágothai_ János _td_ (1912-1994)
- 100 1 _1 _ta_ Szentágothai_ János _td_ (1912-1994)
- 100 1 _1 _ta_ Szentágothai_ János _td_ 1912-1994
- 100 1 _1 _ta_ Szentágothai_ János _td_ 1912-1994

The diagram on the right shows a complex network of interconnected nodes, each representing a different language or identifier for the same entity, with various national flags (Hungary, France, etc.) around it.

2. ábra

Képernyőkép Szentágothai János VIAF-adatlapjáról
 Forrás: <https://viaf.org/viaf/12315483/>

Szentágothai János példájánál maradva, ezt magunk is lekövetthetjük adatlapjának „History of VIAF ID:12315483” fülén, ahol láthatjuk, mikor jött létre az elem, valamikor mikor kerültek hozzáadásra a különféle azonosítók a rekordhoz (pl. ISNI, OSZK, Wikidata stb.).

A VIAF–Wikipédia integráció

2012 nyarán az angol Wikipédia közössége támogatta a *Wikipedians in Residence at OCLC* program és a British Library azon kezdeményezését, hogy egy bot (számítógépes program) segítségével negyedmilliói VIAF-azonosítót importáljanak és kapcsoljanak össze az angol Wikipédia szócikkeivel.^{6,7} A projektről *VIAFbot and the Integration of Library Data on Wikipedia* címmel összefoglaló esettanulmány is készült (Klein–Kyrios, 2013). A tanulmány és a projekt érdekessége, hogy annak kivitelezését egy „szimpla” Wikipédia-szerkesztő (név szerint *Maximilian Klein*) támogatta az OCLC-ből (ez lenne a *Wikipedians in Residence* kezdeményezés lényege, hogy egy dedikált „wikipédista”, tevékenykedik egy projekten az intézménynél).⁸ Az importálás gondos előkészítést követően eredményesen lezajlott, a projekt sikere azonban egyben felvetett egy gyakorlati problémát: mivel az azonosítók importálása csak az

angol Wikipédiába történt, a projekt gyümölcsét a további több száz nyelv Wikipédiája nem élvezhette. Mivel akkor még a projektek közötti szinkronizálás megvalósíthatatlannak bizonyult, a kezdeményezés előnyeit csak korlátozott számú olvasó és Wikipédia-projekt élvezhette.

Szerencsére azonban erre megoldás kínált a Wikidata, amely egyszerre képes tárolni egy személy összes Wikipédia-szócikkének linkjeit, valamint a személy különböző külső (így pl. VIAF-beli) azonosítóit is. Erre építve egy 2013 márciusában elkezdődött projekt keretében 7 különböző katalógus (többek között a VIAF, az ISNI és az amerikai Kongresszusi Könyvtár) közel egymillió rekordját integrálták a Wikidatába. Hamar kialakult a jelenleg is működő VIAF–Wikipédia–Wikidata integráció.

Ezen tanulmány lezárásáig (2021. február 5-éig) a Wikidata több mint 2,6 millió VIAF-azonosítót tartalmaz, és ez a szám egyre csak nő és nő.

Az integráció előnyei

A VIAF egyik visszatérő és kiküszöbölhetetlen problémája, hogy mivel egyszerre sok szervezet „táplálja” adatokkal, óhatatlanul is keletkeznek benne duplumok és hibás elemek. Ezek felkutatásában – az eddigi tapasztalatok alapján – hiánypótló segítséget

jelentenek a Wikipédia, illetve a Wikidata önkéntes szerkesztői. Miközben a VIAF-azonosítók importálása folyamatosan zajlik (a korábban már említett botok és az emberi szerkesztők által), a hibák azonosítása és javítása is folyamatos.⁹

