

A népszámlálásnak, ahogy eddig ismertük, vége van? A regiszteralapú népszámlálásokról

Absztrakt

Az adatforradalomnak köszönhetően jelentős átalakulások zajlottak le az elmúlt évtizedekben, és zajlanak ma is a népszámlálások módszertanában Európa-szerte. Egyre több ország hagyja maga mögött a hagyományos, megkérdezésen alapuló módszert és váltja ki regiszterek használatával. Több európai országban ma már teljes egészében regiszteralapúak a népszámlálások, sok országban kombinált módszert használnak, azonban főként a kelet- és dél-európai országokban, így Magyarországon sem használnak még regisztereket. A regiszterek használatának számos nyilvánvaló előnye van, így például jelentősen olcsóbb, gyorsabb, rugalmasabb, és akár jobb minőségű adatokkal is szolgálhat, ugyanakkor például komoly kihívások elé állítja a statisztikai apparátust, adatvesztéssel járhat, hiszen egyes adatokat nem lehet regiszterekből kinyerni, sérülhetnek az idősorok, és komoly jogszabályi feltételei vannak. Tanulmányunk egy *state-of-art* típusú összegzésére vállalkozik annak, amit a regiszteralapú népszámlálások jelenlegi európai helyzetéről tudunk. Bemutatjuk a regiszterek használatának előnyeit, a lehetséges kihívásokat és az „átállás” feltételeit. Nemzetközi kitekintésünkben az európai helyzet ismertetése mellett három, a népszámlálás során teljes egészében regiszterekre alapozó ország (Dánia, Spanyolország és Szlovénia) módszertanát ismertetjük. Tanulmányunkat a magyarországi helyzet mérlegelésével, valamint a regiszterek használatára való átállást segítő javaslatokkal zárjuk. Úgy látjuk, rövidesen Magyarország is áttérhet egy kombinált módszertanra, ehhez azonban többek között jogszabály-módosításokra, az adatszolgáltatók és a KSH közötti szorosabb együttműködésre és a regiszterek minőségének javítására van szükség.

Kulcsszavak: népszámlálás, census, regiszter, adminisztratív adat

Abstract

The data revolution reinforced major changes in census methodology all over Europe in the last decades. More and more countries leave their traditional, survey-based methods behind, and shift to using registers. In several European countries, censuses are fully register-based, many other countries use a combination of registers and surveying, however, some Eastern- and Southern-European countries (including Hungary) do not use registers in their censuses. There are several obvious advantages of using registers, such as cost-efficiency, speed, flexibility, and even better data quality. On the other hand, it may pose serious challenges on the statistical apparatus, may cause loss of data, as some of the data may not be available in registers, and register-based censuses require a strong legal base. The aim of this study is to give summarize the current state-of-art of registered-based censuses in Europe. We demonstrate its potential advantages, challenges, and the necessary conditions for the use of registers.

Our review involves the description of the European situation, as well as the demonstration of three countries' register-based census methodology (Denmark, Spain, and Slovenia). We provide a discussion of the Hungarian conditions and make proposals on how the Hungarian census could shift to a register-based methodology. We argue that Hungary may soon shift to a combined census methodology, but it requires legislative amendments, stronger cooperation between the data providers and the Statistical Office, and improvement of the quality of the registers.

Keywords: census, register-based, administrative data

Bevezetés

A népszámlálások hagyományos, megkérdezésen alapuló adatait egyre több országban váltják ki regiszteralapú adatokkal. Ausztriában, Szlovéniában és a skandináv országokban már nem is készülnek hagyományos népszámlálások. A kérdőívalapú népszámlálások legfőbb problémája, hogy rendkívül költségesek. De azt is érdemes figyelembe venni, hogy jelentősen átalakult a kutatási környezet az elmúlt évtizedekben. A kérdőíves vizsgálatok kapcsán látható válaszadási hajlandóság csökkenésével párhuzamosan ma a népszámlálási kérdésekre is jóval kevésbé magától értetődő válaszolniuk az állampolgároknak, mint korábban.¹ Emellett azonban a technológiai fejlődésnek és az *adatforradalomnak*² köszönhetően egyre több adminisztratív adat érhető el az állampolgárokról. A leggyakrabban használt ilyen adminisztratív adatok döntő része valamilyen állami regiszterből származik (ingatlan-nyilvántartás, oktatási nyilvántartás stb.), de tudunk egyéb forrásból származó, digitális lábnyomokon alapuló adatbázisok felhasználásáról is (Hollandia, Olaszország, Észtország).

Az adminisztratív adatok felhasználásának több előnye van. Egyrészt jóval olcsóbb, mint a hagyományos módszerek használata, egyes becslések szerint akár tizedére csökkenhetnek a költségek egy teljes mértékben regiszteralapú népszámlálással.³ Az adminisztratív adatok naprakészége azt is lehetővé teszi, hogy nagyobb gyakorisággal frissüljenek a népszámlálási sarokadatok. További előny lehet, ha az adminisztratív adatok minősége jobb, mint a megkérdezésen alapuló adatoké. Bár ez a feltéte-

1 Dygaszewicz 2020.

2 Kitchin 2014; Aragona–Zindato 2016; Németh 2020.

3 Kodaj 2010.

lezés elméletileg megalapozottnak tűnik, mint látni fogjuk, empirikusan gyakran nem alátámasztott a szakirodalom alapján. A regiszteralapú népszámlálásnak ráadásul veszélyei, hátrányai is vannak. A regiszterek használatának és összekapcsolhatóságának komoly jogi feltételei vannak, amelyek megteremtése akadályokba ütközhet. Továbbá, szemben a megkérdezésen alapuló statisztikákkal, a regiszterek adatait nem feltétlenül a statisztikusok „generálják”, így azok nem feltétlenül, vagy kevésbé alkalmasak statisztikai felhasználásra.

Magyarországon az adminisztratív adatok népszámlálási célokra való felhasználása, nemzetközi összehasonlításban, gyerekcipőben jár, amelynek legfőbb oka, hogy jelenleg nem adottak a regiszteralapú népszámlálás feltételei. A jogszabályi környezet jelenleg nem alkalmas a regiszterek adatainak felhasználására és összekapcsolására, ám a KSH kutatói a különböző regiszterek minőségével kapcsolatban is hiányosságokról számoltak be.⁴ A KSH által vizsgált adatbázisok nem teljes körűek, különbözőképpen vannak nyilvántartva, sőt még a rendszereken belül sem feltétlenül egységesek a nyilvántartások.⁵ A KSH mindazonáltal folyamatosan vizsgálja az adminisztratív adatok felhasználását, és ajánlásokat tesz arra a népszámlálások során. A legutóbbi jelentésekben⁶ a soron következő népszámlálásra még a kombinált, azaz megkérdezés- és regiszteralapú adatok egyidejű használatát valószínűsítették a kutatók, és a KSH honlapján⁷ ma is található erre irányuló szándék.⁸

Jelen tanulmány egy *state-of-art* jellegű összegzését kívánja adni az európai népszámlálások módszertanának, különös tekintettel a regiszteralapú módszerre. Tanulmányunk célja bemutatni a regiszteralapú népszámlálásokban rejlő lehetőségeket és a potenciális kihívásokat. A nemzetközi tapasztalatok és a KSH által az elmúlt években összegyűjtött

4 Harcsa 2010.

5 KSH 2017, 64.

6 KSH 2017.

7 KSH Általános Adatkezelési Tájékoztató természetes személyek részére. http://www.ksh.hu/altalanos_adatkezesi_tajekoztato_termeszetes_szemelyek_reszere Letöltve: 2021. április 13.

8 Nehéz ugyanakkor bármit is előre jelezni, hiszen a 2021-re tervezett népszámlálást a koronavírus-járvány miatt 2022 őszére halasztották. Forrás: A kormány 45/2021. (II. 5.) rendelete a veszélyhelyzet idején alkalmazandó további védelmi intézkedésekről szóló 479/2020. (XI. 3.) korm.-rendelet módosításáról. *Magyar Közlöny*, 2021/18, 489. <https://magyarkozlony.hu/dokumentumok/efa2cedfc8de412d4da06cbc09cfaa7d73854183/letoltes> (Letöltve: 2021. április 12.)

információk alapján írásunkat a magyarországi átállás feltételeinek rövid összegzésével, valamint javaslatok megfogalmazásával zárjuk.

A fogalmak tisztázása

A népszámlálások évtizedeken át kizárólag *megkérdezésen alapuló* adatokat tartalmaztak. Az adatgyűjtésben az adatgazda aktívan részt kellett hogy vegyen, az adat jellemzően egy interjúszituációban kérdezőbiztos és kérdőív segítségével keletkezett. A technológiai fejlődés következtében az adatfelvételek fokozatosan digitalizálódtak, ám a domináns továbbra is a megkérdezésen alapuló mód maradt (legyen az személyes vagy online, önkitöltős interjú). Ezt azonban már az 1970-es évek óta Európa-szerte próbálják statisztikusok kiváltani *nem megkérdezésen alapuló* vagy másodlagos adatforrásokkal. Ezen adatok egységes jellemzője, hogy gyűjtésükhöz nincs szükség az adatgazda bevonására, hiszen az adatokat tipikusan valamilyen intézmény tárolja.⁹ A legtöbb esetben ezek az intézmények nem statisztikai intézmények. A nem megkérdezésen alapuló adatok főbb típusait és forrásait az 1. ábrán foglaltuk össze.

