

GYIMESI RÉKA

Különböző nominális források összekapcsolásának módszertani lépései

Bonyhádi esettanulmány a 19. század második feléből

Methodological steps of record linkage – case study from the 19th century Bonyhád

The study aims to introduce the steps of creating a database based on historical nominal sources. Two individual level census materials (from 1857 and 1869) and an online genealogy based on vital registers were linked in case of Bonyhád. While the censuses provide a particularized but static picture of the population, creating a database with linking them to the genealogical records offers a dynamic analysis on a broader spectrum.

Keywords: record linkage, census, genealogy, cadastral map



A tanulmány célja, hogy bemutassa azokat a lépéseket, melyek egy történeti forrásokból származó adatokat összekapcsoló adatbázis létrehozásához szükségesek. Ennek oka, hogy az alapvetően is kisszámú magyar történeti demográfiai irodalom elenyésző része foglalkozik az elemzések alapjául szolgáló adatállományokkal és az azokkal való módszertani munkával, ami nélkül nem lenne lehetőség az elemzések elkészítésére. Jelen esetben a Bonyhádról fennmaradt, két időpontból származó (1857 és 1869) népszámlálási íveket és egy online elérhető genealógiából származó anyakönyvi adatokat kapcsoltunk össze.¹ Míg a népszámlálások egy-egy évből adatgazdag, ám statikus képet nyújtanak a bonyhádi lakosságról, addig a kapcsolódó évtizedek anyakönyvi adataival való összekötés dinamizálja a rendelkezésre álló információk halmazát és az elemzések jóval szélesebb körét teszi lehetővé.

¹ <https://online-ofb.de/bonyhad/?lang=de>. Letöltve: 2022.05.26.

A források bemutatása: a népszámlálások

Magyarországon az első teljeskörű népszámlálást II. József uralkodásának idején, 1784–1787-ben hajtották végre. Annak ellenére, hogy az uralkodó törekvése az ilyen típusú összeírások rendszeressé tétele volt, a következő népszámlálás végrehajtásáig több évtized telt el. Az 1848–1849-es eseményeket követően kialakított centralizált birodalmi rendszerben került sor a következő két összeírásra, melyet Magyarországon is az osztrák hatóságok hajtottak végre.² Az első 1850-ben vette kezdetét, de befejezni csak a következő évben tudták. Az elhúzódozó összeírás, az eszmei időpont hiánya és más módszertani problémák miatt pontatlan eredmények születtek, a kutatás szerint nagyjából 5–6%-osra becsülhető az alulszámlálás mértéke.³ A következő népszámlálást 1857-ben tartották, melynek célja többek között az előző hibáinak kiküszöbölése volt. Ez volt az első olyan összeírás, amikor alkalmazták az eszmei időpontot,⁴ ezzel is arra törekedve, hogy pontosabb eredményeket kapjanak.⁵ E két népszámlálás eredményeit feldolgozó kötet így fogalmaz: *„Az 1850. évi népszámlálás utolsó tagja annak az évszázados törekvésnek, amely az Osztrák Birodalomban és Magyarországon is »feudális-katonai jellegű népösszeírások« címszóval foglalható össze. Az 1857. évi népszámlálás pedig az első tagja a modern, első sorban demográfiai szempontokat szem előtt tartó népszámlálások sorozatának.»*⁶

Bonyhád településről már az 1850–51-es népszámlálásból⁷ is maradtak fenn iratok, de ez a forrásanyag rendkívül töredékes, így a tervezett elemzéshez nem mutatkozott alkalmasnak. Az 1857-es népszámlálás meglévő anyaga⁸ sem teljes, de a lakosság körülbelül kétharmad részét lefedi, ezért ez képezi a vizsgálat kiindulópontját. A felvételi íveken feltüntették az országot, a megyét, a községet, az adott ház birtokosának nevét, a házszámot, az ingatlanban élők nevét és a születési évét. A többi adatot előre meghatározott rovatokba behúzott strigulákkal vették fel, ilyen volt az egyének vallása, foglalkozása, különböző korcsoportokba sorolása,⁹ a családi állapotuk, valamint, hogy jelen vagy távol voltak-e az összeíráskor. Ez

² Nagy 2008. 102.

³ Lencsés 2011. 751.

⁴ Az 1857-es népszámlálás idején az 1857. október 31-i állapotokat írtak össze. Az eszmei időpont alkalmazásának a célja az alulszámlálás elkerülése, a pontos, egyidejű adatok rögzítése.

⁵ Óri – Pakot 2011. 12–13.

⁶ Dányi 1993. 5.

⁷ TML IV. B. 152./b

⁸ TML IV. B. 152./b

⁹ Az 1857-es népszámlálás felvételi ívein a férfiak kormegoszlására részletesebben kérdeztek, mint a nőkére. Míg utóbbiak esetében 7 korcsoportot alakítottak ki, addig a férfiaknál 16 rovat szerepelt, külön kiemelve a 14–21 éves kor közöttieket, ahol minden egyes életév külön oszlopot kapott, hogy részletes adatot gyűjthessenek a (jövőbeli) katonakorúakról.

a forrásanyag 4108 bonyhádíról tartalmazza a fenti adatokat, a hiányzó ívek miatt a 173 és a 352 közötti házszámok alatt élőkről nincsenek információk.

A következő népszámlálást a kiegyezés évében, a Keleti Károly¹⁰ vezetésével létrejövő Statisztikai Szakosztály kezdte el megszervezni. Ebből kifolyólag a végrehajtásban részt vevők köre, az összeírás módja és szemlélete is különbözött az előző két összeírástól.¹¹ Emellett a népszámlálások megszervezése során egyre inkább érvényesült az 1850-es évek második felétől rendszeresen megrendezésre kerülő nemzetközi statisztikai kongresszusokon megfogalmazott tudományos céloknak való megfelelés.¹² A népszámlálást végül 1870-ben bonyolították le, 1869. december 31-i eszmei időponttal. Az összeírás végrehajtását erre kialakított bizottságok végezték, melyek munkáját a helyi közigazgatási szervek is kötelesek voltak segíteni. A Statisztikai Szakosztály az elejétől a végéig felügyelte a műveletet, hiszen az eljárásban részt vevő közegek még gyakorlatlanok voltak, valamint a feldolgozást is nagyban segítette, hogy a hivatal munkatársai egészében átlátták a folyamatot. Az 1869-es népszámlálás során már a tényleges népesség¹³ összeírása mellett döntöttek, valamint törekedtek arra is, hogy ahol lehet, a lakosság adatait önszámlálással gyűjtsék be, de ez végül csak Budapesten valósulhatott meg.¹⁴

Az 1869-es népszámlálás anyaga¹⁵ Bonyhád teljes lakosságáról fennmaradt, vagyis 6036 főről tartalmazza az alábbi adatokat. Összeírták a háztartásokban élők nevét, nemét, születési évét, vallását, családi állapotát, foglalkozását és munkaviszonyának típusát,¹⁶ születési helyét. Megadták, hogy helybeli-e vagy idegen,

¹⁰ Keleti Károlyról bővebben: Rózsa 2014. 363–366.

