

Adatok és kezelésük a hivatalos statisztikában¹

A statisztikai törvény² (Stat tv.) első paragrafusában szerint a „statisztika feladata és célja, hogy valósághű, tárgyilagos képet adjon a társadalom, a gazdaság, a tulajdonviszonyok, a környezet állapotáról és változásairól az államhatalmi és a közigazgatási szervek, valamint a társadalom szervezetei és tagjai számára.” Az elmúlt másfél évszázadban, mióta ez a tudományág önállósodott, ezen kívül számtalan más definíciót is megfogalmaztak, de minden meghatározás lényege, hogy a statisztika az adatgyűjtés és adatfeldolgozás általános tudománya. A statisztikai munka valójában a valóságnál kezdődik, azt képezi le adattá valamilyen definíció mentén, és a folyamat végén a közölt adat a felhasználónál válik információvá.

Az adat nagyon tág fogalom. A KSH módszertani dokumentációjának fogalomtárára (KSH, 2014a) szerint „az adat az információ formalizált módon való megjelenítése, amely alkalmas feldolgozásra, továbbításra, közlésre, értelmezésre.” Egy adatból akkor lesz információ, azaz objektumokra (tényekre, eseményekre, dolgokra, folyamatokra vagy a gondolati világ elemeire) vonatkozó ismeret, ha definiált, azaz pontosan meghatározott, hogy milyen objektum, melyik ismérvére vonatkozik, és milyen nomenklatúra, jegyzék szerint kell értelmezni. A statisztikai adat a „valós világ egyedeinek tulajdonságaira vonatkozó statisztikai megfigyelések, illetve további statisztikai műveletek eredménye” (KSH, 2014), vagyis annyiban specifikus, amennyiben előállítását és értelmezését a statisztikai módszertanok szabják meg.

Az adatok egy speciális fajtája az ún. metaadat, mely más adatokat ír le, illetve határoz meg. A statisztikai metaadatok tartalmazzák a statisztikai adatrendszerben használt fogalmakat, nomenklatúrákat, a felhasznált adatforrások leírását, az adatelőállítás módszertanát. A metaadatok alapján lehet eldönteni, hogy az adatok megfelelnek-e a nemzetközi és hazai standardoknak, kritériumoknak.

¹ Köszönetnyilvánítás: Ezúton szeretnék köszönetet mondani kollegáimnak, dr. Szép Katalinnak, dr. Nagy Eszternek és Nagy Beátának, hogy szakértelmükkel, észrevételeikkel és javaslataikkal hozzájárultak a tanulmány elkészítéséhez.

² 1993. évi XLVI. törvény a statisztikáról.

Jelen tanulmányban a szerző áttekinti a hivatalos statisztikai szolgálat (HSSZ)³ által használt adatforrás-típusokat, felhasználási területeiket és módjukat.

Az adatforrások típusai

Adatforrásnak nevezzük azt az adathalmazt, metaadat-halmazt, adatbázist vagy metaadat gyűjteményt, amiben a szükséges adatok elérhetők (OECD, 2008).

Az ENSZ 5. statisztikai alapelve kimondja, hogy „statisztikai célokra bármilyen adatforrásból – statisztikai adatfelvételekből vagy adminisztratív nyilvántartásokból – származó adatok felhasználhatók. A statisztikai szervezeteknek a forrás kiválasztásánál figyelembe kell venniük az adatok minőségét, időszerűségét, a költségeket és az adatszolgáltatókra háruló terheket.” (ENSZ, 2014)

Statisztikai adatigények megfogalmazását követően manapság, amikor a statisztikai munkában egyre nagyobb hangsúlyt kap a költséghatékonyság és az adatszolgáltatói – különösképpen a vállalkozásokra nehezedő – terhek csökkentése, első lépésként azt kell megvizsgálni, hogy létezik-e már olyan adatforrás, amelyből a szükséges információk átvethetők. Ennek hiányában indokolt csak új statisztikai adatokat gyűjteni.

Egy statisztikai tevékenységet végző intézmény szemszögéből tekintve attól függően, hogy mely szervezet végezte a kapcsolódó adatgyűjtést, az adatforrások rendszere két részre osztható:

- + az elsődleges adatforrások és
- + a másodlagos adatforrások csoportjára.