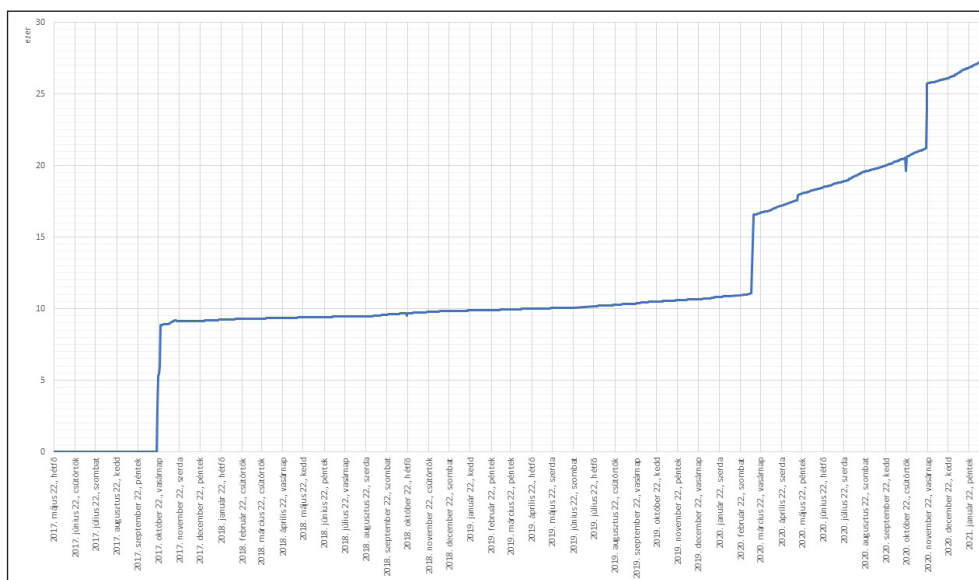
Duplumok esetén ez úgy történik, hogy a VIAF bizonyos időközönként ellenőrzi az olyan Wikidata-elemeket, amelyek egynél több VIAF-azonosítóval rendelkeznek, ugyanis ideális esetben egy elemhez csak egy azonosító tartozhat. Tehát ha egy rekord több azonosítóval is rendelkezik, az azt jelenti, hogy valami nincs rendben a VIAF oldalán; az így felfedezett, duplikált VIAF-azonosítókat össze kell vonni. Az ezt követő Wikidata-importáláskor az így feleslegessé vált VIAF-számok „eltűnnek”, és elemenként már csak egy azonosító marad.

A duplumok mellett számos egyéb, jellemzően összetettebb hibák is fakadhatnak abból, hogy a VIAF különböző katalógusokból készül. Leggyakrabban azonos nevű személyek kerülnek tévesen egy azonosító alá (noha például más korban éltek), így akár matematikai művek is megjelenhetnek egy csupán orvostudományi területen publikáló orvos adatlapjánál.¹⁰ Az ilyen anomáliák és még sok más hiba megtalálásában felbecsülhetetlen segítséget nyújtanak az önkéntes szerkesztők, tehermentesítve a VIAF munkatársait. A VIAF számára a Wikipédiával történő együttműködés egyik leglátványosabb előnye

mégis az olvasók számának növekedése: mint ahogy arra a korábban már említett esettanulmány szerzői is rámutattak (Klein–Kyrios, 2013), a VIAF-azonosítók importálása az angol Wikipédiába látogatószám-növekedést eredményezett a VIAF számára. A projekt keretében importált negyedmillió azonosító szignifikánsan megemelte az oldal – angol Wikipédia felől érkező – látogatóinak számát. Mindez még akkor történt, amikor a Wikidata még nem volt alkalmas hasonló projekt megvalósítására, mára azonban a Wikidatában tárolt 2,6 milliónál is több VIAF-azonosító különböző nyelvű Wikipédiák szócikkeinek millióin keresztül „kattintható” (érhető el).

Petőfi Irodalmi Múzeum – a magyar példa

A Wikidata szép számmal tartalmaz magyar vonatkozású azonosítókat is (pl. Magyar Tudományos Művek Tára-szerzőazonosító, MTA-tagazonosító, MEK-azonosító, KSH településazonosító törzsszám stb.¹¹), ám ezek közül is – a tanulmány szempontjából – a Petőfi Irodalmi Múzeum személynévtére a legjelentősebb. A PIM névtér az alapvető életrajzi adatokon túl az adott személy díjait és kitüntéseit is listázza (ezek saját részadatbázissal is rendelkeznek), ezért nem túl meglepő, hogy a Wikipédia-szerkesztők előszeretettel használják forrásként Wikipédia-szócikkeikben.