¹¹ Óri–Pakot 2011. 14.

¹² <https://www.isi-web.org/index.php/about-isi/what-is-isi/history> Letöltve: 2020. 09. 15.

¹³ 1850-ben és 1857-ben a jogi népesség összeírására törekedtek, részben ez okozta az alulszámlálást is, hiszen ebben az időszakban a honosság kérdése még tisztázatlan volt. A jogi népesség viszont magában foglalta a jelen nem lévő lakosokat is, fogalmát azonban nem határozták meg pontosan, amely komoly problémákat okozott az összeíráskor. Dányi 1993. 31. Az 1869-es népszámlálást követően nyújtottak be törvényjavaslatot a magyar állampolgárság megszerzését szabályozandó. Népszámlálás 1870, 4. Ez lett az alapja az 1879. évi L. törvénycikknek, „A magyar állampolgárság megszerzéséről és elvesztéséről” <http://1000ev.hu/index.php?a=3¶m=5880> Letöltve: 2020. 06. 23.

¹⁴ Thirring 1983. 5–8.; Népszámlálás 1870, 1–4.

¹⁵ TML V. 709./c

¹⁶ A népszámlálási felvételi íveken szereplő leírás szerint: „Munka- vagy szolgálati viszonya: e rovatban ki kell tüntetni vajjon az illető a kijelölt foglalkozásnál önállóan mint vállalkozó, vagy mint segéd működik-e? Vajjon mint tulajdonos, haszonbérlet, évi vagy havi bér vagy napszám szerint dolgozik-e gazdaságban, gyárban, stb.? Mesterlegény, inas-e valamely mesterségnél? Birtokos, segéd, könyvvezető, inas-e valamely üzletben? Szolgálatban van-e háztartásnál, stb.?” – TML V. 709./c

jelen van-e vagy távol, illetve tud-e írni-olvasni.¹⁷ A sort egy megjegyzés rovat zárta, ahová bejegyezheték, ha valaki a katonaság kötelékéhez tartozott, ha távol van, akkor hol tartózkodik, de ide jegyezték fel azt is, ha valaki testi vagy szellemi fogyatékossgal élt. Ezen kívül a felvételi ív tartalmaz egy lakásstatistikai összesítést is, ahol az adott házszám alatti épületről adták meg, hogy hány darab és milyen típusú helyiségből áll, van-e emelete, félemelete és milyen melléképületek tartoznak hozzá. Az utolsó oldalon pedig a háziállatokat vették számba.¹⁸

A fentiekből adódóan a népszámlálások felhasználása során egyrészt tekintettel kell lenni arra, hogy a két összeírás más-más szempontok szerint készült, és minden esetben fontos a forráskritika alkalmazása mellett a különböző keletkezéstörténeteket is figyelembe venni. Emellett ügyelni kell arra is, hogy habár vannak azonos, vagy nagyon hasonló kérdések, de ezek mögött a hasonlóságuk ellenére is különböző fogalomkészlet rejlik, igaz ez például az életkorra,¹⁹ a foglalkozásra²⁰ vagy a honosságra vonatkozóan is.

¹⁷ Az 1869-es népszámlálás az első, melyben felmérték a lakosság írni-olvasni tudásának szintjét. Az alfabetizációt a 6 éven felüli népesség körében vizsgálták, ezt azonban a következő, 1880-as összeírásnál megváltoztatták. Ekkor egy évvel kitolták a korhatárt és a 7 éven felüli lakosságot vették alapul, egyrészt, hogy a külföldi adatokkal összehasonlítható eredményeket kapjanak, másrészt, mert ugyan az iskolakötelezettség 1868-as törvényben is rögzített 6 éves kortól kezdődött, de az első évben még sok gyermek nem tudott írni-olvasni, így ez torzíthatta az 1869-es eredményeket. Népszámlálás 1880. 226.

¹⁸ A tervezett, de ekkor még meg nem valósult mezőgazdasági összeírás helyett mérték így fel a háztartásokban tartott háziállatokat. https://www.ksh.hu/mult_kronologia Letöltve: 2020. 09. 15.

¹⁹ Gondolhatunk itt például az előzetesen meghatározott korcsoportok különböző kialakítására a népszámlálásokban, vagy a két népszámlálás különböző eszmei időpontjából (1857-ben október 31., míg 1869-ben december 31.) fakadó eltérő számításokra. Ez utóbbi az aggregált adatok használatánál okozhat problémákat, jelen elemzés azonban az népszámlálási ívekre támaszkodik és onnan nyeri az életkori adatokat. Dányi Dezső így írt az életkor kapcsán az 1850-es évek népszámlálásait feldolgozó kötetben: „*sajátos, egymással és a következő népszámlálásokkal is nehezen összevethető életkor-, illetve korbevallási statisztikát alakított ki.*” – Dányi 1993. 43.