Elsődleges adatforrások

Az elsődleges adatforrások kategóriájába a HSSZ minden egyes szervénél a saját statisztikai adatgyűjtései tartoznak. A hivatalos statisztikai rendszeren belül az adatgyűjtések történhetnek kikérdezéssel (pl. kérdőívek alkalmazása) vagy megfigyeléssel (pl. árak összeírása). Abból a szempontból, hogy az adatgyűjtés a megfigyelendő statisztikai egységek milyen körét öleli föl, lehet teljes vagy részleges. Teljes lefedettség esetén a sokaság minden egységét számba veszik, míg részleges lefedettség esetén azoknak csak egy részéről gyűjtenek információt. A részleges megfigyelés egy fontos típusa a mintavételen alapuló, reprezentatív megfigyelés. Teljeskörű adatgyűjtésre a legismertebb példa a népszámlálás, az állatszámmlások és számos intézményi adatszolgáltatás. Sok esetben azonban, amikor nem lehetséges vagy nem célszerű a sokaság valamennyi elemét felmérni, a sokaságnak csak egy részéből vesznek mintát. Néha a a kétféle adatgyűjtési módot kombinálják, ha a sokaság egyes elemeit teljeskörűen felmérik (pl. bizonyos mérethatár felett az összes vállalkozást), a sokaság más elemeiből (pl. a kisebb vállalkozásokból) pedig mintát vesznek.

³ A hivatalos statisztikának formális definíciója nem létezik. Jelen tanulmányban ide értjük a statisztikai törvény hatálya alá tartozó tevékenységet. A statisztikai törvény 3.§ (1) szerint „E törvényben meghatározott statisztikai tevékenység ellátása a hivatalos statisztikai szolgálat feladata.”

Statistikai célú adatgyűjtések rendszere

Magyarországon az adatszolgáltatási kötelezettséggel járó statisztikai adatgyűjtéseket – kivéve a törvényben elrendeltek – a hivatalos statisztikai szolgálat országos statisztikai adatgyűjtési programja (OSAP) tartalmazza. Az OSAP keretében gyűjtendő adatok körét a HSSZ bejelentett adatgyűjtési igényei alapján a KSH évente állítja össze az Országos Statisztikai Tanács (OST) bevonásával. A Sattv.-ben kapott felhatalmazás alapján a kormány az éves OSAP-ot rendelet⁴ keretében jelenteti meg. Az OSAP hatálya a jogi személyekre, a saját joga szerint jogképes szervezetekre, valamint a gazdasági tevékenységet folytató természetes személyekre terjed ki. A 2015-ös OSAP összesen 362 tételt⁵ tartalmaz.

Az Infotv⁶ 5. §-a szerint személyes adat akkor kezelhető, ha ahhoz az érintett hozzájárul, vagy ha azt törvény vagy törvény felhatalmazása alapján helyi önkormányzat rendelete kötelező érvénnyel elrendeli. Ennek megfelelően a népmozgalmi⁷ és a halottvizsgálattal kapcsolatos adatgyűjtéseket nem az OSAP, hanem a Sattv. rendeli el, míg a népszámlálásokat és az általános mezőgazdasági összeírásokat esetileg elfogadott törvények.⁸ A statisztikai adatgyűjtési program részeként minden egyéb, személyes adatra vonatkozó adatszolgáltatás önkéntes. A KSH honlapján és a Hivatalos Értesítőben közzé teszi az általa tervezett önkéntes adatgyűjtéseket. 2015-re tíz ilyen tétel szerepel a listán, melyek munkaerőre, háztartási költségvetésre és életkörülményre, magyar állampolgárságot szerzőkre, utazási szokásokra, infokommunikációs eszközök használatára, bevándorlókra és külföldiek magyarországi kiadásaira vonatkoznak.