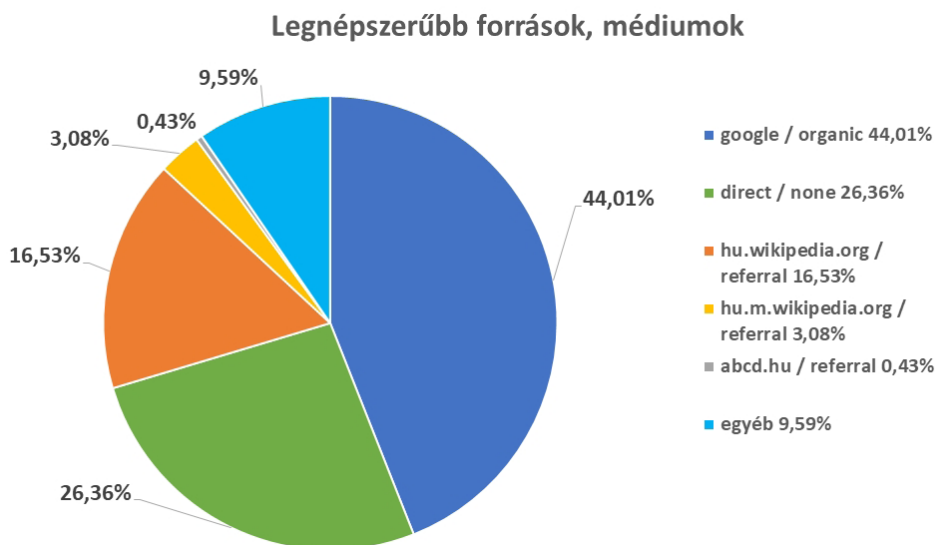
3. ábra

A PIM-névtér-azonosítók számának változása a Wikidatán 2017. május 22. és 2021. február 5. között
A szerző saját gyűjtése és szerkesztése

Forrás: https://www.wikidata.org/wiki/Template:Property_uses

A Wikidata jelenleg (2021. február 5-én) több mint 27,2 ezer PIM-névtér-azonosítót tartalmaz, ezek jelentős része kölcsönösen kapcsolt (oda-vissza linkelt); az olvasó egy PIM-es adatrekord „Wiki” füle alatt találja a személy Wikipédia- és Wikidata-linkjét, míg a PIM-elemek a szócikkek nemzetközi katalógusai között láthatók. A PIM-névtér-azonosítók importálása 2017 őszén egy közel kilencezres köteggel kezdődött. Ezt követően az azonosítókat a szerkesztők rendre kézzel vitték fel, majd több nagyobb adatimportálást követően 2020 szeptemberében az azonosítók száma először átlépte a 20 ezret, novem-

berben pedig a 25 ezret (lásd 3. ábra). A diagramon a Wikidatán található azonosítók számának változása követhető a tulajdonság létrehozásának dátumától (2017. május 22.). Az adatokat a forrásként feltüntetett, a Wikidata-tulajdonságok használatát naplózó oldal laptörténetéből nyertem ki egy lekérdezés segítségével, majd vizualizáltam. Eltérő számítási módszertől és definíciótól függően az azonosítók száma változhat. A szignifikáns, 2020. október 21-én történt ezres visszaesés anomáliájának oka tisztázatlan, azonban ez nem befolyásolja az adatok pontosságát.



4. ábra

A PIM Névtér látogatottságának megoszlása 2019 áprilisa és 2021 februárja között
 Forrás: Petőfi Irodalmi Múzeum/Digitális Bölcsészeti Központ (DBK)

A Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjától kapott, 2019 áprilisa (9986 db PIM-névtér-azonosító a Wikidatán) és 2021 februárja (27 097 db PIM-névtér-azonosító a Wikidatán) közötti adatok tanúsága szerint (lásd 4. ábra) a magyar Wikipédia normál, asztali változatáról (<https://hu.wikipedia.org/>) a vizsgált időszakban a PIM Névtér (<https://opac-nevter.pim.hu/>) forgalmának több mint 15%-a származott, ami így a Google-keresések és a közvetlen (nem más oldalról származó) elérések után a harmadik legjelentősebb forrás volt. Közvetlenül ezután következik a magyar Wikipédia mobil weblapja (<https://hu.m.wikipedia.org/>) körülbelül három százalékkal, megelőzve ezzel a fél százalék körül álló Nemzeti Névtérrel (<https://abcd.hu/>, ennek oka utóbbi béta állapotában keresendő). Összesítve tehát, a vizsgált időszakban a PIM forgalmának nagyjából az

ötödét tették ki a közvetlenül a magyar Wikipédiáról érkező látogatók, ami az együttműködés szempontjából – meglátásom szerint – szép sikernek mondható, különösen, hogy a PIM-névtér-azonosítók száma emellett folyamatos növekedésben van a Wikidatán (lásd 3. ábra). A magyar Wikipédia asztali és a mobil változatról érkező forgalmának jelentős különbsége részben arra vezethető vissza, hogy mobil nézetben a „nemzetközi katalógusok” sablon nem látható az olvasók számára. Ez a jelenség 2021 februárjáig változatlanul fennállt, így az olvasók a PIM-re mutató hivatkozásoknak csupán egy töredékét láthatták. 2021. február 10-én összesen 17 761, a PIM Névtérre mutató linket tartalmaztak a magyar Wikipédia szócikkei, ebből 14 694 a „nemzetközi katalógusok” sablonból, 2909 a „Pim” elnevezésű sablonból, 158 pedig közvetlen hivatkozásból származott.¹²A ko-

rábban már említett probléma miatt, miszerint a magyar Wikipédia mobil változatán (<https://hu.m.wikipedia.org/>) a „nemzetközi katalógusok” sablon nem jelenik meg az olvasók számára, ennek a 14,6 ezer linknek csupán 17,3%-a volt valóban látható mindkét módon a látogatóknak. Összefoglalva tehát, a PIM-névtér-azonosítók importálása a Wikidatába (és ezek megjelenítése a „nemzetközi katalógusok” sablonon keresztül) a korábban már említett VIAF-integrációhoz hasonlóan (Klein–Kyrios, 2013) kimutatható látogatás-szám-növekedést eredményezett a PIM számára, lévén, hogy a hivatkozások döntő többsége – és a látogatók egy része – a Wikidatából, a „nemzetközi katalógusok” sablonból származik.

A megvalósítás

Szerencsés helyzetben vagyunk, hogy a Wikidata indulása óta immár lassan kilenc év telt el, és ez idő alatt adatbázisok százainak importálása történt meg sikeresen a Wikidatába. Ebből kifolyólag bőven találunk jó példákat, kapaszkodókat és működő eszközöket arra, hogyan lenne kivitelezhető egy Nemzeti Névtér-integráció. Mivel 2020-ban a Nemzeti Névtérnek csupán a béta verziója volt elérhető, ezért a továbbiakban csak olyan megvalósítási lehetőségeket veszek számba, amelyek közvetlenül is bevonják a Wikidatát és annak közösségét.

Automatizálás

A Wikidatát bevonó automatizálás során érdemes a Nemzeti Névtérben és a Wikidatában is szereplő közös azonosítókkal dolgozni (pl. PIM, VIAF, ISNI). Vegyük példaképp a Wikidata 27 ezer PIM-névtér-azonosítóval rendelkező elemegyüttesét, majd exportáljuk ezek azonosítóját, magyar címkéjét és PIM-névtér-azonosítóját egy Excel táblába. Hasonlóképp járunk el a Nemzeti Névtér esetében is (mivel a béta verzió azonosítói még nem véglegesek, az automatizálás módjai könnyen változhatnak a jövőben).