²⁰ Egyrészt az 1857-es népszámlálásban előre meghatározott rovatokban tüntethették fel a foglalkozásokat, mely nem alkalmas a pontos foglalkozásszerkezet rekonstruálására. Az 1869-es népszámlálásban pedig a számlálók írták be az adott személyek hivatását és foglalkozási viszonyát. A két népszámlálás foglalkozásra vonatkozó adatai így egymással közvetlenül nem összehasonlíthatók. Erről a forráskritikai kérdéséről részletesen: Pozsgai 2000. 170–172. Az 1869-es népszámlálás eredményeiben már az összefoglaló kötet elkészítésekor hibák tűntek fel. Az összeírás során ugyanis a földművelésből élők egy része a személyi szolgálatot teljesítők kategóriájába került. Ennél még jelentősebb torzításhoz vezetett az, hogy a keresők feleségei, gyermekei is ebbe a csoportba kerültek, ahelyett, hogy a háztartásfő foglalkozása szerinti besorolást kaptak volna, mint eltartottak. Mindez az adatok kritikával való kezelésére hívja fel a figyelmet, valamint arra, hogy a földművelésből élőknek nagyobb a foglalkozás szerinti megoszlásban való aránya, mint ahogy az eredmények ezt tükrözik. Országos szinten ez közel 10%-os eltérést jelentett, az összeírás szerint 68,4%, míg valójában 75–78% volt a földművelésből élők aránya 1869-ben Magyarországon. Thirring 1983. 19–20.; Népszámlálás 1870. 264.

Az anyakönyvi adatok – online genealógia

A bonyhádi lakosságról összegyűjtött népszámlálásokban elérhető sokszínű adategyüttest az egyes felekezetek által vezetett anyakönyvekben rögzített életeseményekkel lehet dinamizálni. A kereszteleseket, a házasságkötéseket és a halálozásokat tartalmazó forrásoknak az adott időpontban a településen élők adataival való összekapcsolása olyan információhalmazzá hoz létre, mely rendkívül sokrétű elemzést tesz lehetővé. Az anyakönyvi adatok felhasználása egy online elérhető genealógia segítségével történt, mely már az egyes egyénekhez kapcsolva tartalmazza az életeseményeket – természetesen abban az esetben, ha ezek az információk elérhetők az adott személyről.

Az interneten mindenki számára hozzáférhető genealógia a német *Verein für Computergenealogie* által gondozott projekt része, mely célja, hogy családtörténeti kutatásokhoz alkalmas forrásokat gyűjtsön össze és adjon közre.²¹ A weboldal számos német tartomány településein túl más országokból származó adatokat is tartalmaz, így lengyel, román, szlovák és cseh városok mellett magyar helységekből származó anyakönyvek alapján is készítettek genealógiai adatbázist.²² A bonyhádi *Ortsfamilienbuch* Tarnai Péter kutatásaira alapozva készült el.²³

Egy teljesen kitöltött genealógiai személyi adatlap²⁴ tartalmazza az egyén nevét, születési és halálozási dátumát, felekezetét, foglalkozását, keresztszülei nevét, valamint a lakóhelyére vonatkozó adatokat különböző években. Ezt követően feltüntetik a családját, férjét vagy feleségét – ha több volt, akkor mindet egymás alatt – a házasságkötés időpontjával, valamint az ő születési és halálozási dátumát, emellett a házasságokból született gyerekeik adatait is. Ezt követi a származási család megjelölése, a szülők és a testvérek nevei és születési-halálozási adatai. A legtöbb esetben azonban nem érhető el minden információ egy adott emberről, így vannak kevésbé vagy szinte alig kitöltött adatlapok²⁵ is, amikor csak egyetlen forrás tesz említést valakiről, de a családtagjait például nem lehetett azonosítani.

²¹ <https://www.compgen.de/> Letöltve: 2022. 12. 15.

²² <https://ofb.genealogy.net/> Letöltve: 2022. 12. 15.

²³ <https://online-ofb.de/famlist.php?ofb=bonyhad> Letöltve: 2022. 12. 15.

²⁴ Ilyen például Jakob Schmidt adatlapja, melyen szerepel a felesége és a házasságukból született nyolc gyermekük. Hozzá kapcsolták a szüleit is, valamint megtalálható az oldalán nyolc testvére is. Ezeken felül több más kiegészítő adat is szerepel az adatlapon, a felekezete, foglalkozása, lakcíme, hogy kik voltak a tanúk a házasságkötéskor stb. <https://online-ofb.de/famreport.php?ofb=bonyhad&ID=I72494&nachname=Schmidt&lang=de> Letöltve: 2023.03.12.

²⁵ Példának hozhatjuk Adam Schmidt adatlapját, melyen csak a felesége neve és a házasságkötésük dátuma szerepel, más személyt vagy adatot nem tudtak az anyakönyvek alapján hozzá kötni. <https://online-ofb.de/famreport.php?ofb=bonyhad&ID=I49582&nachname=Schmidt&modus=&lang=de> Letöltve: 2022.05.26.

Ilyen különböző kitöltöttségű személyi adatlapokból összesen több mint 90.000 áll a kutatók rendelkezésére az *Ortsfamilienbuch Bonyhád* oldalon. Az online genealógia 1729 és 1984 közötti római katolikus, evangélikus, református és kisebb részben izraelita anyakönyvi adatokat tartalmaz Bonyhádról és a környező településekről, tanyákról.²⁶

Az adatbázis elkészítésekor a felhasznált genealógia még bővítés alatt állt, annak készítője folyamatosan tölti fel az anyakönyvek feldolgozása során nyert új adatokat. Az utóbbi időben főként a zsidókról születtek új bejegyzések, a római katolikus és a protestáns anyakönyvekből származó adatok köre már csak minimálisan bővült, ezek feldolgozása jóval előrehaladottabb volt, mint az izraelita forrásoké. Mivel közel százezer egyén különböző életeseményei kerültek rögzítésre több mint két évszázadnyi időtartamot felölelve, így elkerülhetetlen, hogy a genealógia is tartalmazzon forráskritikára szoruló elemeket. Többször előfordult, hogy egy egyén kétszer szerepelt, egyszer a származási családjával (szülők, testvérek) összekötve, egy külön adatlapon pedig a feleségével és gyerekeivel. Ez abban az esetben vált bizonyíthatóvá a népszámlálási adatokból, ha az adott egyén például a szülei háztartásában maradt megházasodása után, így egyértelműen látszik, hogy mind a származási, mind a prokreációs családja hozzátartozik. Egy másik tipikus probléma – ami szinte elkerülhetetlen az anyakönyvek használatából és annak keletkezési sajátosságaiból fakadóan –, hogy ha egy már többgyermekes család költözött a településre máshonnan, akkor nagy esély volt arra, hogy ezek a máshol született gyermekek nem szerepelnek a genealógiában, hiszen máshol lettek megkeresztelve. Akár teljesen kimaradhattak az adatállományból, ha a család néhány év múlva tovább költözött, anélkül, hogy a gyermekek életében bármilyen anyakönyvezésre kerülő esemény történt volna (meghaltak vagy megházasodtak). Bekerülhettek azonban, ha a településen maradtak, és például megházasodtak. Ebben az esetben a házassági anyakönyvből is pótolhatók a születési adatok, és ha meg vannak jelölve a házasulandók szülei is, akkor jó eséllyel keletkezett egy adatlapjuk a genealógiában.