Másodlagos adatforrások

Másodlagos adatforrásnak tekintünk minden olyan adatállományt, amely esetében az adatgyűjtő személy, szervezet nem azonos azzal, amelyik elemzi vagy használja. Ez különbözteti meg az elsődleges adatoktól, amelyek esetében a két funkció megvalósítója azonos. Egy adott, statisztikai tevékenységet végző szervezet szempontjából másodlagos adatforrások közé sorolandó valamennyi, más szervezettől átvett adatgyűjtemény függetlenül attól, hogy eredetileg mi volt az adatgyűjtés célja. A másodlagos adatforrásokon belül három nagy csoport különíthető el:

- HSSZ-en belüli (OSAP-ban rögzített) statisztikai adatgyűjtések,
- adminisztratív adatok,

4 288/2009. (XII. 15.) Korm. rendelet az Országos Statisztikai Adatgyűjtési Program adatgyűjtéseiről és adatátvételeiről. A rendeleter évente módosítják, és a mellékletei tartalmazzák a hivatalos statisztikai szolgálat szerveinek aktuálisan érvényes közvetlen adatgyűjtéseit és statisztikai célú adatátvételeit, valamint a gyűjthető adatköröket.

5 Egy tétel alatt egy közvetlen statisztikai adatgyűjtést vagy egy, nem statisztikai célra gyűjtött adatállomány átvételét értjük.

6 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról.

7 Születés, halálozás, házasságkötés, bejegyzett élettársi kapcsolat létesítése, élettársi nyilatkozat nyilvánításba vétele, házasság felbontása, valamint érvénytelenné nyilvánítása, bejegyzett élettársi kapcsolat felbontása, megszüntetése, a nyilvántartott élettársi kapcsolat megszűnése, valamint lakcímváltozás tartozik a népmozgalmi események közé.

8 Pl. 2009. évi CXXXIX. törvény a 2011. évi népszámlálásról; 2010. évi XXIV. törvény a 2010. évi általános mezőgazdasági összeírásról.

- egyéb másodlagos adatforrások,
 - vállalkozások, által gyűjtött adatok,
 - HSSZ-en kívüli statisztikai adatgyűjtések,
 - big data és
 - bármi egyéb.

HSSZ-en belüli (OSAP-ban rögzített) statisztikai adatgyűjtések

A Stattu.-ben megfogalmazott célok elérése érdekében a KSH nemcsak összeállítja az OSAP-ot, hanem egyúttal megszervezi a HSSZ-en belül az OSAP keretében gyűjtött adatok átadását, átvételét is. Ezzel elősegíti a statisztikai céllal gyűjtött adatok hatékony felhasználását. A 2015-ös OSAP-ban szereplő 362 tételből 293 eredményeit nemcsak az elrendelő szervezet hasznosítja, hanem a HSSZ más szerve is átveszi. A két legnépszerűbb adatgyűjtés a KSH 1845. Éves gazdaságstatisztikai jelentése és a 2235. Havi teljesítménystatisztikai jelentése. Ezeket 4-4 más intézmény veszi át. Az aktuális adatátvételi igények a KSH honlapján olvashatók.⁹

Meg kell azonban jegyezni, hogy a HSSZ-en belül csak korlátozottan léteznek olyan közös szabványok, kritériumok vagy ajánlások, amiket a statisztikai munkafolyamatoknak és termékeknek ki kell elégíteni, sőt a HSSZ tagszervezeteinél független statisztikai szervezeti egység is csak elvétve létezik. Ezért az OSAP-adatgyűjtések sem tartalmilag, sem módszertanilag nem képeznek koherens egységet. Másodlagos felhasználásuk során a teljeskörű statisztikai adatgyűjtéseket ugyanúgy kell kezelni, mint egy nemstatisztikai adatállományt.

Adminisztratív adatok

A legerjedtebb meghatározás szerint az adminisztratív adatforrás közfeladatot ellátó szervezet által fenntartott adatgyűjtemény, amit azért gyűjtenek és használnak, hogy eleget tegyenek a közterhekkel, (nyilvántartási, engedélyezési, jogosultsági) eljárásokkal, szolgáltatásokkal, juttatásokkal kapcsolatos feladataiknak. Az adatokat tehát eredetileg jog és kötelezettség keletkeztetésével, az egyénnel szemben valamilyen joghatás kiváltásával összefüggésben gyűjtötték. Az adminisztratív adatok jellegzetessége, hogy gyűjtésük célja a célcsoport teljes lefedettsége, minden egyes egyed megfigyelése, továbbá, hogy a későbbi statisztikai hasznosítás ellenére az adatgyűjtés és a feldolgozás módszerét alapvetően továbbra is az adatgazda szervezet határozza meg (*Brackstone, G. J., 1987*). Az adminisztratív adatok gyűjtését és karbantartását mindig jogszabály rendeli el, ami részleteiben írja le a gyűjtendő adatok körét, a nyilvántartás által használt fogalomrendszert és sok esetben az alkalmazott osztályozási rendszereket is.