A nem megkérdezésen alapuló adatforrások legnagyobb része *adminisztratív adat*, azaz „a közigazgatási/közfeladatot ellátó szervezet hazai jogszabályban vagy EU-s követelményben meghatározott igazgatási feladatának végrehajtása során keletkező adat, gyűjtését jogszabály rendeli el. Idetartoznak a nyilvános, a közhitelű és egyéb nyilvántartások, valamint a hatósági ellenőrzési, gazdasági vagy egyéb tevékenységgel járó adatgyűjtések.”¹⁰ A *regiszterek* tehát azok az adatbázisok, amelyek rendelkezésre állnak ezeknél az intézményeknél. Fontos, hogy – a megkérdezésen alapuló adatokkal szemben – ezeket jellemzően nem statisztikai hivatalok, és nem statisztikai célra gyűjtik, ám céljuk szintén a teljes körű adatgyűjtés¹¹.

⁹ KSH 2017, 8.

¹⁰ KSH 2017, 118.

¹¹ Bár az adatok nem statisztikai céllal készülnek, a regiszteradatok egy része a népszámlálási céloktól függetlenül is részét képezheti a KSH adatgyűjtéseinek, például a nemzeti számláknak.

A nem megkérdézésen alapuló adatok egy további szelete a *magán-szektorban képződő adminisztratív adat*. Ezek annyiban különböznek az adminisztratív adatoktól, hogy nem közfeladatot ellátó szervezetnél képződnek, jogilag nem minősülnek statisztikai adatforrásnak, ám a cégek jogszabályi kötelezettsége gyűjteni őket.¹² Ilyenek a közüzemi szolgáltatók, telekommunikációs szolgáltatók adatbázisai, amelyek szintén közel teljes körű gyűjtések eredményei, és akár jól alkalmazhatóak statisztikai célokra.

A harmadik nagy csoportja a nem megkérdézésen alapuló adatoknak ugyancsak a magán-szektorban képződik, ám nem jogszabályi előírásból, hanem organikus¹³ vagy éppen üzleti céllal. Az adatforradalom nyomán ma korábban nem elképzelhető mennyiségű digitális adat termelődik elképesztő gyorsasággal nap mint nap. A gyakran *big data*ként¹⁴ hivatkozott adattömeg legfontosabb megkülönböztető jele az előző kategóriákhoz képest, hogy kevésbé vagy egyáltalán nem strukturált.¹⁵ Ugyanakkor például olyan „digitális lábnyomok”, mint a pénzügyi, ügyviteli folyamatokból, a szenzorokból, a nyomkövető eszközökből (GPS), a viselkedést nyomon követő szolgáltatásokból vagy a közösségi médiából származó adatok statisztikai felhasználásában nagy lehetőségek rejlenek.

Megkérdézésen alapuló népszámlálásnak nevezzük tehát azokat a cenzusokat, ahol az adatok teljes egészében közvetlenül a lakosoktól kerülnek begyűjtésre tipikusan egy kérdőív segítségével. *Regiszteralapú népszámlálásnak*¹⁶ minősülnek azok a népszámlálások, amelyek a megkérdézésen alapuló adatokat *teljes egészében* kiváltva, *minden* adatot regiszterekből gyűjtene. Azt a megközelítést, amely az adatok egy részét regiszterekből, másik részét hagyományos, megkérdézésen alapuló adatgyűjtésből származtatja, *kombinált népszámlálásnak* hívjuk. Megemlítjük még a *gördülő*

12 KSH 2017, 8.

13 Groves 2011.

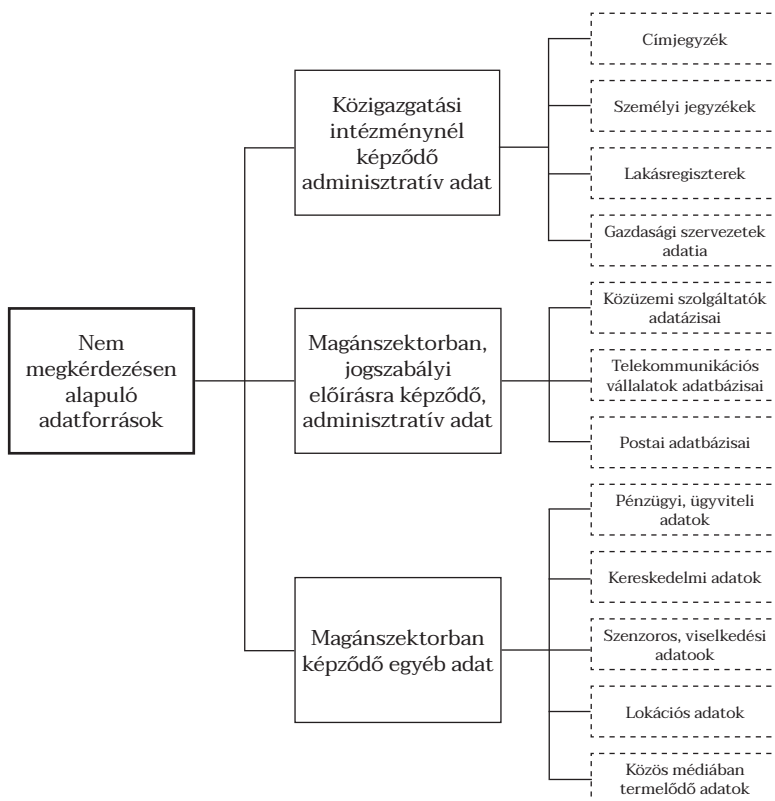
14 Kitchin 2015.

15 Dessewfy–Láng 2015; Kmetty 2018.

16 A KSH nem használja ezt a kifejezést, helyette az adminisztratív adatforrásokon vagy nyilvántartásokon alapuló népszámlálást használja. Érvelésük szerint a regiszter fogalma Magyarországon erősen szűkíti a szóba kerülhető nyilvántartások körét (KSH 2017). Tanulmányunkban azért használjuk mégis a regiszteralapú népszámlálás fogalmát, mert a nemzetközi szakirodalomban ez terjedt el.

népszámlálást,¹⁷ amelyet Franciaországban fejlesztettek ki. Az eljárásban időben elhúzódó mintavételt alkalmaznak, amelynek keretében nem egy időpontban kérdezik meg az ország összes lakosát, hanem egyes években csak bizonyos kiválasztott települések lakosaitól gyűjtenek adatokat, és ezt követően a mintavétel „továbbgördül” más települések lakóira. A teljes periódus ötéves.

1. ábra. A nem megkérdezésen alapuló adatok főbb típusai és példák azok forrásaira



Forrás: KSH 2017. Saját szerkesztés.

17 Durr 2005.

Lehetőségek és kihívások

Ahogy a bevezetőben már érintettük, a regiszteralapú népszámlálás egyszerre hordoz magában lehetőségeket és kihívásokat. Ebben a fejezetben ezeket járjuk körbe. A lehetőségek és előnyök kapcsán a költségeket, az időtényezőt, a minőséget és a gyűjtött adatok körét lehet kiemelni.

Költségek: A regiszteralapú népszámlálások egyik legnagyobb előnye, hogy jelentős költségcsökkenést eredményezhetnek az adatgyűjtésben. A költségcsökkenés mértéke értelemszerűen attól függ, hogy milyen mértékű a regiszterek felhasználása, a nemzetközi tapasztalatok azonban azt mutatják, hogy egy kombinált népszámlálás 22 százalékos költségcsökkenést eredményezhet egy hagyományos népszámláláshoz képest.¹⁸ A kérdőíves megkérdezések teljes kiváltása jó minőségű regiszterekkel ugyanakkor akár 98 százalékos csökkenést is eredményezhet.¹⁹ Fontos azonban hozzátenni, hogy az átállás, az új adatgyűjtési rendszer felállítása magas egyszeri költségekkel járhat (például pilot mérések formájában), így elképzelhető, hogy rövid távon még nem látható megtérülés.²⁰ Az átállásnál ráadásul nemcsak a gyűjtést végző statisztikai intézménynél, de az adatszolgáltató intézményeknél is megjelenhetnek extra költségek.²¹

Idő: Egy másik előnye az adminisztratív adatok használatának, hogy jelentősen gyorsíthatja az adatfelvételt. Ismeretes, hogy a kérdőíves felvételek rendkívül hosszú időt vesznek igénybe. A teljes átállás drámai gyorsulást eredményezhet, mivel a teljes terepmunka megspórolható, de egy kombinált népszámlálás ideje is rövidülhet azáltal, hogy kevesebb adatot kell gyűjteni primer adatfelvétellel. Ugyanaz elmondható azonban itt is, ami a költségeknél, hogy az új rendszer felépítése nyilvánvalóan sok tervezést és időt vesz igénybe az első egy-két adatgyűjtésnél.²² Az időhöz kapcsolódó másik előny, hogy az adminisztratív adatok használata megteremtheti a lehetőségét a gyakoribb, évi vagy akár folyamatos adatszolgáltatásnak. Ennek természetesen az a feltétele, hogy a regiszterek folyamatosan frissüljenek. Már az évi rendszeresség is komoly pozitív

¹⁸ ENSZ 2014.

¹⁹ ENSZ 2014.

²⁰ KSH 2017.

²¹ Összefoglalásért lásd: KSH 2017.