Mindez felhívja a figyelmet egy olyan problémára, amelyet sem a népszámlálások, sem az anyakönyvek felhasználásával nem tudunk orvosolni: a vándorlás ekkor még nem igazán létező adminisztrációjára. Ebben az időben ugyanis nem rögzítették, ha valaki elhagyta a települést, ezt csak más források segítségével – például rendszeresen vezetett lélekösszeírások alapján – lehet rekonstruálni, ami Bonyhád esetében nem áll a rendelkezésünkre. Ebből fakadóan az elemzés a meglévő források felhasználásával próbál következtetéseket levonni a vándorlásra vonatkozóan. Ez az oka annak, hogy míg az 1857-es népszámlálás az elemzés kiindulópontjaként szolgál, addig az 1869-es népszámlálást egyfajta lezárásként

²⁶ <https://online-ofb.de/bonyhad/?lang=hu> Letöltve: 2022.05.26.

használjuk az elemzésben. A nyomon követett egyénekkal, akikről megfelelő adatokkal rendelkezünk – tehát megtaláltuk őket a genealógiában is és a népszámlálásokban is – elméletileg három dolog történhetett a két időpont között: meghaltak (a genealógiában szerepel a halálozásuk dátuma), a településen éltek (szerepelnek az 1869-es népszámlálásban), ha pedig egyik sem, akkor feltételezhetjük, hogy elvándoroltak.

A források összekapcsolásának módszere

Ahhoz, hogy a rendelkezésre álló adatokat használni, elemezni lehessen, a forrásokat valamilyen módon össze kellett kapcsolni egymással.²⁷ A két vizsgált népszámlálás Tolna Vármegyei Levéltárban megtalálható anyaga az eredeti struktúra megőrzésével került digitalizálásra, míg egy harmadik adatállományt a genealógiából származó anyakönyvi adatok képeztek. A források ismertetéséből már kiderült, hogy a két népszámlálás rovatai nem teljesen azonosak, ráadásul az összeíró szervek, a nyelv²⁸ és a kérdések mögött rejlő fogalomkészlet is más-más jellegű a két összeírásnál. Az online genealógiának a népszámlálásokhoz képest teljesen más szerkezete van, miközben mindhárom forrásban potenciálisan szerepelhetnek ugyanazok az egyének. A kutatás egyik legnagyobb feladata ezeknek az embereknek az azonosítása volt. Ennek a folyamatnak az első lépése az adatok megtisztítása, az adatállományok rendezése volt, hiszen amellet, hogy a történész feladata az eredeti forrás megőrzése, jelen esetben a további elemzéshez alkalmas adatállományokat kellett létrehozni. A változtatások – az eredeti verziók megőrzése mellett – az ún. R programnyelvben²⁹ történtek, mely használhatónak bizonyult az adatbázis elkészítésének további lépéseire és annak elemzéséhez is.

Az adatok összekapcsolásához szükség volt mindhárom adatállományban egyéni azonosítóval (ID) ellátni a személyeket. Az online genealógiában már rendelkeztek egy ilyen ID-val az egyének, míg az 1857-es és 1869-es népszámlálásban

²⁷ Ezt a folyamatot, a személyek különböző forrásokban való azonosítását és összekapcsolását, angolul *record linkage*-nek nevezik és bőséges nemzetközi szakirodalom áll rendelkezésre erről a módszerről. A több évtizedes múltra visszatekintő családrekonstrukció módszere is ezen alapul (Henry, Wrigley, Schofield – Magyarországon: Andorka 1987). A teljesség igénye nélkül néhány közelmúltban ezt a módszert alkalmazó cikk: Richards et al. 2014; Feigenbaum 2016. Magyarul: Andorka 1988; Benda 2006; Pakot 2013. Az elemzésben felhasznált források összekapcsolásának problémájáról részletesen, példákkal: Gyimesi 2019.

²⁸ Az 1857-es népszámlálás német nyelven, míg az 1869-es magyar nyelven készült.

²⁹ <https://www.r-project.org> Letöltve: 2022. 01. 27. Az R egy statisztikai számításokhoz és adatok vizualizációjához használható szoftverkönyvtár, mely bárki számára ingyenesen hozzáférhető.

szükség volt ezek létrehozására. A következő fontos lépés annak eldöntése volt, hogy mely változókat lehet, illetve érdemes felhasználni arra, hogy potenciális találatokat kapjunk egy adott forrásban szereplő egyénhez. Olyan mindhárom adattállományban szereplő ismérvekre volt szükség, melyek alkalmasak az egyének minél pontosabb azonosítására.

Egyértelműnek tűnik, hogy ilyen adat lehet az *egyén neve*. Ez a rovat képezte a keresések alapját, de ennek kapcsán meg kell jegyezni, hogy több probléma is felmerült a név mint azonosító használatát illetően. Egyrészt a források természetükből fakadóan is tartalmazhatnak hibákat, elírásokat, amelyek száma a digitalizálás során akaratlanul vétett elütésekkel is gyarapodhatott. Másrészt a források eltérő nyelvének köszönhetően – az egyik népszámlálásban a német, míg a másokban a magyar nyelvet használták, az anyakönyvekben pedig mindezek mellett a latin is gyakori – ugyanannak a névnek többféle variációjával is találkozhatunk.³⁰ Igaz lehet ez a vezetéknevek³¹ és a keresztnévek³² esetében is. Mivel több tízezer adatról van szó, így a nevek alapján történő keresés automatizálásra került, melyhez a program karaktértávolságot számított, így minden egyes elütés vagy írásmódbeli különbség jelentőséggel bírt. Az ilyen jellegű különbségek mellett az azonos írásmódban szereplő nevek is okozhattak problémát, hiszen voltak rendkívül gyakran előforduló nevek: a genealógiában például több mint 180 Müller Johann és 100-nál is több Schmidt Johann szerepel.