Az adminisztratív adat lényege az adat egyedi azonosíthatósága, hisz az adott egyedre származtathatnak következményeket. A statisztikai adat esetén viszont az egyedi azonosítás lehetősége csak eszköz az adatgyűjtés és feldolgozás során, mert a sokaságot kívánjuk jellemezni.

A statisztikai felhasználás szempontjából az adminisztratív adatok erősségét és gyengeségét mutatja az 1. tábla összehasonlítva a népszámlálással és az egyéb statisztikai adatgyűjtésekkel.

⁹ A hivatalos statisztikai szolgálat szerveinek egymás közötti 2015. évi adatátvételi http://www.ksh.hu/docs/hun/info/adatgyujtes/2015/HSSZ_egymuk_adatvetelei.pdf

1. táblázat: A népszámlálás, egyéb statisztikai adatgyűjtések és a népességre vonatkozó adminisztratív adatok mint statisztikai adatforrások összehasonlítása

Szempont		Népszámlálás	Egyéb statisztikai adatgyűjtés	Adminisztratív adatok
1.	Lefedtettség	A népesség teljes lefedettségét célozza.	Bizonyos népességcsoportokat kizárhat.	A célpopulációt a közigazgatás igényei határozzák meg.
2.	Tartalom	Az adatok széles köre kiterjedt keresztosztályozást tesz lehetővé.	Rendszerint csak egy szűk témakörre szorítkozik, de sokkal részletesebben, mint egy népszámlálás.	A közigazgatási cél szempontjából igényelt adatokra szorítkozik.
3.	Fogalmak, definíciók	A társadalmi, gazdasági elemzések elvárásaira alapulnak.		A közigazgatás igényei határozzák meg.
4.	Kistérsületi becslések	A teljes lefedettség eredményeként előállnak.	Általában nem állíthatók elő.	Előállíthatók, ha az egyedi rekordok a kistérsületek szintjén beazonosíthatók.
5.	Minőségellenőrzés	A hibák minimalizálása érdekében megtervezhető.	A kisebb méretet a népszámlálásnál szigorúbb ellenőrzést tesz lehetővé.	A közigazgatási hatóság végzi, és lehet, hogy a kulcsváltozók kivételével nem fordít rá elég figyelmet.
6.	Költség	Drága.	Egy adatgyűjtés viszonylag olcsó, de a két népszámlálás közötti időszak kumulált költsége nagy lehet.	Viszonylag olcsó, ha a kezdeti adatfeltöltés költsége a statisztikát nem érinti.
7.	Gyakoriság	5 - 10 év.	Témától függően lehet éves, negyedéves, havi.	A közigazgatási feladattól függően az éves és a napi többszöri gyakoriság között lehet, sőt akár folyamatos is.
8.	Időszerűség	Néhány hónaptól több év témától és részletzettségtől függően.	Rendszeres adatgyűjtés esetén néhány hét, ad-hoc adatgyűjtés esetén néhány hónap.	A közigazgatási eljárástól függ. Az éves adatok átvételre ellenőrzött formában gyakran csak sok hónap után van lehetőség.
9.	Stabilitás	A változások azoknak a statisztikusoknak a felelőssége, akik a felhasználói igények teljesítéséért is felelnek.	Ismétlődő adatgyűjtés esetén ritka a változás, hogy az időbeli összehasonlíthatóság biztosított legyen.	Jogi, szabályozási vagy eljárási módosítások következtében törtenhet változás.
10.	Adatszolgáltatói terhelés	Súlyos, de ritka.	Általában kicsi, de a mintába választottak számára súlyos.	Nincs további terhelés.