²² ENSZ 2018.

változást jelentene Magyarországon a jelenlegi tízéves ritmushoz képest, különösen annak fényében, hogy egyre nagyobb igény mutatkozik a friss statisztikák használatára minden területen.²³

Minőség: A megkérdezések kihagyása és az adminisztratív adatok használata mellett, hogy csökkenteni mind a megkérdezettek, mind az adatfelvevők terheit,²⁴ potenciálisan korrigálhatja a megkérdezések mérési hibáit. Ilyen az a probléma, hogy a népszámlálási kérdőívekre annak ellenére sem válaszol minden állampolgár, hogy kötelező (*teljes nemválaszolás*), illetve a kérdőívre válaszolók egyes kérdéseknél megtagadják a válaszolást (*részleges nemválaszolás*).²⁵ A hiányzó adatokat súlyozással és imputálással szokás korrigálni, ám még ezek ellenére is számolni kell bizonyos mértékű torzítással a nemválaszolásnak köszönhetően. A nemválaszolási trendek ráadásul arra utalnak, hogy a válaszadási hajlandóság fokozatosan és következetesen csökken világszerte.²⁶ A módszertani szakirodalomban jól dokumentált, hogy a kérdőíves interjú milyen potenciális torzításokat implikál, milyen hibák merülhetnek fel a válaszadó és a kérdezőbiztos oldalán (például társadalmilag elvárt válaszadás). Ezen lehetséges hibaforrásokra nem térünk ki részletesebben, így csak azt hangsúlyoznánk, hogy az adminisztratív adatok gyűjtése során ezek a hibák nem merülnek fel. *A viselkedésalapon vagy beavatkozásmentesen* gyűjtött adatok nagy előnye, hogy az adatgyűjtő nem avatkozik bele az adat születésébe, és az egyének sincsenek tudatában az adatok létrejöttének. Ez sok szerző szerint forradalmi az empirikus szociológia és a népszámlálók számára.²⁷

Új adatok: A regiszterek olyan új adatokhoz juttathatják az adatgyűjtőket, amelyeket megkérdezéssel nem tudnának gyűjteni. Ennek oka, hogy a válaszadók egyszerűen nincsenek birtokában bizonyos információknak (például internetes adathasználat mértéke), vagy nem feltétlenül pontos, hiányos a tudásuk (például energiafogyasztás, egészségügyi események).

23 KSH 2017.

24 Aragona–Zindato 2016.

25 Schulte Nordholt 2018.

26 Luiten és szerzőtársai 2020.

27 Savage–Burrows 2007; Aragona–Zindato 2016.

Több adat: Egy kérdőíves megkérdezés természetes korlátját jelenti, hogy az interjúhossz adott, nem lehet 2-3 órás kérdőíveket felvenni. A regiszterek használatánál nincs ilyen korlátozás.

Szakpolitika-független: A regiszteralapú adatok használatának egy további előnye, hogy míg a megkeresésen alapuló adatgyűjtés során a ki-gyűjtendő információ köre politikai döntések függvénye volt, addig a regiszterek adattartalma adott.²⁸ Ahogy Rudas érvel: „Ebben az értelemben a *big data* nemcsak nagy lehetőség, hanem bizonyos értelemben véve korlát is az államigazgatás működésében.”²⁹

Kooperációs nehézségek: A regiszteralapú adatok használatának hátrányai is vannak. Az egyik gyakran emlegetett nehézség, hogy szemben a hagyományos gyakorlattal, amely során az adatgyűjtő statisztikai intézmény egyedül végzi a gyűjtést, minimálisan kooperálva más intézményekkel, addig a regiszteralapú népszámlálás több intézmény szoros együttműködését feltételezi. Ez egyrészt bonyolítja a folyamatokat, másrészt az állami intézményektől való függés veszélyeztetheti az adatszolgáltatás minőségét.³⁰

A hátrányok között több olyan pont is szerepel, amit az előnyöknél is felsoroltunk.

Fogalmi különbségek: A tény, hogy más, nem statisztikai intézmények szolgáltatják az adatokat, gyakran azt eredményezi, hogy ezek az adatszolgáltatók más-más fogalmakat és definíciókat használnak. Ez komoly problémákat vethet fel az idősoros összehasonlítások vagy a különböző forrásból származó adatok összehasonlítása során. Ahogy az ENSZ ajánlása is fogalmaz, a statisztikai hivatalok minden esetben az adataik koherenciájának és folytonosságának megőrzésére törekcsenek, így mérlegelniük kell, hogy megéri-e az az ár, amit az adminisztratív adatok használatával fizetnek.³¹ A statisztikai hivatalok és az adatszolgáltató intézmények jó kooperációjával mindazonáltal hosszú távon egységesíthetőek a fogalmak. Hasonló probléma, hogy a regiszterek gyakran nem ugyanarra az időpontra tartalmazznak adatokat, ami az adatok összekapcsolásakor szintén gondot okozhat.

²⁸ Galántai 2015.

²⁹ Galántai 2015, 104.

³⁰ ENSZ 2018.

³¹ ENSZ 2018, 15.

Idő: Az időtényező kapcsán nemcsak előnyt, hanem egy potenciális hátrányt is kiemelhetünk. Mivel egyes regiszterek lassabban frissülnek, egy adott évre vonatkozó adatgyűjtés lezárásához akár egy évet is várni kell, míg minden regiszteradatunk rendelkezésre állhat. Ez gyengítheti a népszámlálások naprakészességét.

Minőség: Láttuk, hogy az adminisztratív adatok használatával egyes mérési hibák eltűnnek, és számos olyan adathoz férhetünk hozzá, amihez hagyományos módszerrel nem. Az adminisztratív adatok használata azonban ezzel egyidejűleg adatminőség-romláshoz is vezethet. Az egyik kapcsolódó probléma, hogy a regiszterekben található információ gyakran nem kellően részletes, azaz elnagyoltabbak az adatok, mint a statisztikusok megszokott kategóriái.³² Egyes adatforrásoknál felmerülhet az is, hogy bizonyos aktorok pontatlanul/hibásan gyűjtenek egyes adatokat. Az egészségügyi vagy akár a bűnügyi adatok kapcsán számos ilyen kritika született az elmúlt években.³³

Teljeskörűség: A minőséghez részben kapcsolódó további probléma lehet, hogy az adatok gyakran nem a teljes populációra vonatkoznak.³⁴ Ráadásul vannak olyan adatok, amelyek nem találhatóak meg regiszterekben, viszont hagyományos módszerrel gyűjthetőek lennének. Ilyen például Magyarországon a vallási hovatartozás és az etnikum. Az interneten képződő digitális adatoknál sem világos jellemzően, hogy milyen populációra vonatkoznak, azaz számottevő fedési hibát implikálhatnak.³⁵ Az interneten képződő adatoknál nagyon fontos megvizsgálni az adatok keletkezésének kontextusát és körülményeit (például algoritmusok), hiszen ezek alapvetően befolyásolják az adattömeg érvényességét.³⁶ Mindazonáltal az ilyen típusú adatok minősége igen változó. Az iparág számottevő erőfeszítéseket tett az elmúlt évtizedben a minőségi kritériumok felállítására annak érdekében, hogy minél több ilyen adat legyen felhasználható népszámlálási célokra.³⁷

32 Schulte Nordholt 2018.

33 Erdei 2013.

34 ENSZ 2018, 16.

35 Giczi–Szóke 2018.

36 Kmetty 2018.

37 European Statistical System Committee 2014.

Jogi kérdések: Nem feltétlenül hátrány, de minden bizonnyal kezelendő probléma a regiszteralapú népszámlálásoknál az adatokhoz való hozzáférés jogi kérdése, illetve az adatbiztonság. Törvényileg biztosítani szükséges a statisztikai hivatal számára az adatokhoz való hozzáférést és azok statisztikai felhasználhatóságát. Ez a közintézményeknél tárolt adatok esetében számos európai országban, így Magyarországon is biztosított. Nagyobb problémát jelent a magáncégek tulajdonában lévő adatok felhasználása, ezek ugyanis többnyire, ha személyes adatot tartalmaznak, nem hozzáférhetőek. A jogi nehézségek közé tartozik az egyéni szintű adat-összekapcsolást megnehezítő/megtiltó szabályozási környezet is.

Társadalmi támogatottság: A megkérdezésen alapuló népszámlások természetes módon nagy érdeklődésre tartanak számot, hiszen minden egyes állampolgárt személyesen érint az adatfelvétel. A tapasztalatok azt mutatják, hogy a regiszteralapú adatgyűjtésre való áttérés határozott visszaesést eredményez a nyilvánosság és a média népszámlálás iránti érdeklődését tekintve. Ez nem meglepő annak fényében, hogy kérdés nélkül, a nyilvánosság számára észrevétlenül zajlik az adatgyűjtés. Ideális esetben ez nem azt jelenti, hogy a népszámlálások eredményei és azok felhasználása ne lennének továbbra is fontosak, ám az adatgyűjtés módszertana jóval kisebb figyelmet kap egy regiszteralapú cenzusnál. Az is felmerülhet továbbá, hogy legitimációs vita alakul ki a regiszterek felhasználásáról, mert például a társadalom egy részében ellenérzéseket kelt, ha központi szervezetek különböző adatbázisokhoz, szenzitív egyéni adatokhoz férnek hozzá, és ilyeneket használnak fel.

1. táblázat. A regiszteralapú népszámlálások előnyei és hátrányai

<i>Előnyök</i>	<i>Hátrányok</i>
Jelentősen alacsonyabb költségek	Több intézmény kooperációja szükséges
Gyorsabb adatgyűjtés	Lehetséges fogalmi különbségek a regiszterek között
Gyakoribb adatgyűjtés	Akár rosszabb minőségű, hiányzó adatok
Akár jobb minőségű adatok	Jogszabályi feltételei vannak
Új adatok	Technikai feltételei vannak
Függetlenedés a politikától	Szükséges a közvélemény támogatása

A regiszteralapú népszámlálások feltételei

Az adminisztratív adatok statisztikai célú felhasználásának számos olyan feltétele van, melyek nem feltétlenül vonatkoznak a hagyományos felvételekre. Ezeket az alábbiakban mutatjuk be vázlatosan.³⁸

Jogalap: A regiszteralapú népszámlálásnak számos jogi feltétele van. A legfontosabb, hogy a jogi környezetet úgy kell kialakítani, hogy a statisztikai hivatal hozzáférjen az adatokhoz, és jogi felhatalmazást kapjon azok további használatára. Az Európai Unió statisztikai rendelete³⁹ a 2015-ös módosítással megkönnyítette a jogalap megteremtését, ám további országspecifikus törvénymódosítások lehetnek szükségesek. A magánvállalatoknál képződő adatok felhasználása ugyanakkor jellemzően más megítélés alá esik, Magyarországon például „az egyéb adatforrások tekintetében, ha azok személyes adatot tartalmaznak, az Infotv. [...] 5. §-a a mérvadó, más esetben a polgári törvénykönyv magán-, üzleti és egyéb titkokra vonatkozó rendelkezéseit kell figyelembe venni.”⁴⁰ Tudunk azonban olyan országokról, ahol a statisztikai törvény vonatkozó részeit magáncégekre is kiterjesztették (például Szlovénia, Észtország, Kanada).