Wikidata-azonosító	Magyar címke	PIM- névtér-azonosító
Q662084	Szentágothai János	PIM71401
Q81219	Petőfi Sándor	PIM67655
Q316556	Arany János	PIM41404

5. ábra

*Három személy Wikidata-azonosítója, címkéje, illetve PIM-névtér-azonosítója – példa
A szerző saját gyűjtése és szerkesztése*

Nemzeti Névtér-azonosító	Címke	PIM- névtér-azonosító
266842	Szentágothai János	PIM71401
281804	Petőfi Sándor	PIM67655
267858	Arany János	PIM41404

6. ábra

*Három személy Nemzeti Névtér-azonosítója, címkéje, illetve PIM-névtér-azonosítója – példa
A szerző saját gyűjtése és szerkesztése*

Ezt követően már csak pár lépésre van szükség ahhoz, hogy PIM-névtér-azonosítók szerint rendezve megkapjuk az importáláshoz szükséges Nemzeti Névtér–Wikidata párokat. A „Q662084” elemhez (Szentágothai) így a „266842”, míg a „Q81219” elemhez (Petőfi) a „281804” Nemzeti Névtér-azonosítót importálnánk.

Ugyanígy járhatunk el az ISNI vagy VIAF esetében is; mivel a Wikidata milliányi ilyen azonosítót tárol, segítségükkel jelentős mennyiségű Nemzeti Névtér-adatlapot csatolhatunk a megfelelő Wikidata-elemhez.

Az olyan rekordok esetében, ahol a párosítás nem megoldható külső azonosítók segítségével, továbbra is lehetőség van a név és a születési/halálozási dátum, illetve hely szerinti párosításra. Ez szintén jelentős mennyiségű importálást jelenthet (természetesen szigorúan ügyelve az azonos nevé, illetve a megtévesztően hasonló adatokkal rendelkező személyekre). Ami a Wikidatába való importálást illeti (tehát amikor onnét visszalinkelünk a Nemzeti Névtérre), a *QuickStatements* segédeszköz gyors és stabil keretet biztosít több ezer azonosító felvételére.¹³ Az eszköz segítségével összetett változtatások széles skálája végezhető el, így az alapvető adatimportálás mellett számos egyéb felmerülő igény kiszolgálására is alkalmas.¹⁴

Mix'n'match

Az automatikus importálásra valamilyen okból nem alkalmas (pl. hiányos adatokkal rendelkező) elemek kézi párosításához rendelkezésünkre áll a Wikidata úgynevezett *Mix'n'match* eszköze.¹⁵ Ennek segítségével felhasználóbarát és hatékony módon vihetők fel azonosítók a Wikidatába. A könnyű használhatóságot szolgálja a „játékmód”, amelyben többek között láthatjuk a párosítandó rekord adatait és azt, hogy az

adott entitásnak van-e már Wikipédia-szócikke vagy Wikidata-adatlapja. Amennyiben igen, egyetlen kattintással össze is kapcsolhatjuk őket. Amennyiben viszont nincs, ugyanitt létre is hozhatunk neki egy Wikidata-elemet.

Jelenleg több olyan „Mix’n’match game” is zajlik, ahol magyar adatbázisok párosításában lehet segíteni. Ilyen például a *Magyar Katolikus Lexikon*, az

Elektronikus Periodika Adatbázis vagy épp a Magyar Tudományos Akadémia köztestületi tára. Mindez a 7. ábrán látható módon a gyakorlatban – *Huszár Zoltán* példáján, akinek még nincs Wikipédia- vagy Wikidata-lapja. A Mix’n’match elsődleges célcsoportja azon önkéntes szerkesztők köre, akik önszorgalomból vagy csupán kedvtelésből segítenének a kézi párosításban.

HAS member ID Action ▾

ID of the data-sheet of members of the Hungarian Academy of Sciences

Huszár Zoltán 🔍

Entry	12025463
Catalog ID	23967
Catalog description	19-20. századi magyar történelem és muzeológia, a Duna-medence története
Type	human [Q5]

Enter Q number of matching item

Set Q New item N/A

Search

[Search Wikidata](#) | [Search hu.wikipedia](#) | [Google-search Wikipedias](#) | [Google-search Wikisource](#) | [Google-search Wikidata](#)

Wikidata search results

No matches found

hu.wikipedia search results

7. ábra

Képernyőkép az MTA-azonosítókat párosító Mix’n’match felületéről
 Forrás: <https://mix-n-match.toolforge.org/#/catalog/249>