Forrás: Brackstone G J: Issues in the Use of Administrative Records for Statistical Purposes. Survey Methodology, June 1987 Vol. 13, No. 1, pp. 28-43 Statistics Canada

A két kulcsfontosságú előny, vagyis az adminisztratív adatokra épülő rendszer fenntartásának költséghatékonyasága és az adatszolgáltatói terhek csökkentése mellett még továbbiakat is érdemes megemlíteni. Tekintve, hogy az igazgatási adatnak mindig naprakésznek kell lenni, ezért ezek alapján gyakrabban lehet a statisztikai mutatókat előállítani, ami különösen jelentős az olyan komplex és költséges, ezért csak ritkán végrehajtható adatgyűjtések tekintetében, mint a népszámlálás. A rövid távú mutatók esetében is

számottevő előny adódik abból, hogy sok idő nyerhető az igazgatási adatok használatával (ENSZ, 2011).

Egyéb másodlagos adatforrások

Vállalkozások és kutató helyek adatai

Az ENSZ 5. statisztikai alapelve azt jelenti, hogy a statisztika szempontjából a fentiekén kívül akár vállalkozások által gyűjtött adatok és kutató helyek statisztikai adatgyűjtéseinek eredményei is szóba jöhetnek. Az előbbi esetében megoszlik a nemzetközi gyakorlat az adatok elérhetősége szempontjából. A szlovén statisztikai törvény¹⁰ például úgy rendelkezik, hogy a hivatal és az egyéb hivatalos adat-előállítók felhasználhatják az állami és a magánszektor különböző hivatalos és egyéb adminisztratív adatgyűjtéseit (feljegyzések, nyilvántartások, adatbázisok stb.) egyedi azonosításra alkalmas adatait, ha azokat törvény alapján vagy az egyén írásbeli hozzájárulásával vezették. Észtország statisztikai törvénye¹¹ is általános felhatalmazást ad a magán adatforrásokhoz is. Ezeket az adatforrásokat azokban az esetekben lehet statisztikai célokra hatékonyan felhasználni, amikor – hasonlóan az adminisztratív adatokhoz – vagy általános nyilvántartásként egyetlen szervezet vezet, vagy több adatgazda esetén jogszabály rendel el a nyilvántartás tartalmát és a használt fogalmakat, definíciókat és osztályozási rendszereket. Magánvállalkozások által vezetett, statisztikai szempontból felhasználható nyilvántartások a hitel ügynökségek, üzleti elemzők, közművek adata, telefonkönyvek, telefonszolgáltatók, hűségkártya rendszert működtető kiskereskedők stb.

Kutatóintézetek és egyéb kutató helyek adatai (pl. szociológiai felmérések, epidemiológiai felmérések, közvélemény kutatások) ellenőrizhető és megfelelő minőség esetén a statisztikai szervezetek adatforrásai között szintén felhasználhatók, ha a jogszabályi feltételek adottak, illetve az adatvédelmi szabályok nem sérülnek.

Big data

A statisztikai hasznosítási potenciálokat tekintve nem lehet kikerülni, hogy megemlítjük itt az utóbbi idők legnagyobb kihívását, a naponta keletkező, óriási nagyságú digitális adattömeget, a „big data”-t. Az elnevezés arra utal, hogy az egyes adatállományok mérete terabájtban, sőt petabájtban mérhető, és ezért az általánosan használt szoftver eszközök nem képesek elfogadható időn belül átvenni és feldolgozni őket. A big data hatékony kezelése különleges technológiát igényel.

Ez a típusú adat egyrészt a hagyományos üzleti tevékenységek nyomán keletkezik, másrészt olyan újfajta adatforrásokban, mint például a közösségi médiumok. Tipikus big data az elárúsítóhelyeken felszerelt terminálok (POS) és a banki ATM-ek által generált adatok, de a Google-keresések, blog- és Facebook-bejegyzések, lájkok, mobiltelefonok GPS-koordinátái és a YouTube videók is.

A big data túlnyomóan az emberi viselkedés nyomait követi figyelemmel, és az adatok összekapcsolásával és elemzésével a különböző viselkedések előre jelezhetőek, ezért főként

¹⁰ National Statistics Act. <http://www.stat.si/dokument/5186/NationalStatisticsAct.pdf>

¹¹ Official Statistics Act. <https://www.riigiteataja.ee/en/eli/ee/506012015002/consolide/current>

társadalomkutatásra használható. Meg kell azonban jegyezni, hogy a big data-ból természeti jelenségekre vonatkozó információk is nyerhetők, pl. úrfelvételek, meteorológiai álmások mérései vagy egyéb műszeres megfigyelések.