A következő szóba jöhető ismérv, mely a név alapján kapott találatok pontosításához – a forrástól függően eltérő mértékben – használható az *egyén életkora* vagy *születési éve*. Ennek kapcsán azonban számolni kell a korszakban még nagyon magas analfabetizmusból fakadó azon problémával, hogy a népszámlálásban sokszor nem jól szerepel ez az információ, hiszen az egyén saját maga sem volt tisztában pontos életkorával. Ezt jól mutatja az 1869-es népszámlálás adataiból készült országos szintű korfa is, melyen a kerek életkoroknál tapasztalható kiugró értékek utalnak a pontatlan korbevallásra.³³ Az életkor tekintetében az anyakönyvi bejegyzések szolgálták megbízhatóbb adatokkal, hiszen az abban szereplő keresztelés időpontját feltételezhetően az eseménnyel közel egyidőben

³⁰ Az egyének dolgozatban előforduló névhasználata kapcsán fontos leszögezni, hogy az általunk használt névváltozat nem kíván reflektálni a személyek etnicitására. Egyrészt mivel egyik felhasznált forrás sem tartalmazza az egyének nemzetiségét, másrészt pedig nincsenek olyan kiegészítő forrásaink, melyek erre a komplex kérdésre vonatkozóan információval szolgálnának.

³¹ Például: Seib – Szeib – Seip – Szaib, Letscher – Letser, stb.

³² Ugyanaz a személy a forrás keletkeztetőjétől függően szerepelhet eltérő névváltozatokkal, pl.: Erzsébet – Elisabeth(a) – Örzse, Franz – Ferenc, Eugen – Jenő, Adalbert – Béla stb.

³³ Magyarország népességének száma nemek és életkor szerint – 1870-es év. <https://www.ksh.hu/interaktiv/korfak/orszag.html> – Letöltve: 2022. 05. 26.

jegyezték fel.³⁴ Sok esetben azonban a születés időpontját nem a keresztelési anyakönyvből vitték fel a genealógiába, hanem más forrásból (házasságkötés, halálozás) számolták vissza, ahol megadták az egyén életkorát. Az ilyen módon kiszámított születési év esetében külön jelzés szerepel a genealógiában,³⁵ itt is szükséges a forráskritika alkalmazása.

Emellett az *egyének vallása* is szerepel mindegyik forrásban, így ez is felhasználható volt a minél pontosabb találatok és az azonosítás érdekében. Ebben az esetben is szükség volt az adatok előzetes tisztítására és egységesítésére, hogy mindhárom forrásban ugyanolyan elnevezéssel szerepeljen az adott felekezet.³⁶ A római katolikus és a zsidó anyakönyvi bejegyzésekben minimális volt a pontatlanság, mivel ez a két felekezet a vizsgált korszakban saját egyházi adminisztrációval rendelkezett a településen. Az evangélikus és a református anyakönyvezés viszont nem vált ennyire külön, így többször találkozhatunk egymásnak ellentmondó bejegyzésekkel a népszámlálásokban és a genealógiában (valaki az egyik típusú forrásban evangélikusként, míg a másikban reformátusként szerepel).³⁷ Mivel a népszámlálás önbevallásra épülő adatokon alapul, a felekezet tekintetében valószínűleg az ezekben szereplő adatok tekinthetők pontosabbnak, ellentétben az életkori adatokkal, ahol ez éppen megbízhatatlanabb eredményre járt.³⁸

Az azonosításhoz használt változók kiválasztását követően létre kellett hozni azokat a *szabályokat*, melyek alapján a program megalkot egy listát arról, hogy egy adott személyhez potenciálisan kiket lehetne hozzákapcsolni a másik két forrásból. A népszámlálások minden egyes egyénéhez megkerestük azokat a genealógiában szereplő személyeket, akik megfeleltek az alábbi kritériumoknak. 1) Az 1857-es népszámlálás kapcsán kizártuk azokat a genealógiában talált személyeket, akiknek a halálozási dátuma 1857 előtti, vagyis a népszámlálás időpontjában nem lehettek a településen. Ugyanezt megtettük az 1869-es népszámlálás esetében is a megfelelő

³⁴ A korban magas csecsemő- és gyermekhalandóság miatt törekedtek a mielőbbi megkeresztelésre.

³⁵ Pl. *berechnet* szó a dátum előtt. Még pontatlanabb lehet az adat, ha valamihez viszonyítva adják csak meg a dátumot (pl.: nem tudják a férj halálozási dátumát, de a feleség halálakor a nő már özvegy, így a férj halálozási dátumánál annyi szerepel, hogy *vor [a feleség halálozási éve]*, de azt nem tudni, hogy ez pontosan mikor történt.)

³⁶ Az egységesítést követően 5 különböző elnevezés szerepelt a forrásokban: római katolikus, református, evangélikus, zsidó és görögkatolikus. A népszámlálások digitalizása során már azonos módon került rögzítésre a felekezeti hovatartozás, míg a genealógiában használt rövidítéseket (pl.: rk [= római katolikus], ev.ref., calv, helv [= református], stb.) fel kellett oldani.

³⁷ A szakirodalom szerint a reformátusok a hidasi gyülekezethez tartoztak ekkor (Tolna megye kézikönyve, 196.)

³⁸ Nem zárhatjuk ki azt a lehetőséget sem, hogy egyik forrás sem téved. Előfordulhat, hogy születésekor még valakit egy bizonyos vallás szerint kereszteltek meg, de a házasságkötéséig, vagy éppen akkor áttért, így a népszámlálás során valóban lehetett már más felekezet tagja.

halálozási évet alkalmazva. 2) A nemnek és a vallásnak³⁹ mindkét forrásban meg kellett egyeznie. 3) A forrásokban szereplő születési évek között a fent jelzett pontatlanságokat figyelembe véve +/- 5 évnyi távolságot engedélyeztünk. Ezeknek a kritériumoknak a meghatározásával sok olyan egyént ki lehetett zárni, akik biztosan nem lehettek azonosak a népszámlálásban szereplő személlyel. Hozzá kell azonban tenni azt is, hogy mivel a genealógiában is sok olyan egyén van, akinél hiányoznak bizonyos adatok, így ha a szűréshez használt adat hiányáról volt szó, akkor nem hagytuk ki az adott egyént automatikusan a listából, hiszen egy információ hiánya alapján nem lehetett egyértelműen kizárni az azonosságot.