A közvélemény támogatása: Mint említettük, a hagyományos népszámlálások nagy nyilvánosságot kapnak, és társadalmi támogatottságuk magasnak mondható. A regiszteralapú adatgyűjtésre való áttérés legitimációs veszélyeket rejthet magában, amennyiben az adminisztratív adatok felhasználása ellenérzéseket kelt az emberekben. A tapasztalatok azt mutatják, hogy ez a kérdés kulturálisan erősen meghatározott. Olyan országokban, mint Hollandia, az adatok felhasználására vonatkozóan gyakorlatilag konszenzus van a lakosság körében,⁴¹ míg a poszt szocialista országokban – a negatív történelmi tapasztalatok (megfigyelés) miatt – ez nem feltétlenül ilyen egyértelmű. A statisztikai hivatalok kommunikációja az adatok biztonságára és felhasználására vonatkozóan kulcsfontosságú egy ilyen átállásnál.

Együttműködés az intézmények között: Jó minőségű adminisztratív adatok gyűjtése csak úgy képzelhető el, ha a statisztikai hivatal és az adat-

38 Az összegzés az alábbi tanulmányokon alapul: ENSZ 2018; Gárdos 2015; KSH 2017.

39 Az Európai Parlament és a Tanács 223/2009/EK rendelete az európai statisztikákról.

40 Gárdos 2015, 37.

41 Schulte Nordholt 2018.

közlő intézmények szorosan együttműködnek. Gyakran ez kormányzati „segítséggel” történik, például egy központi határozat formájában, amely előírja az adatközlő intézményeknek a statisztikai hivatal munkájának segítését. A nem kormányzati szerveknél ez nehezebben biztosítható, esetükben más ösztönzőket kell alkalmazni.

Jó minőségű regiszterek, egységes azonosítók, adat-összekapcsolás: Az adminisztratív adatok használatának vélhetőleg leggyakoribb kerékkötője a jó minőségű regiszterek és az egységes azonosítórendszer hiánya (például Magyarország). Az adminisztratív adatok gyűjtése csak abban az esetben kifizetődő, ha nem jár adatminőség-romlással. Ez pedig csak jó minőségű és teljes lefedettséget nyújtó adatbázisokkal képzelhető el. Amennyiben a regiszterek elemeinek azonosítása nem egységes kóddal történik, az adatok összekapcsolása csak nagyon nehezen, vagy egyáltalán nem lehetséges. Ez legalább három szinten: az egyének; az épület, lakás; valamint a gazdasági szervezetek szintjén jelenhet meg.

Technikai feltételek: Mivel nagy mennyiségű adatot szükséges integrálni különböző informatikai rendszerekből, a regiszterek használatára való áttérés gyakran komoly informatikai fejlesztéseket igényel, továbbá az informatikai munkaerő fejlesztését, bővítését elsősorban a statisztikai hivataloknál, de akár az adatközlő intézményeknél is.

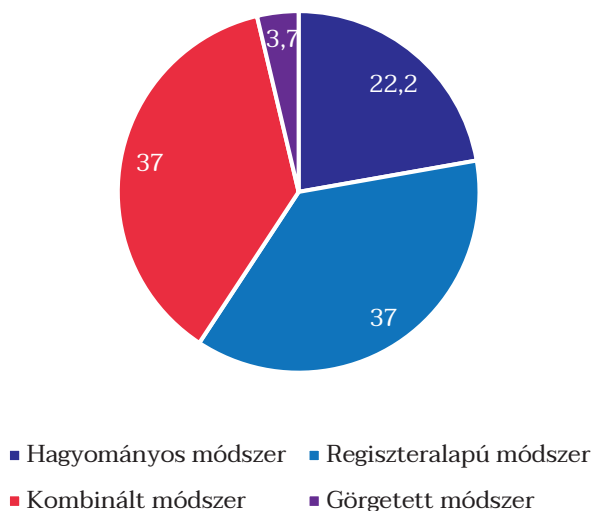
Módszertani sokféleség az unióban

Míg az idődimenzió tekintetében viszonylag korán – az 1870-es évek elején – konszenzus alakult ki az európai nemzetek között a tízéves⁴² kérdéses gyakoriságot illetően, az adatok összehasonlítását érintő tartalmi egységességre való igény csak a 2000-es évek elején alakult ki. Mára ugyan a lekérdezés idejét tekintve (amennyiben az adott országban ez releváns), illetve a rendeletek következtében a kérdések tartamát illetően is egységesség uralkodik az Európai Unió országain belül, az adatok rögzítése szempontjából sokféleség jellemzi a tagországokat. Az adatfelvételhez alkalmazott módszerek négy csoportot képeznek: elsőként említhető

42 A kezdeti, tízévente történő mérés kiegészült az ötévente lebonyolításra kerülő mikrocenzussal, amelyet hazánkban 1963 óta kérdeznek le.

a hagyományos kérdőíves adatfelvétel (online és/vagy offline), a második a kombinált módszer, amely során már alkalmaznak nyilvántartásokat, azonban szükségesnek vélik a hagyományos módszer megtartását is. Az adatfelvételek módjának harmadik típusát a kizárólag regiszteralapon történő lekérdezés képezi, a negyedik pedig a görgetett módszer. Az országok közötti sokféleség abban is tetten érhető, hogy a kombinált módszer tekintetében sem beszélhetünk egységességről, hiszen egyes országok a teljes lakosságon elvégzik a regiszterekből ki nem nyerhető adatok mérését, míg mások eseti mintavételt alkalmaznak. Emellett a regiszteralapú mérésre való átállást tekintve is vegyes módszerekkel közelítenek a tagországok: némelyik ország egyikről a másik hullámra tér át, más országokban ez több időt és erőforrást igénylő folyamat, így egy-két köztes hullámot követően valósul meg a teljes átállás. Az 2. és 3. ábrán, valamint a mellékletben az ENSZ által közölt 2020-as népszámlálási forduló módszertani megoszlása látható az Európai Unió országai közt.

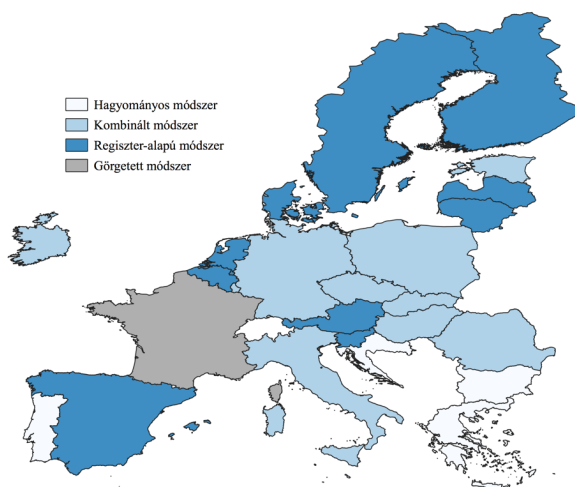
2. ábra. A népszámlálások (tervezett) módszerei az európai uniós országok körében a 2020–22-es népszámlálás alapján (%)



Forrás: ENSZ 2020, saját gyűjtés, saját szerkesztés

A 2020–2022-es (hiszen ebben az évben rendhagyó módon a 2022-es év is részét képezi a fordulónak) népszámlálás során – előreláthatólag – az uniós országok 37 százaléka regiszteralapú mérést végez, és ugyanilyen arányban végeznek vagy terveznek végezni kombinált, tehát regiszteralapú és hagyományos módszerű mérést. Az országok ötöde (22,2 százalék) a hagyományos módszerrel fogja lebonyolítani a népszámlálást. Az uniós országok közül egyedül Franciaország az, amely nem sorolható egyik említett kategóriában sem, ők a görgetett módszert alkalmazzák. Ebben a listában Magyarországot a kombinált módszert használókhoz soroltuk, hiszen a KSH szerint a 2022-es népszámlálás már ezzel a módszerrel készül.⁴³ A regiszteralapú módszert alkalmazó országok között főként észak- és nyugat-európai (skandináv és balti) országokat találunk, míg a kombinált vagy a hagyományos módszert alkalmazók között déli és posztoszocialista országokat. Az országok közötti különbségeknek számos történeti, kulturális oka lehet,⁴⁴ ezek részletezésére jelen tanulmányban nem térünk ki.

3. ábra. A népszámlálások módszerei az európai uniós országok körében a 2020–22-es népszámlálás alapján térképen ábrázolva



Forrás: ENSZ 2020, saját gyűjtés, saját szerkesztés

43 KSH Általános Adatkezelési Tájékoztató természetes személyek részére. http://www.ksh.hu/altalanos_adatkezesi_tajekoztato_termeszetes_szemelyek_reszere Letöltve: 2021. április 13.

44 Karttunen 2001, 247.

Három ország, három út

A következőkben három, részben vagy egészben regisztereket használó európai ország népszámlálási módszertanának átalakulását mutatjuk be. Célunk konkrét példákkal illusztrálni a regiszterek lehetséges felhasználási lehetőségeit. Elsőként Dánia, majd Spanyolország, végül pedig Szlovénia esetében mutatjuk be, hogy váltottak át regiszteralapú népszámlálásra. Azért esett választásunk erre a három országra, mert történelmük, méretük és gazdasági körülményeik igen különbözőek, így bemutatásuk jól szimbolizálhatja, hogy eltérő adottságokkal és utakon, de el lehet jutni a regiszterek extenzív használatáig.