Az adminisztratív adat mint statisztikai adatforrás

Nemzetközi jogszabályok, ajánlások

A skandináv országokban az adminisztratív adatok rendszerszerű felhasználása a statisztikában egészen a 60-as és 70-es évek fordulójáig vezethető vissza (Statistics Finland 2004). A nemzetközi szervezetek (ENSZ, Eurostat) az 1990-es évek közepén kezdtek azzal foglalkozni, hogy hogyan lehetne az adminisztratív adatok statisztikai felhasználását koordinálni. Ezeknek az erőfeszítéseknek az első jeleként értékelhető, hogy az ENSZ-alapelveknek az 1994-ben megjelent első verziója már úgy említi az adminisztratív adatokat, mint a statisztika lehetséges adatforrásait. A közösségi statisztikákról szóló 322/97/EK rendelet és a helyébe lépő európai statisztikákról szóló 223/2009/EK európai parlamenti és tanácsi rendelet egyformán szabályozta a közigazgatási nyilvántartásokhoz való hozzáférést. Bár a jogszabály úgy rendelkezett, hogy a tagországok érintett szervezeteinek olyan mértékben kell hozzáférniük a közigazgatási rendszerükön belül az adminisztratív adatforrásokhoz, amilyen mértékben ezek az adatok az európai statisztikák fejlesztéséhez, előállításához és terjesztéséhez szükségesek. Ezzel együtt a helyi szabályozást a tagállamok hatáskörébe utalta.

A rendelet 2015. áprilisi módosításában az egyik legfontosabb változás az adminisztratív nyilvántartásokhoz való statisztikai célú hozzáférést érintette. A hatályos jogszabály szerint az érintett intézmények jogosultak a közigazgatási nyilvántartásokhoz való azonnali és ingyenes hozzáférésre és azok felhasználására. Adatok átadásakor csatolni kell a metaadatokat is. Kötelezővé teszi a statisztikai szervek bevonását a nyilvántartások kialakításába, fejlesztésébe és megszüntetésébe, valamint az európai statisztikák előállítása szempontjából lényeges közigazgatási nyilvántartásokra vonatkozó standardizálási tevékenységekbe. A statisztikai hivataloknak és a közigazgatási nyilvántartások tulajdonosainak ki kell alakítaniuk a szükséges együttműködési mechanizmusokat.

Az Európai Bizottság 2005-ben ajánlásként jelentetett meg egy szabályzatot „Az európai statisztika gyakorlati kódexe” (ESR, 2011) címmel. A gyakorlati kódex célja az európai statisztikák iránti általános bizalom biztosítása azáltal, hogy megállapítja az európai statisztikák fejlesztésének, előállításának és terjesztésének módját a statisztikai rendeletben meghatározott statisztikai elvekkel és a statisztika legjobb nemzetközi gyakorlataival összhangban. A kódexet 2011-ben megújították, és az új változatban hangsúlyossá vált az adatgazdák és a statisztikai szervezetek közötti együttműködés az igazgatási nyilvántartások statisztikai célú felhasználhatósága érdekében.

A gyakorlati kódex 15 alapelvét 80 ismervben fogalmazták meg részletesen, és az adminisztratív adatforrásokra vonatkozó 9 ismerv a 2., 8., 9. és 10. elvénél jelenik meg az alábbiak szerint.

2. alapelv: Felhatalmazás adatgyűjtésre.

2.2. ismerv: A statisztikai szervezetek jogszabályi felhatalmazással rendelkeznek az adminisztratív nyilvántartások adataihoz való statisztikai célú hozzáférésre.

8. alapelv: Megfelelő statisztikai eljárások.

8.1. ismerv: Ahol az európai statisztikák adminisztratív nyilvántartáson alapulnak, ott az adminisztratív célú alkalmazások definíciói és fogalmai jól közelítik a statisztikai célú fogalmakat.