A szűkítést követően, a kritériumoknak megfelelt személyek között történt a következő – a nevek közötti karaktertávolságon alapuló – keresés. Ennek során nagyobb hangsúlyt kapott a vezetéknevben mutatkozó azonosság, mint a keresztnév kapcsán fennálló egyezés. Ennek oka egyrészt az volt, hogy a források viszonylag kisszámú, de gyakran előforduló keresztnévet tartalmaznak.⁴⁰ Emellett a keresztnevekre megalkottunk egy „szótárat” is,⁴¹ ami alapján a program érzékelte, hogy egy adott név becézett formája (pl.: Erzsébet – Örzse – Ense) vagy más nyelvű változata (pl.: Elisabeth[a]) ugyanannak a névnek minősül, így a túl nagy karaktertávolságok miatt potenciálisan kieső keresztnévváltozatokat nem vesztettük el.

A név alapján történő azonosításhoz még egy, a forrásadottságokból származó problémát orvosolni kellett, ez pedig a nők vezetéknevének megváltozása a házasságkötést követően. A genealógiában értelemszerűen minden nő a születési nevével szerepelt, azonban, ha a népszámlálás időpontjában már férjhez mentek, akkor a férjük vezetéknevét viselték, a leánykori nevüket pedig nem jegyezték fel az összeírásokban. Ezen felül megnehezítette a házas nők azonosítását, hogy a házasságkötési adatok nélkül nem lehetett tudni, hogy melyik népszámlálásban milyen vezetéknevvel kell keresni őket. Ha hajadonok voltak, akkor a születési nevük szerepelt az összeírásban, de megházasodhattak akár már az első népszámlálás előtt, akár a kettő között, sőt, megözvegyülhettek (jellemzően ekkor is a volt férjük vezetéknevét viselték) vagy újránházasodhattak, így akár a két népszámlálásban két különböző – a lánykori nevüktől eltérő – vezetéknevvel is szerepelhettek. Ennek orvoslására létrehoztunk egy új változót mindkét népszámlálás adatállományában, mely figyelembe vette a genealógiában rögzített házasságkötéseket, és ha bármelyik

³⁹ A felekezeti azonosság mint kritérium, néhány esetben a valós találat kizárásával járt a korábban említett református – evangélikus bejegyzések különböző forrásokban való eltérése miatt, ezeket az eseteket manuális ellenőrzés során javítottuk.

⁴⁰ A genealógiában 11601 Johann, 6306 Katharina, 5423 Elisabeth és 5414 Anna szerepel. Ezek a leggyakrabban előforduló keresztnévek, igaz ezekhez még hozzáadódik az anyakönyvek különböző nyelvhasználata miatt 650 János, 56 Katalin és 1002 Erzsébet is.

⁴¹ A forrásokban előforduló névváltozatokat egymáshoz rendelő táblázat, mely több mint 80 keresztnév minimum 2, maximum 6 különböző formáját tartalmazza.

népszámlálást megelőzően történt házasságkötés, akkor a férj vezetéknevét rögzítettük ebben az új oszlopban. Ennek köszönhetően a népszámlálások idején házas vagy özvegy nők esetében a nevek közötti karaktertávolság kiszámítása a férjük vezetéknevének használatával történt.⁴²

A program tehát készített egy listát minden egyes népszámlálásban szereplő egyénhez, melyben az szerepelt, hogy ki(k) az(ok) a genealógiában, aki(k) a nevük, a születési évük, a vallásuk és a nemük alapján leginkább hasonlítanak az adott személyhez. A nem szerinti szűrés használatára azért volt szükség, mert vannak olyan férfi–női névpárok, amik nagyon hasonlóak egymáshoz, esetleg csak 1–2 karakternyi távolság van közöttük (pl.: Josef – Josefa), amiket így ki lehetett zárni a találati listáról.

A legtöbb esetben ez a lista nem egyelemű, vagyis a már ismertett forrásokból származó pontatlanságok, bizonytalanságok és hiányosságok miatt a legjobban megalkotott szabályrendszerrel sem születhetne meg egy teljesen automatizált összekapcsolás. Így szükség volt beiktatni a munkába egy manuális szakaszt, mikor a program által előállított listáról az eddig használt ismérveken túlmutató információk használatával minden kétséget kizáróan megállapítást nyert a különböző forrásokban szereplő egyének közötti azonosság. Ennek elvégzéséhez egy webes felület került létrehozásra, mely lehetővé tette a kutatói döntés meghozatala után az egyének összekapcsolását.

Ebbe az adatbáziskezelő rendszerbe került beolvasásra mindhárom adatállomány (a két népszámlálás és a genealógia) és került meghatározásra az, hogy melyik forrás szolgál a keresés alapjául, és melyekben keressük a lehetséges találatokat. Először az 1857-es, majd az 1869-es népszámlálásban szereplő egyéneket kerestük a genealógiában, ezen keresztül a két népszámlálást egymással is összehasonlítottuk annak érdekében, hogy egy adott személyről minél több információ rendelkezésre álljon.

Az összekapcsolás minden egyes, az adott népszámlálásban szereplő személy esetében létrehozott külön oldalon történt, ahol az egyén azonosításához szükséges alapadatokon túl (név, születési év, stb.) az is látszott, hogy kikkel élt egy háztartásban a népszámláláskor. Ez, vagyis a rokonok, családtagok ismerete segítette a potenciális találatok közötti „válogatást”. Az oldal alsó részén pedig a genealógiából

⁴² Két forráskritikai megjegyzést kell tennünk: egyrészt sok esetben nem szerepelt a házasságkötés pontos dátuma, ezt nyilvánvalóan nem tudtuk orvosolni, ebben az esetben a férj azonosítása után, manuálisan tudtuk a feleséget is összekapcsolni a genealógiában szereplő önmagával. Másrészt problémát okozott az az eset is, amikor a feleségnek házasságkötés előtt már született gyermeke, ami a genealógia adatait tartalmazó weboldal felépítéséből fakadóan csak úgy kerülhetett fel a nő személyi adatlapjára, mintha családja lenne, viszont a házastárs rovat ilyenkor üres maradt, szintén üres házasságkötési dátummal, így a program a később kötött házasságokat nem tudta azonosítani. Ezt szintén manuális kereséssel és összekapcsolással oldottuk meg.

származó lehetséges, legközelebbi találatok szerepeltek, a saját adataikon túl a genealógiában szereplő összes rokonuk megjelölésével (házastárs(ak), gyerek(ek), szülők, testvér(ek)). Erre azért volt szükség, mert egy adott személy kapcsán nem lehetett előre tudni, hogy a népszámlálás az élete melyik szakaszában rögzíti őt, még gyerekként él-e egy háztartásban a szüleivel vagy már saját családjával él külön stb. Így tehát az egyén azonosítása a családtagok adatainak segítségével történt meg minden kétséget kizáróan.⁴³

Ezt a döntést azonban csak manuális ellenőrzést követően lehetett meghozni, melynek rögzítésére is alkalmas volt a webes felület. Így a különböző forrásból származó, egy személyhez tartozó adatok az egyéni azonosítók (ID) mentén összekapcsolásra kerültek és további elemzés alá vonhatókká váltak.