Dánia

Dánia – Magyarországot egy évszázaddal megelőzve – 1769 óta készít népszámlálásokat, továbbá az egyike azoknak az országoknak, amelyek meglehetősen korán áttértek az adminisztratív nyilvántartásokból származó adatok alapján készült cenzusokra. Dániában 1981 óta kizárólag ezzel az eljárással készülnek a cenzusok, melyek technikai és módszertani kivitelezésére alig több mint tíz évre volt szükségük. Ennek megfelelően az első statisztikai kiadvány, amely az adminisztráción alapuló népszámlálás módszertanát taglalja, szintén Dániából származik 1994-ből.⁴⁵ Az adminisztratív alapon működő népszámlálások módszerének kidolgozásához minden országban szoros együttműködés szükséges a regisztereket üzemeltető hatóságok, a közigazgatási szervek és a statisztikai hivatalok között.⁴⁶ Emellett fontos alapfeltételnek tekinthető az adatok összekapcsolását lehetővé tevő jogi környezet is.

Dániában 1966-ban került bevezetésre egy statisztikai törvény, amely lehetővé tette a korábban kormányzati fennhatóság alá tartozó statisztikai hivatal függetlenedését. A törvény teljes hozzáférést és felügyeletet biztosított a rendelkezésre álló nyilvántartásokhoz a hivatal számára, emellett célul tűzte ki az adatfelvételi és -rögzítési módszerek korszerű-

⁴⁵ Wallgren–Wallgren 2007, 31.

⁴⁶ Valente 2010, 2.

sítését. Ez a törvény nagyban hozzájárult a tizenöt évvel később megvalósuló átálláshoz, és következményeként a dán kormány lefektette, hogy az 1970-ben lebonyolított népszámlálás lesz az utolsó, amely a hagyományos módszerrel kerül rögzítésre.⁴⁷

A jogi feltételek biztosítását követően a technikai és az adminisztrációs feltételek kidolgozása következett. Elsőként egy nyilvántartó regiszter létrehozására volt szükség, a korábbi mérések és a rendelkezésre álló nyilvántartások adatai alapján. Így 1968-ban létrejött az ún. *Központi népesség-nyilvántartás* (CPR). A nyilvántartás tartalmazta a lakosság nagyságát, az állampolgárok születési helyét és idejét, nemét, családi állapotát, lakcímét és állampolgárságát. A 70-es években a CPR-be további nyilvántartások is becsatornázásra kerültek, mint például orvosi adatok, az aktuális születések és halálozások, azóta pedig folytonosan frissül a helyi nyilvántartások által közölt adatokkal. Az adatok aktualitása éves szintű, jó minőségű és friss adatokat tesz lehetővé. A CPR rendszerének legjelentősebb újítása, hogy minden állampolgárhoz egyedi azonosító számot rendeltek, amely felváltotta és egységesítette a korábbi azonosítási eljárásokat a közigazgatási rendszer egészében, továbbá kizárta a rendszerek közti duplikálás lehetőségét.⁴⁸

Az 1975-re tervezett népszámlálást elhalasztották a gazdasági miniszter és a dán statisztikai hivatal döntése értelmében, ehelyett egy népszámláláshoz hasonló mérést bonyolítottak le 1976-ban, amely már kombinált módszerrel zajlott. A cenzusszerű mérés során a legnagyobb nehézségnek a lakhatási információk hiánya, a munkahelyek címekre vonatkozó információk hiánya, illetve a munkavállalókra vonatkozó konkrét tevékenységi információk hiánya bizonyult. Emellett adathiányokat regisztráltak az állampolgárok iskolai végzettsége és jelenlegi foglalkozása kapcsán is. Számos egyeztetés és tárgyalás zajlott a következő időszakban az illetékes hatóságokkal, hogy a hiányosságok kijavítására még az 1980-as népszámlálást megelőzően sor kerülhessen. Az 1978-as év fordulópontnak bizonyult a dán népességszámlálások történetében, hiszen ekkor jött létre a *Központi lakás- és épület-nyilvántartás*, amelynek kidolgozásában fontos szerepe volt a statisztikai hivatalnak, hiszen kulcsfontosságú volt,

47 Thygesen 2015, 381.

48 Borchseniu 2011, 2.

hogy a census statisztikai igényeit kielégítse. Ezzel az új nyilvántartással megoldódott a fent említett lakhatási információk hiányának problémája. A másik említett probléma a tevékenységi köröket érinti, melynek kiküszöbölése érdekében *munkaerő-kutatást* végeztek, a munkahelyeket érintő adatproblémák kapcsán pedig *munkahelyprojektet* bonyolítottak le 1978-ban.

Dániában a hetvenes években a statisztika több területén beindult az adminisztratív források alkalmazása: az oktatás, a szociális szolgáltatások és a bűnelkövetések terén is, azután ez az átalakulási folyamat folytatódott a nyolcvanas években, mikor az egészségbiztosítás, a kórházi kezelések és a szociális támogatások nyilvántartásai is csatlakoztak.

Miután az Európai Unió elfogadta a benyújtott eredményeket, 1980-ban Dánia lebonyolította az első adminisztratív forrásokon alapuló népszámlálást.⁴⁹ A dánok sikere főként abban rejlik – más országokkal szemben – hogy az „*internetalapú adminisztrációs bumm*” előtt tették ezt az egységesítést elősegítő lépést, így ez a kódrendszer képezhette a későbbi rendszerek alapját, nem pedig fordítva. Az évek során a *CPR* rendszerébe becsatornázásra kerültek többek között a cégekre, adózásra, biztosításra, oktatási rendszerre, fizetési információkra és munkanélküli-ellátásra vonatkozó további nyilvántartások is.

Spanyolország

Spanyolországban a censusok története 1785-ig nyúlik vissza. Az ország helyzete a népességszámot tekintve sajátságos, hiszen a regiszteralapú népszámlálást lebonyolító országok között az egyik legnépesebbnek mondható. A spanyolok már a 90-es évek közepe óta dolgoznak a regiszteralapú népességszámlálás kialakításán, melynek érdekében 1996-ban létrehozta egy népesség-nyilvántartó rendszert *Padrón* néven.⁵⁰ A *Padrón*t más országok népesség-nyilvántartó rendszereitől eltérő módon a Spanyol Statisztikai Hivatal (INE) koordinálja, nem pedig az állami közigazgatás valamely szerve.⁵¹

49 Thygesen 2015, 383.

50 Valle–Jiménez–Julián 2020, 187.

51 Argüeso 2021, 1.

A népszámlálásoknak alapvetően két kérdésről kell információt szolgáltatniuk: egyrészt az országban élő emberek számáról régiók szerint is értelmezhető részletességgel, másfelől részletes adatokat kell hogy biztosítsanak az állampolgárok főbb demográfiai jellemzőiről, például az aktuális tevékenységi státuszukról, az országba való megérkezésük évéről és az iskolai végzettségükről. Spanyolországban a *Padrón* létezésének köszönhetően pontos adatok állnak rendelkezésre a népességről, így a spanyol rendszer igazi kihívása a másik kritériumnak való megfelelés volt.⁵² A rendszer mára már havi szinten tart nyilván friss adatokat, tartalmazza a nemre, születési helyre és időre, nemzetiségre, iskolai végzettségre és a regiszterek összekapcsolása tekintetében kulcsfontosságú nemzeti azonosító számra vonatkozó információkat a spanyol lakosságról. Adatforrásként működnek az önkormányzatok, az anyakönyvi hivatalok, illetve a konzulátusok is.

A 2001-es hullám kulcsfontosságúnak tekinthető a spanyol cenzusok történetében, hiszen a népszámlálás adatait rögzítő kérdőívre már előre nyomtatásra kerültek a korábbi fordulók adataiból származtatott információk – megkönnyítve a terepmunka feladatát –, amit a személyes lekérdezés során kiegészítettek a további adatokkal. A 2001. évi népszámlálás során már folyamatban volt az adminisztratív és statisztikai nyilvántartások integrálása, mind az egyének, mind a lakásszámok vonatkozásában. A teljes átállásig azonban számos megoldandó feladat áll az INE előtt, egyfelől a lakossági nyilvántartás egységesítése az összekapcsolás megvalósítása érdekében, másfelől pedig a jogszabályi környezet kidolgozása. A terepmunka csökkentése lehetővé tette egy átláthatóbb és ellenőrizhetőbb statisztikai rendszer létrehozását, ami előnyösnek bizonyult az adatok pontosságát és aktualitását illetően.⁵³ Az adatok integrálása a következő években folytatódott: a 2001-es népszámlálási adatokkal, a választási összeírással, az ingatlanokkal, a születési, a házassági és a haláleseti nyilvántartásokkal.⁵⁴ Ennek köszönhetően a 2001-es adatok bevonásával és a *Padrón* segítségével 2011-ben már kevert módszert alkalmaztak a népszámlálás lebonyolítására a spanyolok. A lakosok 12 százalékát kérdezték meg kérdőíves módszerrel, a népesség további részé-

52 Valle–Jiménez–Julián 2020, 188.

53 Maruri–Domingo 2020, 116.

54 Valle–Jiménez–Julián 2020, 188.

nek adatait azonban már a szinte teljes egészében a nyilvántartásokból származtatták.⁵⁵ A spanyol népszámlálás kérdőíve 22 fő kérdést tartalmaz, amelyek 8 további blokkra csoportosíthatók alapvető demográfiai adatok alapján: migráció, családi állapot, lakás összetétele, tanulmányok, igénybe vett szolgáltatások, termékenység, gazdasági aktivitás, mobilitás. A 2011-es mérést követően, 2016-ban lebonyolításra került egy pilot kutatás, melynek célja a 22 fő kérdés lefedése volt, és az arról való meggyőződés, hogy a nyilvántartáson alapuló népszámlálás képes-e reprodukálni legalább azonos részletességgel a 2011-es népszámlálást.