8.7. ismerv: Az adminisztratív adatok statisztikai céloknak való jobb megfelelése érdekében a statisztikai szervezetek részt vesznek az adminisztratív adatok tervezésében.

8.8. ismerv: A statisztikai szervezetek az adminisztratív adatok tulajdonosaival megállapodásokat kötnek, amelyek tartalmazzák a felek közös elköteleződését az adminisztratív adatok statisztikai célra való használatára.

8.9. ismerv: Az adatminőség biztosítása érdekében a statisztikai szervezetek együttműködnek az adminisztratív adatok tulajdonosaival.

9. alapelv: Az adatszolgáltatói teher nem túlzott mértékű.

9.4. ismerv: Az információgyűjtés megkettőzésének elkerülése érdekében, amennyiben lehetséges, adminisztratív forrásokat használnak.

9.5. ismerv: A többszörös információgyűjtés elkerülése érdekében a statisztikai szervezetek között általánosan elfogadott az adatmegosztás.

9.6. ismerv: Az adatszolgáltatói teher csökkentése érdekében a statisztikai intézmények olyan lépéseket szorgalmaznak, amelyek lehetővé teszik az adatforrások összekapcsolását.

10. alapelv: Költséghatékonyság.

10.3. ismerv: Előremutató lépések történnek az adminisztratív nyilvántartások használhatóságának javítására és a közvetlen adatfelvételeket szolgáló források korlátozására.

Hazai tapasztalatok

A hazai folyamatokat jellemzi, hogy a Stattu. a KSH feladatai között már az 1993. évi elfogadásától felsorolta ugyan az adminisztratív adatgyűjtések fogalmi és osztályozási rendszerének kialakításában való közreműködést,¹² de a hivatal jogosultsága az azonosításra alkalmas, nem saját gyűjtésből származó adatok felhasználására¹³ csak az 1999-es módosítással került be a jogszabályba.

A nemzetközi tendenciákkal összhangban az utóbbi másfél évtizedben hazai szintén is felerősödött az adminisztratív adatok rendszerszerű, statisztikai célú felhasználása. Ezeknek az adatforrásoknak az előtérbe kerülését elsősorban az informatikai fejlődés segítette elő, hiszen kiszorultak a papíron vezetett nyilvántartások az elektronikus rendszerek javára, ami megkönnyítette az adatok mozgatását, kezelését.

A hazai jogrendszer hozzáigazítása a módosított európai statisztikai rendelethez a közeli jövő feladata. Az adminisztratív adatok statisztikai felhasználása szempontjából elsősorban a nyilvántartásokat elrendelő jogszabályokban kell feloldani a statisztikai felhasználás korlátait.

Az OSAP-tételeknek közel háromnegyede közvetlen statisztikai adatgyűjtés, a többi ún. statisztikai célú adatátvétel, amiknek a döntő része adminisztratív adat átvételét jelenti. Ebből azonban nem az következik, hogy a hivatalos statisztikai rendszerben összesen használt adminisztratív adatkörök száma a 100-at sem éri el, mert sok esetben az

¹² 6.§ (1g) pontja.

¹³ 21. § (4) bek.

átadás-átvétel jogi kerete nem az OSAP, hanem egy szerződés vagy megállapodás. Pl. a KSH a nem-statisztikai adatátvételeinek csak 45%-át teszi közzé az OSAP-ban.

Az adminisztratív források minőségének rendszerszerű megközelítése azt igényli, hogy beazonosítsuk és nyilvántartásba vegyük őket. Jelentősen hátráltatja az adminisztratív nyilvántartások statisztikai célú felhasználását, ha ezeknek az adatrendszereknek nincs egy regisztere, ami információt szolgáltat az igazgatásban fellelhető adatok köréről, azok tulajdonságairól.

A nyilvántartások regisztere azonban nemcsak a statisztikai hasznosítás szempontjából fontos, hanem az állami és önkormányzati nyilvántartások együttműködése tekintetében is. Ezért a 2015. január 1-jén hatályba lépett interoperabilitás törvény¹⁴ 2016 januárjára létrehozta a nyilvántartások elektronikus információs rendszerét. A végrehajtásról szóló kormányrendelet¹⁵ szerint a regiszter tartalmazza a nyilvántartás és a nyilvántartó azonosító adatait, a nyilvántartás vezetéséről és adattartalmáról rendelkező jogszabályra hivatkozást, a nyilvántartott adatok megnevezését és a nyilvántartás együttműködésének módját a többi nyilvántartással.