Összegzés

Szerencsére sokan gondoljuk úgy, hogy egy Nemzeti Névtér–Wikidata között megvalósuló együttműködés mindkét fél számára kölcsönös előnyökkel járna (a projektcélok között szerepel a más tudásadatbázisokkal való együttműködés). Ehhez azonban fontos belátni, hogy a két oldal nem konkurens egymásnak. Mint az a Nemzeti Névtér nevében és céljaiban is egyértelműen megfogalmazódik, mindenekelőtt „nemzeti”, tehát magyar vonatkozású kíván lenni. Ezzel szemben a Wikidata nem függ nyelvtől, így a benne tárolt adatok világunk egészére, nem pedig egy-egy nemzetre fókuszálnak. A Nemzeti Névtér három névtere (földrajzi, személyi és testületi) ebből fakadóan gyaníthatóan mindig teljesebb lesz,

mint a Wikidata ezen szegmense, miközben a nem magyar vonatkozású, generikus entitások tekintetében a Wikidata jóval részletesebb. Ez azonban nem probléma, hanem inkább lehetőség. A Nemzeti Névtér jelenleg még csak béta állapotú, de folyamatosan fejlődik és bővül, a globális névterekkel – köztük a Wikidata-val – való integrációja pedig a projekt tervei között szerepel. A korábban már vizsgált nemzetközi trendek alapján az integráció előnyei közül a következőket emelném ki legfontosabbként:

- A Nemzeti Névtér látogatottsága és ismertsége jelentősen nőhet. A Nemzeti Névtér adatlapjainak megjelenítése a Wikipédia-szócikkekben nemcsak az olvasók elérésében, hanem idővel *A Nemzeti Névtér értelméről és hasznáról* című dokumentum által is említett potenciális közreműködők („kom-

petens laikusok, lokálpatrióták”) bevonásához is hozzájárulhat.

- A Nemzeti Névtér adataival (születési dátum, hely, irányítószám stb.) jelentősen bővíthetnek a Wikidata magyar vonatkozású elemei, ezáltal minden olyan internetes szolgáltatás is, amely ezekre támaszkodik (pl. Google, Siri stb.).
- Ha a Wikipédia-szócikkekben a szerkesztők egy kattintással eljuthatnak a Nemzeti Névtérhez, akkor annak információihoz is könnyebb lesz hozzáférni, azokat forrásként felhasználni, ezáltal növe ezek felhasználtsága.

Összességében úgy vélem, hogy egy Nemzeti Névtér–Wikidata–Wikipédia együttműködésben jelentős potenciál rejlik, amivel (a Wikidatát és a Wikipédiát egy szereplőnek tekintve) mindkét félnek érdemes foglalkoznia.