A 2016-os mérés eredményei bizakodásra adnak okot olyannyira, hogy a következő mérést – a tervek szerint – teljes mértékben adminisztratív módon e rendszer segítségével tervezik lebonyolítani. A pilot tanulsága alapján az Európai Unió által megkövetelt összes információt (születési hely és idő, lakhely, nem, családi állapot, gazdasági aktivitás, iskolai végzettség, országba való érkezés)⁵⁶ tartalmazza a spanyol regisztrálapú népszámlálás. Vannak azonban más változók, amelyeket nem vagy nem elég részletesen tartalmaznak a kötelező érvényű uniós előírások, ugyanakkor nagy hagyománnyal rendelkeznek, és relevánsnak mondhatók Spanyolországra nézve. Ilyen opcionális változónak tekinthetők például a migrációval kapcsolatos változók és a háztartástagok közti kapcsolatokra vonatkozó változók, így az INE szakemberei ezen adatokat más forrásokból vagy különböző források összekapcsolásából kísérelték meg kinyerni.⁵⁷ A migrációs adatok pontosabb megismerése érdekében olyan, szinte futurisztikusnak tekinthető felvetés is megfogalmazódott, mely szerint a nyilvántartások erre vonatkozó adatait nemzetközi szinten is érdemes lehet összekapcsolni.⁵⁸

A 2011-es népszámláláshoz hasonlóan tehát a 2021-es fordulónál is alapnak lesz tekinthető a *Padrón* és a 2016-os mérés adatai. A 2021-es mérés során az adatokat az alábbi módokon tervezik kinyerni az egyes forrásokból Spanyolországban. A nemre, korra, születési helyre, állampolgárságra és az országba érkezés idejére vonatkozó adatok a *Padrón* nyilvántartásából származtathatóak, a családi állapot az adóhivataltól

55 Maruri–Domingo 2020, 108.

56 Európai Bizottság 2017, 14.

57 Valle–Jiménez–Julián 2020, 189.

58 Argueso 2021, 11.

és a főbb statisztikai jelentésekből. A gazdasági aktivitás változóját több forrás összehangolásából nyerik, mint például a társadalombiztosítási nyilvántartás, munkanélküliségi nyilvántartás, állami segélyek nyilvántartása. A foglalkozásra vonatkozó adatokat a társadalombiztosítás és a kölcsönösségi nyilvántartás adataiból, az oktatásra vonatkozóakat az oktatási hivatal adataiból. Az ingóságokra vonatkozó adatokat az adóhivataltól, a tartózkodási helyre vonatkozó adatokat pedig az ingatlan-nyilvántartások adataiból.⁵⁹

Szlovénia

Az első népszámlálást Szlovénia területén 1857-ben tartották. A szlovén népszámlálások végigkövetésekor azonban nem mellőzhetjük a tényt, hogy az ország éveken át Jugoszlávia részét képezte, ami hatást gyakorolt a népszámlálások lebonyolítására is. A jugoszláv népszámlálások egyik sajátossága volt, hogy területspecifikusan kellett kialakítani a kérdéseket mind nyelvileg, mind a megfogalmazást tekintve, amely rontotta a későbbi évek népszámlálásainak összehasonlíthatóságát, emellett az adatok folytonosságára hatással voltak a világháborúk okozta határváltoztatások is.⁶⁰ Szlovénia területe részét képezte az Osztrák–Magyar Monarchiának és Jugoszláviának, történelmében nagy hangsúlyt kapott tehát az eltérő nemzetek együttélése. A jugoszláv népességszámlálás azonban politikailag befolyásolt volt, így az etnikai kisebbségekre vonatkozó adatokról – amilyen a nemzeti hovatartozás, a vallás, az anyanyelv és a családban használt nyelv – nem állnak rendelkezésre megfelelő minőségű adatok ebből az időszakból.⁶¹

Az ország már az 1970-es évek eleje óta készült a regiszteralapú mérésekre való átállásra, melynek első állomásaként említhető, hogy a Szlovén Statisztikai Hivatal (SURS) által kidolgozásra kerültek a nyilvántartásokat illető jogszabályok. Kezdetben négy fő regisztert alakított ki a helyi közigazgatás és a SURS: a *Központi népesség-nyilvántartás*, a *Területi egységek nyilvántartása*, a *Foglalkoztatás-nyilvántartás* és a *Cégnyilvántartás*. Fontos

59 Valle–Jiménez–Julián 2020, 190.

60 Josipovič 2015, 160.

61 Flander 2008, 1021.

pontját képezte az átállásra való felkészülésnek, hogy a 90-es évek fordulóján a kérdőíveket személyre szabottan előre kinyomtatták a lakosok korábbi mérésekből származó adataival. A kitöltendő kérdőíveken szerepelt a lakosok azonosító száma, neve, lakcíme és a népszámlálási körzet kódja. Az évek során némi átalakításon ment át a nyilvántartási rendszer, mígnem minden releváns nyilvántartás elemét képezte, így 2002-ben lebonyolították az első kombinált nép- és lakásszámlálást Szlovéniában, tehát az adatokat részben adminisztratív és statisztikai forrásokból, részben pedig helyszíni összeírás alapján gyűjtötték össze.⁶²

A regiszteralapú népszámlálások során az egyik legfontosabb feladat a személyekre, a háztartásokra és a lakásokra vonatkozó adatok egyedi azonosítókkal való összekapcsolása volt. Az adatok összekapcsolásánál azonban Szlovénia azzal a kihívással szembesült, hogy a *Központi népesség-nyilvántartásban* a többlakásos épületekben élő személyek lakhatási száma nem felel meg a helyi adatoknak, megközelítőleg 400 000 személy adata hiányzott, ami a többlakásos házban élő lakosság több mint felét jelenti, ez a változó pedig kulcsfontosságú volt a lakások személyekhez és háztartásokhoz való illesztése szempontjából.

Ennek orvoslására a belügyminisztérium és a SURS egyfelől módszertani megoldásokat dolgozott ki a hiányzó lakásszámok automatikus meghatározására, például összekapcsolással, a tulajdonosok és bejegyzett lakóhelyük alapján. Másfelől hivatalos levelet küldtek a többlakásos épületben élő háztartások számára, amelyek lakószáma nem volt ismert, hogy jelentsék ezt az információt. Ennek következtében a 2011-es 75 százalékos adatról jelentősen megnőtt a lakcímek helyes nyilvántartásának aránya, hiszen 2015-re már 94 százalék volt. A nemzeti statisztikai törvény értelmében csak statisztikai adatokra alapozva nem lehetséges adminisztratív nyilvántartásokat létrehozni, következésképpen a lakások 2002. évi népszámlálási adatai nem használhatók fel erre a célra. 2007-ben azonban létrehozták az addig hiányzó helyi *Központi lakás- és épület-nyilvántartást* is, amelynek alapját képezik az összeírt ingatlanok, a szlovén földmérési és térképészeti hatóság adatai, valamint a már rendelkezésre álló források (geodéziai kataszter, bírósági telekkönyv). A lakószámot és a lakcímet is tartalmazza az adatbázis, így létrejöhett a szlovén sajátosságként említ-

62 Flander 2008, 1014.

hető *Elektronikus háztartás-nyilvántartás* is. Más nyilvántartás-alapú népszámlálást alkalmazó országokban nincs ilyen magas színvonalú adatforrás a háztartások szerkezetéről.⁶³

A 2002-es adatokat referenciapontként alkalmazva, illetve bevonva számos adminisztratív nyilvántartást,⁶⁴ 2011-re Szlovénia egyike lett annak a kilenc európai országnak,⁶⁵ amely a népszámlálást regiszteralapú módszerrel bonyolította le. Szlovénia átállását megkönnyítette, hogy a SURS számára biztosított a hozzáférés az állami nyilvántartási rendszerekhez, továbbá a vonatkozó jogszabály nemcsak a közigazgatási célú adatokhoz biztosít hozzáférést, hanem a vállalkozások adminisztratív adataihoz is.⁶⁶

A korábbi évek népszámlálása jellemzően a tavaszi időszakban volt esedékes, azonban Szlovénia a mérések eszményi időpontját a regiszteralapú mérésre való átállást követően áthelyezte a naptári év elejére, január 1-re. Ennek okaként a más nyilvántartásokkal való koherenciát jelölték meg, hiszen számos adminisztratív forrás kapcsolódik a naptári évhez.⁶⁷

A hazai helyzet

Magyarországon a népszámlálást az első népszámlálás óta (1784–85) – így a legutóbbi, 2011-es cenzust is – megkérdezésen alapuló módszerrel végzik. 2011-ben regiszterek csak előkészítési és adatpótlási célokra kerültek felhasználásra. A megkérdezés hibrid módszerrel zajlott: a lakosoknak lehetőségük volt online kitölteni a kérdőívet, azokat pedig, akik nem töltötték ki online, kérdezőbiztosok keresték fel otthonukban. A KSH 2007-ben kezdte meg a regiszteralapú népszámlálás megvalósíthatóságának feltérképezését, azóta meglehetősen világossá váltak a megvalósulás

63 ENSZ 2018, 59.

64 Többek között a Központi Népeségregisztert, a Háztartásregisztert, a Foglalkoztatási regisztert, a belügyminisztérium adatait, hivatalos statisztikai adatbázisokat, a hatósági és az ingatlan-nyilvántartást, a Szlovénia által vezetett cégnylvántartást és a jogi nyilvántartásokat.