Miután a két népszámlálásban szereplő 10.000 embert összehasonlítottuk a genealógiában szereplő 90.000 fővel, az alábbi eredményeket sikerült elérni. Az 1857-es népszámlálásban összeírt 4108 emberből 2376 személynek, míg az 1869-es népszámlálásban szereplő 6036 főből 4063-nak találtuk meg a genealógiában a párját. Összesen 1448 egyénről mondható el az adatbázisunkban, hogy mindkét népszámlálásban és a genealógiában is sikerült azonosítani. Figyelembe véve a korban jellemző magas csecsemő- és gyermekhalandóságot, a forrásadottságokat és az összekapcsolás szabályrendszerének összetettségét, ez az eredmény meghaladta az előzetes várakozásokat.

Összegzés

A feldolgozni kívánt népszámlálások keletkezéstörténetének és jellemzőinek ismertetését követően bemutattuk az online genealógia adatait és azok struktúráját. Az ezeken alapuló adatbázis létrehozásának legfontosabb kérdése, hogy milyen módon sikerült a három különböző adatállományban szereplő egyéneket összekapcsolni. Ebben a tanulmányban bemutattuk ennek a menetét, hogy milyen egyéni ismérveket használtunk fel ehhez, milyen megoldandó problémák merültek fel a források adatbázisba rendezése során. Külön kitértünk az adatbázis létrehozásának technikai részére, ismertetve az ezt lehetővé tévő online felületet, valamint elkülönítettük a folyamat automatizálható és manuális részét, amelyek aránya meghatározó lehet ilyen méretű adatállományokkal dolgozva.

⁴³ Az egyének népszámlálásban szereplő adatai és családtagjai segítségével történt a genealógiában való azonosítás. Dürr Ádám a feleségével, Dürr Gertrúddal szerepelt az 1869-es népszámlálásban. A neve, a neme, a születési éve és a felekezete szerint kiadott találatok között szerepel Ifjabb Dürr Ádám János, akiről az elérhető adatok összevetése után kiderül, hogy ugyanaz a személy, mint akit a népszámlálásban azonosítani szeretnénk. Így az adatbáziskezelő rendszer felületén az automatizált találatok manuális ellenőrzése után elvégezhetjük az összekapcsolást.

HISTORY

A népszámlálásban szereplő személy, akit azonosítani szeretnénk a genealógiában

House number: 17 Family number: 1 ID number: 280 Sur name: Dür First name: Ádám Relhead: 101

Laslett: 3a Year of birth: 1797 Sex: 1

Dür Ádám

A népszámlálásban szereplő személy népszámlálásban szereplő családtagja(i)

GENEALOGY TO 1869S

Felsőége: Dür Gertrúd

A kritériumok alapján megtalált lehetséges találatok a genealógiából

GENEALOGY INDIVIDUAL	DISTANCE	MATCH	FULLNAME	SEX	BYEAR	FAMID	FAMID23 ID	CHILD1 NAME	CHILD1 ID	CHILD2 NAME	CHILD2 ID	CHILD3 NAME	CHILD3 ID	FATHER NAME	FATHER MOTHER ID	BROSIS NAME	BROSIS ID
168214	1	Yes	Johann Adam JUN. DÜRR	1	1792	Katharina HUFNAGEL Gertrud ELLER NA	16821572312NA	Jakob DÜRR,Johann DÜRR,Josef DÜRR,Juliana DÜRR,Adam JUN. DÜRR	168216, 168617, 169807, 170794, 171554	None	Johann DÜRR Katharina FAIK	1841.4116094	Adam DÜRR,Anna Elisabeth DÜRR,Anna Maria DÜRR,Elisabeth DÜRR,Johann DÜRR,Magdarena DÜRR,Susanna	128460, 174796, 19514, 122740, 168256, 168442, 173986, 174220, 168181			

GenealogyTo1869 object (26609)

Genealógiában szereplő személy azonosítója (ID), távolság, találat, személy neve, neme, születési éve

Házastárs(ak) neve és ID-ja, az egyes házasságokból született gyerekek neve és ID-ja

1. ábra: Az összekapcsolás folyamata Dür Ádám példáján

Ugyan nemzetközi kutatásokban nem újkeletű a különböző forrásokban szereplő egyének összekapcsolása és az ezek végrehajtásának módszeréről szóló tanulmányok is rendszeresen látnak napvilágot,⁴⁴ magyar adatokon végzett hasonló kutatásokról, különösen annak végrehajtásának módszeréről kevesebb információ áll rendelkezésünkre.⁴⁵ Általában hasonló típusú forrásokat kapcsolnak össze a kutatók – például lélekösszeírásokat, anyakönyvi adatokat, hogy az egyének életében bekövetkező demográfiai eseményeket nyomon tudják követni –, míg jelen kutatásban egymástól struktúrájában, céljában és jellegében is eltérő forrásokat kötöttünk össze. A népszámlálások genealógiai adatokkal való összekapcsolásának módszertani ismertetője támpontul szolgálhat azoknak, akik hasonló jellegű kutatást kívánnak folytatni. A két népességösszeírás és az anyakönyvi adatokat tartalmazó genealógia együttes használata ugyanis jóval komplexebb elemzéseket tesz lehetővé, mint amikor csak egy-egy népszámlálás anyaga maradt fent egy településről.⁴⁶