Jegyzetek és irodalmi hivatkozások

1. Az aktuális statisztikákért ld.: Wikidata Site Stats == Wikimedia Metrics [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2021. Hozzáférhető: <https://grafana.wikimedia.org/d/000000162/wikidata-site-stats?orgId=1> [Megtekintve: 2021.02.04.]
2. Wikidata: Introduction == Wikidata [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2012–. A lap utolsó módosítása: 2020.10.29. Hozzáférhető: <https://www.wikidata.org/wiki/Wikidata:Introduction> [Megtekintve: 2021.02.13.]
3. Szentágothai János == Wikipédia : a szabad enciklopédia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2001–. A lap utolsó módosítása: 2020.12.24. Hozzáférhető: https://hu.wikipedia.org/wiki/Szent%C3%A1gothai_J%C3%A1nos [Megtekintve: 2021.02.13.]
4. Szentágothai János == PIM Névtér [online]. Budapest: Petőfi Irodalmi Múzeum, 2021. Hozzáférhető: <https://resolver.pim.hu/auth/PIM71401> [Megtekintve: 2021.02.17.]
5. Szentágothai János VIAF adatlapja == Virtual International Authority File [online]. Dublin [Ohio]: OCLC, 2021. Hozzáférhető: <https://viaf.org/viaf/12315483/> [Megtekintve: 2021.02.13.]
6. Wikipedia: Authority control integration proposal/RFC == Wikipedia : the free encyclopedia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2001–. A lap utolsó módosítása: 2018.01.02. Hozzáférhető: https://en.wikipedia.org/wiki/Wikipedia:Authority_control_integration_proposal/RFC [Megtekintve: 2021.02.13.]
7. User: VIAFbot == Wikipedia : the free encyclopedia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2001–. A lap utolsó módosítása: 2020.05.06. Hozzáférhető: <https://en.wikipedia.org/wiki/User:VIAFbot> [Megtekintve: 2021.02.13.]
8. Wikipedian in residence == Wikipedia : the free encyclopedia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2001–. A lap utolsó módosítása: 2021.01.31. Hozzáférhető: https://en.wikipedia.org/wiki/Wikipedian_in_residence [Megtekintve: 2021.02.10.]
9. KLEIN, Maximilian – KYRIOS, Aley. VIAFbot and the Integration of Library Data on Wikipedia. == The Code4Lib Journal [online], (2013) 22. ISSN 1940-5758. Hozzáférhető: <https://journal.code4lib.org/articles/8964> [Megtekintve: 2021.02.13.]
10. A hasonló hibák gyűjtése is egy dedikált lapon történik: Wikidata:VIAF/cluster/conflating entities == Wikidata [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2012–. A lap utolsó módosítása: 2021.01.23. Hozzáférhető: https://www.wikidata.org/wiki/Wikidata:VIAF/cluster/conflating_entities [Megtekintve: 2021.02.13.]
Itt érdemes megemlíteni a Wikidata VIAF-kikötésmegszegéseit gyűjtő, automatikusan frissülő listáját, ahol jól követhetők az eszközölt javítások: Wikidata: Database reports/Constraint violations/P214 == Wikidata [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2012–. A lap utolsó módosítása: 2021.02.03. Hozzáférhető: https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P214 [Megtekintve: 2021.02.13.]
11. A Magyarországgal kapcsolatos tulajdonságokat az alábbi oldal sorolja fel: Template: Hungary properties == Wikidata [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2012–. A lap utolsó módosítása: 2021.01.20. Hozzáférhető: https://www.wikidata.org/wiki/Template:Hungary_properties [Megtekintve: 2021.02.06.]
12. A PIM-névtér-azonosítót megjelenítő, nemzetközi katalógusok sablont tartalmazó szócikkek listája: Kategória: Wikipédia-szócikkek PIM-névtér-azonosítóval == Wikipédia : a szabad enciklopédia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2003–. Hozzáférhető: https://hu.wikipedia.org/wiki/Kategória:Wikipédia-szócikkek_PIM_azonosítóval [Megtekintve: 2021.02.10.]
A Pim sablont tartalmazó szócikkek listája: A(z) „Sablon: Pim” lapra hivatkozó lapok == Wikipédia : a szabad enciklopédia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2003–. Hozzáférhető: <https://>

hu.wikipedia.org/wiki/Speciális:Mi_hivatkozik_erre/Sablon:Pim [Megtekintve: 2021.02.10]

A PIM névterére hivatkozó szócikkek listája:

Keresés reguláris kifejezéssel == Wikipédia : a szabad enciklopédia [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2003–. Hozzáférhető: <https://hu.wikipedia.org/wiki/undefined?sort=relevance&search=insource%3A%22https%3A%2F%2Fopac-nevter.pim.hu%22> és <https://hu.wikipedia.org/wiki/undefined?sort=relevance&search=insource%3A%22http%3A%2F%2Fopac-nevter.pim.hu%22> [Megtekintve: 2021.02.10]

13. QuickStatements [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2021. Hozzáférhető: <https://quickstatements.toolforge.org/> [Megtekintve: 2021.02.13.]
14. Help:QuickStatements == Wikidata [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2012–. A lap utolsó módosítása: 2021.02.10. Hozzáférhető: <https://www.wikidata.org/wiki/Help:QuickStatements> [Megtekintve: 2021.02.13.]
15. Mix'n'match == Meta [online]. San Francisco [Kalifornia]: Wikimedia Foundation, 2001–. A lap utolsó módosítása: 2020.12.14. Hozzáférhető: <https://meta.wikimedia.org/wiki/Mix%27n%27match> [Megtekintve: 2021.02.13.]