65 Dánia, Finnország, Hollandia, Izland, Belgium, Svédország, Norvégia és Ausztria mellett.

66 KSH 2017.

67 ENSZ 2018, 58.

feltételei.⁶⁸ Az elemzések fő tanulsága, hogy Magyarországon továbbra is komoly akadályokba ütközik a regiszterek népszámlálási célokra való felhasználása. Ennek két fő oka a jogi környezet hiányosságai, valamint a regiszterekkel kapcsolatos hiányosságok. Mindazonáltal a KSH honlapján elérhető információk szerint a 2022-re halasztott népszámlálás során már név felvételére is sor kerül, s ennek nyomán kombinált módszert terveznek használni.⁶⁹

Ahogy utaltunk rá korábban, a hatályban lévő magyarországi törvények alapvetően megteremtik a lehetőséget az adminisztratív adatok statisztikai célú felhasználására, kisebb lényegi módosítások azonban szükségesek lennének. A Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala (KEKKH) kezelésében álló személyiadat- és lakcímnyilvántartás használatát és a személyes adatok statisztikai célú összekapcsolását a 2013. évi CCXX. törvény (Interoptv.) 345/2014. (XII. 23.) rendelete lehetővé teszi. Mindezt nagyban elősegíti az Európai Unió 2015-ös rendelete is.⁷⁰ A jogosítványok tehát adottak, ám „azok érvényesítése továbbra is a partnerek együttműködési hajlandóságától függ”.⁷¹ Hiányoznak azok a rendelkezések is, amelyek előírják, hogy a regiszterek kialakítása, fejlesztése csak a KSH együttműködésével lehetséges. Mint látni fogjuk, ez utóbbi hiány szorosan összefügg a hazai regiszterek minőségi problémáival. További nehézség, hogy az egyéb, nem közigazgatási intézmény által gyűjtött adatok használatát jelenleg szigorú adatvédelmi törvények védik.⁷²

Ami a regisztereket illeti, az egyik alapvető probléma magyar sajátosság: nincs Magyarországon egységes személyazonosító rendszer. Minden lakost legalább három azonosító kódszám azonosít (személyi szám, adóazonosító jel, tajszám), és ezek nem feleltethetők meg egymásnak.⁷³ Mivel azonban az adattartalom között van átfedés, az adatok összekap-

68 KSH 2017; „Felkészülés a 2011-es népszámlálásra” című európai uniós projekt; Harcsa 2010.

69 KSH Általános Adatkezelési Tájékoztató természetes személyek részére. http://www.ksh.hu/altalanos_adatkezelesi_tajekoztato_termeszetes_szemelyek_reszere Letöltve: 2021. április 13.

70 Az Európai Parlament és a Tanács 223/2009/EK rendelete az európai statisztikákról.

71 KSH 2017, 93.

72 Giczi–Szőke 2018.

73 KSH 2017, 22.

csolása elvileg megoldható technikailag, például mesterséges azonosítók bevezetésével.⁷⁴

A KSH jelenleg 19 regisztert tart nyilván népszámlálási célokra reálisan felhasználható forrásként.⁷⁵ A fő adminisztratív forrás a KEKKH kezelésében lévő személyiadat- és lakcímnnyilvántartás, ahonnan a fontosabb demográfiai adatok átvehetőek a bejelentett lakcímmel rendelkezőkre vonatkozóan.⁷⁶ Fontosabb forrás emellett a földhivatal nyilvántartása, az OEP-adatbázis, az Országos Nyugdíjbiztosítási Főigazgatóság (ONYF) adatbázisa, az Adóadatbázis és az Oktatási információs rendszer. Általános tapasztalat, hogy ezek az adatbázisok „nem teljes körűek, nyilvántartási módjuk különböző, olykor egy rendszeren belül országosan sem egységesek”.⁷⁷ Az egyéb, digitális forrásból származó adatok (mobiltelefon-adatok, szenzoradatok) felhasználására a KSH végzett elemzéseket, ezek statisztikai felhasználhatósága azonban egyelőre jogi és anyagi korlátokba ütközik.⁷⁸

Nem térünk ki az egyes regiszterekkel kapcsolatos speciális minőségi problémákra, itt csak a legjelentősebb hiányosságokat említjük meg. A KEKKH adatbázisa nyomán egyáltalán nem áll rendelkezésre adat a *bejelentés nélkül élők*ről és a *bejelentés nélkül dolgozók*ról. Nem áll rendelkezésre továbbá információ olyan változókról, mint az *iskolai végzettség*, a *család/háztartásszerkezet*, a *szenzitív kérdések* (anyanyelv, nemzetiség, vallás, fogyatékossgal élőkre vonatkozó kérdések). A lakcímadatok hiányosságai kritikusak lehetnek, hiszen feltételezhető, hogy a hiányzó adatok nem véletlenszerűen, hanem szisztematikusan hiányoznak egyes országrészekben (például határ mentén dolgozók). Ezek a nem ismert, vélhetőleg szisztematikus hiányok lényegében minden statisztikát torzíthatnak, beleértve a lakosságszámot, de bármilyen területi szintű statisztikát is. Nem áll továbbá rendelkezésre jelenleg regiszter a magyarországi *lakásállományról*. A földhivatal tulajdonában lévő információk lekérése rendkívül költséges, statisztikai célra alig alkalmas, így a KSH a 2011-es népszámlálás címállományát használja.

74 Kodaj 2010.

75 Részletes listáért lásd: KSH 2017, 63.

76 Részletes listáért lásd: KSH 2017, 69.

77 KSH 2017, 64.

78 Giczi–Szöke 2018.

Mint ismertettük, a regiszteralapú népszámlálások kulcsa az intézmények közötti együttműködés. A KSH tapasztalatai arról árulkodnak, hogy hazánkban ezen a téren is sok tennivaló. Bár léteznek intézményesült formái az együttműködésnek (együttműködési megállapodások), egyfajta *kulturális* fordulatra is szükség lenne, ami az adatgazda intézmények oldalán a kooperációra való nyitottság, a transzparencia (adatgyűjtéssel kapcsolatos információhiány) általánossá válását jelentené az adatbázisok kialakításának teljes folyamatára vonatkozóan. Mint utaltunk rá, ez jogi eszközökkel is ösztönözhető lenne.

Mindazonáltal olyan feltételről is tudunk, mely már teljesül hazánkban. A KSH beszámolója szerint az utóbbi évek informatikai fejlesztései nyomán a regiszterek fogadására való átállás technikai feltételei biztosítottak.⁷⁹

Összegzés

Tanulmányunk címében megfogalmazott kérdésünkre tehát a válasz: a népszámlálásnak, ahogy ma ismerjük, (lassan) vége van. A legtöbb európai ország (és számos ország Európán kívül is) népszámlálási gyakorlatában részben vagy teljesen áttért a regiszterek használatára. Jó hír, hogy hazánk is tervez áttérni a kombinált adatgyűjtésre, ami, mint láttuk, az első lépés lehet egy teljes egészében regiszteren alapuló népszámlálási gyakorlathoz. Tanulmányunkban részletesen vizsgáltuk a regiszteralapú népszámlálás előnyeit és lehetséges veszélyeit. Úgy véljük, amennyiben rendelkezésre állnak jó minőségű regiszterek, és az áttérés jól tervezett, a regiszteralapú népszámlálással összességében nyerhet egy ország. Arról nem is beszélve, hogy a nemválaszolási trendek miatt hosszú távon valószínűleg elkerülhetetlen az átállás.

A regiszterek használatára való áttérés folyamata ugyanakkor több, mint technikai változás. A technológiai fejlődés és a digitalizáció nyomán létrejövő új adatok és adatgyűjtési lehetőségek alapjaiban formálják

⁷⁹ KSH 2017, 70.

át a társadalomkutatást, és a társadalmat magát.⁸⁰ Számos szerző érvel úgy, hogy az adatforradalom válságba sodorta mind a hivatalos statisztikákat,⁸¹ mind az empirikus szociológiát.⁸² Az biztosnak látszik, hogy a statisztikai hivatalok nem hagyhatják figyelmen kívül a digitalizáció okozta társadalmi változásokat, és folyamatosan követniük kell az újonnan képződő digitális adatok felhasználási lehetőségeit. Ennek hiányában a hivatalos statisztika kiüresedik, és legitimitását veszti. Itt csak egy ilyen példát említünk meg. Ahogy Davies⁸³ érvel, a hivatalos statisztika hagyományos fogalmainak egy része legalábbis széttöredezett. Az identitáspolitika széles körű terjedése nyomán ma már az emberek jelentős része nem a hagyományos demográfiai kategóriákban értelmezi magát (például alsó középosztálybeli férfi). Az identitásképző tényezők köre jelentősen kibővült, kezdve a gendertől a politikai preferenciákon át a kedvenc focicsapatig vagy popsztárig.⁸⁴ Ezzel nem azt akarjuk mondani, hogy a hivatalos statisztikáknak ezentúl ezekkel a kategóriákkal kellene foglalkozniuk, de amennyiben nem tartják az ujjukat a társadalmi változások ütőerén, úgy valós veszély a hagyományos kategóriák kiüresedése. Fontos hangsúlyozni tehát, hogy a regiszterek használata, és különösen az adatforradalom nyomán képződő digitális adatok használata alapvető gondolkodásbeli, módszertani, illetve – némileg túlzóan – *paradigmatikus* változásokat igényel a hivatalos statisztikákat művelők között.