Források

- Ortsfamilienbuch Bonyhád <https://online-ofb.de/famlist.php?ofb=bonyhad>
 Népszámlálás 1870 = A Magyar Korona Országában az 1870. év elején végrehajtott népszámlálás eredményei a hasznos házi állatok kimutatásával együtt, kiad. Országos Magyar Királyi Statistikai Hivatal. Pest, 1871.
 Népszámlálás 1880 = A Magyar Korona Országában az 1881. év elején végrehajtott népszámlálás eredményei némely hasznos házi állatok kimutatásával együtt. Országos Magyar Királyi Statistikai Hivatal. Pest, 1882.
 TVL IV. B. 152./b = Tolna Vármegyei Levéltár IV. B. 152./b Tolna Megyei Cs. Kir. Megyehatóság iratai 1850–1860. A népszámlálás iratai 1850, 1857. 9–13. doboz
 TVL V. 709./c = Tolna Vármegyei Levéltár V. 709./c Közigazgatási iratok 1850–1949. Bonyhád nagyközség iratai, 1869. évi népesség és háziállatok összeírása. 512–513. doboz

⁴⁴ A történeti adatok összekapcsolásával sokáig főleg a társadalmi és földrajzi mobilitást vizsgálták a kutatók, majd a technológia fejlődésével egyrészt a vizsgálható mobilitás spektruma több generációra nőtt, másrészt új témák vizsgálata is lehetővé vált. Ezekről részletesen beszámol: Ruggles – Fitch – Roberts 2018.

⁴⁵ Különböző forrásokból származó életesemények összekapcsolásával létrejövő adatbázisokat elemez Öri Péter is annak érdekében, hogy különböző demográfiai folyamatokat mutathasson be hosszmenti adatokon, pl.: Öri 2022. 453–479. Szintén különböző forrásokban azonosít egyéneket Ballabás Dániel is, aki a főnemesség rokoni, gazdasági és politikai kapcsolatainak kutatásához használja fel a saját maga által elkészített adatbázisokat. Ballabás 2020.

⁴⁶ Mint például Mohács esetében: Gyimesi 2014.

Felhasznált irodalom

- Andorka 1987 = Andorka Rudolf: Adalékok az ormánsági egyke történetéhez Vajszló és Besence református anyakönyveinek családrekonstrukciós vizsgálata alapján. Budapest (1987), 57–96. Történeti Demográfiai Füzetek 3.
- Andorka 1988 = Andorka, Rudolf: A családrekonstrukciós vizsgálat módszerei. Budapest, 1988.
- Ballabás 2020 = Ballabás Dániel: Az 1911. évi gazdacímtár adatai a magyar főnemesség genealógiai kapcsolathálózatának tükrében. In: Kövér György – Pogány Ágnes – Weisz Boglárka (szerk.): Hálózat és Hierarchia. Magyar Gazdaságtörténeti Évkönyv, 2020. Budapest, 2020. 23–50.
- Benda 2006 = Benda Gyula: A magyarországi családrekonstrukciós vizsgálatok mérlege. In: Benda Gyula: Társadalomtörténeti tanulmányok. Budapest, 2006. 67–76.
- Bunovác 2005 = Bunovác Dezső (szerk.): Tolna megye kézikönyve. Magyarország régiói – Dél-Dunántúli régió. Kaposvár, 2005.
- Dányi 1993 = Dányi Dezső: Az 1850. és 1857. évi népszámlálás. Budapest, 1993.
- Feigenbaum 2016 = Feigenbaum, James: A Machine Learning Approach to Census Record Linking. 2016. <https://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf>
- Gyimesi 2014 = Gyimesi Réka: Mohácsi háztartás-rekonstrukció: Az 1869-es népszámlálás felvételi íveinek feldolgozása. Demográfia 57. (2014:2–3), 183–212.
- Gyimesi 2019 = Gyimesi Réka: Történeti demográfia lokális keretek között: Korlát vagy lehetőség? In: Lengvári István – Pilkhoffer Mónika – Vonyó, József (szerk.): Az ember helye, a hely embere: emberközpontú történetírás, helytörténeti kutatás. A helytörténetírás módszertani kérdései. Pécs, 2019. 326–343.
- Lencsés 2011 = Lencsés Ákos: Felekezet és statisztika. Vigilia, 76. (2011), 751–757.
- Nagy 2008 = Nagy Mariann: Nemzetiség és vallás a Habsburg monarchiában az 1850-es években (az 1850/51-es népszámlálások) In: Kupa, László (szerk.): Vallás és etnikum Közép-Európában. Pécs, 2008. 102–113.
- Őri 2022 = Őri, Péter: Parental loss in 18th–19th century Hungary: the impact of the parents' widowhood and remarriage on their children's survival, Zsám-bék, 1720–1850. *The History of the Family* 27. (2022)4, 53–479.
- Őri – Pakot 2011 = Őri Péter – Pakot Levente: Census and Census-like Material Preserved in the Archives of Hungary, Slovakia and Transylvania. Rostock, 2011. 1–57.
- Pakot 2013 = Pakot Levente: Nemek és nemzedékek. Demográfiai reprodukció a 19–20. századi Székelyföldön. Budapest, 2013. (KSH NKI Kutatási Jelentések 95.)

- Pozsgai 2000 = Pozsgai Péter: Család, háztartás és tulajdon Torna vármegyében a 19. század második felében. Történeti Demográfiai Évkönyv. Budapest, 2000. 166–224.
- Richards et al. 2014 = Richards, Laura – Antonie, Luiza – Areibi, Shawki – Grewal, Gary – Inwood, Kris – Ross, Andrew: Comparing Classifiers in Historical Census Linkage. 2014 IEEE International Conference on Data Mining Workshop. Shenzhen, 2014. 1086–1094.
- Rózsa 2014 = Rózsa Dávid (főszerk.) – Kovács Csaba – Nádudvari Zoltán – Nemes Erzsébet – Pásztor Angelika (szerk.): Portrék a magyar statisztika és népeştudomány történetéből: Életrajzi lexikon a XVI. századtól napjainkig. Budapest, 2014. 363–366.
- Ruggles – Fitch – Roberts 2018 = Ruggles, Steven – Fitch, Catherine A. – Roberts, Evan: Historical Census Record Linkage. *Annual Review of Sociology* 44. (2018), 19–37.
- Thirring 1983 = Thirring Lajos: Az 1869–1980. évi népszámlálások története és jellemzői: 1869–1910. Budapest, 1983.