Felhasznált irodalom

- ALLISON-CASSIN, Stacy – SCOTT, Dan. Wikidata: a platform for your library's linked open data. == The Code4Lib Journal [online], (2018) 40. ISSN 1940-5758. Hozzáférhető: <https://journal.code4lib.org/articles/13424> [Megtekintve: 2021.02.13.]
- ANDERSON, B. Clifford. Introduction to Wikidata: The Wikipedia of Facts. == BENNETT, Miranda, szerk. Seventy-second Annual Conference of the American Theological Library Association. Chicago: American Theological Library Association, 2018. p. 307–309. ISSN 0066-0868
- HICKEY, Thomas B. – TOVES, Jenny A. == Managing Ambiguity In VIAF. == D-Lib Magazine [online]. 20. (2014) 7/8. ISSN 1082-9873. Hozzáférhető: <https://doi.org/10.1045%2Fjuly2014-hickey> [Megtekintve: 2021.02.12.]
- A kódexektől a magára ismerő szövegig [elektronikus dok.] == Délmagyar.hu. Feltöltve: 2019.12.02. Hozzáférhető: [https://www.delmagyar.hu/kultura/helyi-kultura/a-kodexektol-a-](https://www.delmagyar.hu/kultura/helyi-kultura/a-kodexektol-a)

magara-ismero-szovegig-4745004/ [Megtekintve: 2021.02.12.]

MANSKE, Magnus [et al.]. GeneDB and Wikidata. == Wellcome Open Research [online], 4. (2019) 114. ISSN 2398-502X. Hozzáférhető: <https://doi.org/10.12688/wellcomeopenres.15355.2> [Megtekintve: 2021.02.13.]

Nemzeti Névtér béta [elektronikus dok.] == A Nemzeti Névtér honlapja. Feltöltve: 2019.05.13. Hozzáférhető: <http://abcd.hu/2019/05/13/nemzeti-nevter-beta/> [Megtekintve: 2021.02.13.]

A Nemzeti Névtér értelméről és hasznáról [elektronikus dok.] == A Nemzeti Névtér honlapja. Feltöltve: 2019.05.13. Hozzáférhető: <http://abcd.hu/2019/05/13/a-nemzeti-nevter-ertelmeroles-hasznarol/> [Megtekintve: 2021.02.13.]

PELLISSIER TANON, Thomas [et al.]. From Freebase to Wikidata: The Great Migration. == WWW ,16: Proceedings of the 25th International Conference on World Wide Web. Genf: International World Wide Web Conferences Steering Committee, 2016. p. 1419–1428. ISBN 978-1-4503-4143-1

PINTSCHER, Lydia [et al.]. Strategy for Wikidata as a platform [elektronikus dok.], [2019. augusztus] Hozzáférhető: https://upload.wikimedia.org/wikipedia/commons/d/d3/Strategy_for_Wikidata_as_a_platform.pdf [Megtekintve: 2021.02.13.]

TURKI, Houcemeddine – SHAFEE, Thomas – HADJ TAIEB, Mohamed Ali et al. Wikidata: A large-scale collaborative ontological medical database. == Journal of Biomedical Informatics, (2019) 99. ISSN 1532-0464. Hozzáférhető: <https://doi.org/10.1016/j.jbi.2019.103292> [Megtekintve: 2021.02.13.]

VAN VEEN, Theo. Wikidata: Wikidata: From „an” Identifier to „the” Identifier. == Information Technology and Libraries, 38. (2019) 2., 72–81. ISSN 2163-5226. Hozzáférhető: <https://doi.org/10.6017/ital.v38i2.10886> [Megtekintve: 2021.02.13.]

VIAFbot Edits 250,000 Wikipedia Articles to Reciprocate All Links from VIAF into Wikipedia [elektronikus dok.] == Az OCLC honlapja. Feltöltve: 2012.12.07. Archivált változat hozzáférhető: <https://web.archive.org/web/20130609200557/http://www.oclc.org/research/news/2012/12-07a.html> [Megtekintve: 2021.02.13.]

Beérkezett: 2021. február 13.