Sajnos Magyarország az új digitális adatok használata és általában az adminisztratív adatok használata terén is komoly lemaradásban van európai viszonylatban. Bár számottevő erőfeszítések történtek az elmúlt évtizedben a jogi környezet szabályozása terén, szükséges lenne törvényileg szabályozni az adatgazdák együttműködését a KSH-val. Ez elősegítené, hogy már az adatgyűjtés során olyan struktúrájú és minőségű adatok képződjenek a különböző intézményeknél, amelyek később jól összekapcsolhatóak és statisztikai elemzésre alkalmasak.

Megítélésünk szerint továbbá a jelenleg elérhető hazai regiszterek oly mértékben hiányosak, hogy a megkérdezésen alapuló adatok ezen regisz-

80 Aragona–Zindato 2016

81 Davies 2017; Németh 2020.

82 Savage–Burrows 2007.

83 Davies 2017.

84 Németh 2020, 192–193.

terekkel való kiváltása irreális adatminőség-romláshoz vezetne egy népszámlálásnál, még akkor is, ha csak az adatok egy része kerülne kiváltásra regiszterekkel. A fenti jogszabályi változtatásokkal egyidejűleg szükséges egy olyan átfogó *program* elindítása, amely ezen hiányosságok kezelését célozza. A legnagyobb deficit a bejelentés nélkül élők kimaradása. Kérdés például, hogy a NAV adónyilvántartásaiból, OEP-adatokból vagy lokális önkormányzati statisztikákból lehet-e pótolni ezt az adatot.

Amennyiben a regiszterek használatának fenti, technikai akadályai elhárulnak, fontosnak tartjuk az adminisztratív adatok népszámlálási célokra való felhasználásának aktív *lakossági kommunikációját*. Nincsenek adataink a lakossági véleményekről, de úgy sejtjük, hogy az államszocializmus negatív tapasztalatai miatt a lakosság egy részében ellenérzéseket kelthet, ha személyes adataikat állami intézmények felhasználják. A megkérdésesen alapuló népszámlálásnak ráadásul több évszázados hagyománya van Magyarországon, így az áttérés során fontos hangsúlyozni, hogy pontosan miért szükséges változtatni, és milyen előnyei származhatnak ebből az egyéneknek.

Irodalom

- Aragona, B. – Zindato, D. 2016: Counting People in the Data Revolution Era: Challenges and Opportunities for Population Censuses. *International Review of Sociology*, 26, 3, 367–385. <https://doi.org/10.1080/03906701.2016.1244926>.
- Argüeso, A. 2021: *2021 population census and migration statistics in Spain. Why exchanging microdata?* Socio-Demographic Statistics. INE, Spanyolország. https://www.ksh.hu/dgins2017/papers/dgins2017_session1_es.pdf
- Borchsenius, L. 2001: From a Conventional to a Register-based Census of Population. *Les Recensements après*.
- Davies, W. 2017: How Statistics Lost Their Power – and Why We Should Fear What Comes Next. *Guardian*. <https://www.theguardian.com/politics/2017/jan/19/crisis-of-statistics-big-datademocracy>
- Dessewfy, T. – Láng, L. 2015: Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika*, 92–93, 155–168.
- Durr, J. M. 2005: The French new rolling census. *Statistical Journal of the United Nations Economic Commission for Europe*, 22, 1, 3–12.
- Dygaszewicz, J. 2020: Transition from traditional census to combined and registers based census. *Statistical Journal of the IAOS*, 36, 1, 165–175. <https://doi.org/10.3233/SJI-190566>.
- ENSZ 2013: *High-Level Group for the Modernisation of Statistical Production and Services. What does “big data” mean for official statistics?* <http://www1.unece.org/stat/platform/pages/view-page.action?pageId=77170622>

- ENSZ 2014: *Measuring Population and Housing. Practices of UNECE countries in the 2010 round of censuses*. New York – Geneva, <https://unece.org/measuring-population-and-housing>.
- ENSZ 2015: *Report of the Global Working Group on Big data for official statistics*. <http://unstats.un.org/unsd/statcom/doc15/2015-4-BigData.pdf>
- ENSZ 2018: *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses*. <https://doi.org/10.18356/48165223-en>
- ENSZ 2020: *Censuses of the 2020 round. Plans and practices of UNECE countries for the population and housing censuses of the 2020 round*. <https://statswiki.unece.org/display/censuses/Censuses+of+the+2020+round>
- Erdei, G. 2013: A bűnügyi statisztika valóság tartalma. *Military Science Review/Hadtudományi Szemle*, 6(3), 119–125.
- Európai Bizottság 2017: *Az Európai Statisztika Gyakorlati Kódexe*. Luxembourg, Európai Bizottság. <https://ec.europa.eu/eurostat/documents/4031688/9394119/KS-02-18-142-HU-N.pdf/5a4e298a-40e4-472c-aadf-5b831841a634>
- European Statistical System Committee 2014: *ESS Big Data Action Plan and Roadmap 1.0. 22nd Meeting of the European Statistical System Committee*. Lettország.
- Flander, A. O. 2008: A regiszteren alapuló népszámlálás lehetőségei és kihívásai – Szlovénia esete. *Statisztikai Szemle*, 86. évfolyam, 10–11. szám
- Galántai, J. 2015: Interjú Rudas Tamással. *Educatio*, 3, 103–107.
- Gárdos, É. 2015: Adatok és kezelésük a hivatalos statisztikában. *Statisztikai Szemle*, 3, 27–39.
- Giczi, J. – Szőke, K. 2018: Official Statistics and Big Data. *Intersections*, 4, 159–152. <https://doi.org/10.17356/ieejsp.v4i1.408>
- Groves, R. M. 2011: Three Eras of Survey Research. *Public Opinion Quarterly*, 75, 5, 861–871. <https://doi.org/10.1093/poq/nfr057>
- Harcza, I. 2010: Műhely-konferencia az adminisztratív adatállományok statisztikai célú hasznosításáról – különös tekintettel a népszámlálásra. *Statisztikai Szemle*, 88(5), 549–558 http://www.ksh.hu/statszemle_archive/2010/2010_05/2010_05_549.pdf
- Josipovič, D. 2015: Slovenia and the Census: From the 20. Century Yugoslav Counts to the Register-Based Census of 2011. *Contemporary Southeastern Europe*, 2(2), 159–175.
- Karttunen, S. 2001: How to make use of census data in status-of-the-artist studies: Advantages and shortcomings of the Finnish register-based census, *Poetics*, 28(4), 273–279. [https://doi.org/10.1016/S0304-422X\(01\)80004-8](https://doi.org/10.1016/S0304-422X(01)80004-8)
- Kitchin, R. 2015: *What Does Big Data Mean for Official Statistics?* Egyesült Királyság, Discover Society.
- Kitchin, R. 2014: *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Egyesült Királyság, SAGE Publications Ltd. <https://doi.org/10.4135/9781473909472>
- Kmetty, Z. 2018: A szociológia helye a Big Data-paradigmában és a Big Data helye a szociológiában. *Magyar Tudomány*, 179(5), 683–692. <https://doi.org/10.1556/2065.179.2018.5.11>
- Kodaj, K. 2010: A regiszteralapú népszámlálás feltételei Magyarországon. *Statisztikai Szemle*, 88(9), 990–1001.
- KSH 2017: *Másodlagos adatforrások használata a statisztikában. Műhelytanulmányok 11.* Magyarország.
- Luiten, A. – Hox, J. – de Leeuw, E. 2020: Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics*, 36, 3, 469–487. <https://doi.org/10.2478/jos-2020-0025>
- Németh, Zs. 2020: A hivatalos statisztika válsága az adatforradalomban. *Replika*, 117–118, 179–210. <https://doi.org/10.32564/117-118.7>

- Savage, M. – Burrows, R. 2007: The coming crisis of empirical sociology. *Sociology*, 41, 5, 885–899.
- Schulte Nordholt, E. 2018: The usability of administrative data for register-based censuses. *Statistical Journal of the IAOS*, 34, 4, 487–498. <https://doi.org/10.3233/SJI-180425>
- Thygesen, L. 2015: The use of administrative sources for censuses: Merits and challenges. *Statistical Journal of the IAOS*, 31(3), 381–389.
- Treviño, R. – Domingo, A. 2020: ¿Adiós al censo en España?: elementos para el debate. REIS: *Revista Española de Investigaciones Sociológicas*, (171), 107–124. doi:10.5477/cis/reis.171.107
- Valente, P. 2010: Census taking in Europe: how are populations counted in 2010. *Population & Sociétés*, 467, 1–4.
- Valle, J. L. V. – Jiménez, A. A. – Julián, M. P. 2020: Moving towards a register based census in Spain. *Statistical Journal of the IAOS*, 36(1), 187–192. <https://doi.org/10.3233/sji-190516>
- Wallgren, A. – Wallgren, B. 2007: *Register-based statistics: administrative data for statistical purposes (Vol. 553)*. Egyesült Királyság, John Wiley & Sons.

Melléklet

A népszámlálások módszerei az európai uniós országok körében a 2020–22-es népszámlálás alapján az adott módszerre való átállás évszámával

<i>Hagyományos módszer</i>	<i>Kombinált módszer</i>	<i>Regisztralapú módszer</i>	<i>Görgetett módszer</i>
Bulgária	Csehország (2011)	Ausztria (2006)	Franciaország (2004)
Ciprus	Észtország (2021)	Belgium (1991)	
Görögország	Írország (2020)	Dánia (1981)	
Horvátország	Lengyelország (2011)	Finnország (1981)	
Málta	Luxemburg (2020)	Hollandia (1971)	
Portugália	Magyarország (2022)	Lettország (2001)	
	Németország (2010)	Litvánia (2019)	
	Olaszország (2020)	Spanyolország (2016)	
	Románia (2020)	Svédország (2011)	
	Szlovákia (2020)	Szlovénia (2011)	

Forrás: ENSZ 2020, saját gyűjtés, saját szerkesztés