A törvény egyúttal létrehozta az adatmegnevezések jegyzékét is, amely a megnevezéseken kívül tartalmazza a hozzá tartozó értelmezés leírását is. Így a közigazgatás szintjén fokozatosan felépül egyfajta metarendszer.

Mindezek megteremtik a lehetőséget a nyilvántartások monitorozására és rendszeres értékelésére az egyes adatállományoknak a statisztikai adat-előállításban betölthető értéke szempontjából. Ha a tartalmilag alkalmas adatok minősége nem megfelelő, a statisztikusok és az adatgazda közösen kialakíthatják a nyilvántartás fejlesztési koncepcióját azért, hogy az a statisztikai felhasználásra alkalmas legyen.

Az adminisztratív adatok felhasználási területei a statisztika előállításának folyamataiban

Adminisztratív adatok gyakorlatilag az összes statisztikai munkafázisban alkalmazhatók.

♦ Statisztikai regiszter, felvételi keret

A statisztikai regiszter egy adott sokaság teljeskörű nyilvántartása, mely az egyes egységek azonosító és elérhetőségi adatait, valamint a felhasználástól függően egyéb adatokat is tartalmaz. A regiszterek karbantartása főként adminisztratív adatforrásokon alapul. A statisztikai regiszter egyik fő felhasználási módja, hogy mintavételi keretként szolgál a statisztikai adatgyűjtésekhez. Sok esetben az adminisztratív nyilvántartás már önmagában is alkalmas erre a célra.

♦ Statisztikai adatgyűjtések teljes vagy részleges kiváltása

Az adminisztratív nyilvántartásoknak ez a felhasználási módja eredményezi a legtöbb erőforrás megtakarítását, hiszen a legköltségesebb munkafázist, az adatgyűjtést egyszerűsíti vagy rövidíti le, illetve teszi szükségtelenné. Ha adminisztratív adatokkal váltunk ki statisztikai adatgyűjtést, arra is lehet számítani, hogy az elsődleges adatgyűjtő az adatokat bizonyos mértékig már ellenőrizte, tehát az adatelőkészítési munkafázis is olcsóbb

¹⁴ 2013. évi CCXX. törvény az állami és önkormányzati nyilvántartások együttműködésének általános szabályairól.

¹⁵ 142/2015. (VI. 12.) Korm. rendelet az állami és önkormányzati nyilvántartások együttműködésének általános szabályairól szóló 2013. évi CCXX. törvény végrehajtásáról.

lesz. Azt, hogy az adatgyűjtés helyén milyen ellenőrzések és javítások történtek meg, az adatgazdával történő együttműködés keretében lehet megismerni.

- **Editálás és imputálás**

Mind az adminisztratív, mind pedig a statisztikai adatgyűjtési folyamatban történhet olyan hiba, ami rontja az adatok minőségét, vagy adathiányt okoz. Ezek szükség szerinti javítása, pótlása adminisztratív adatállományokból származó kiegészítő információk bevonásával történhet.

- **Közvetett használat a becslésekben**

Az adminisztratív adatok további felhasználási lehetősége a becslések finomítása, rendelkezésre álló adatok súlyozása és az eredmények érvényességének ellenőrzése, validálása.

- **Tájékoztatás kiegészítése, gazdagítása**

Az adminisztratív adatgazda gyakran maga is számol statisztikai mutatókat, amiket a statisztikai szervezet az alapadatok ismerete nélkül a saját tájékoztatási rendszere bővítésére vesz át.

A közvetlen statisztikai adatgyűjtés és a másodlagos adatok felhasználásának feltételei¹⁶

Közvetlen statisztikai adatgyűjtések

A statisztikai adatgyűjtés szabályait a Stattv. tartalmazza. Statisztikai adatgyűjtéseket adatszolgáltatási kötelezettség előírásával, vagy önkéntes adatszolgáltatás alapján lehet végrehajtani. Amint a fentiekben szó volt róla, adatszolgáltatási kötelezettséget az adatszolgáltató típusától függően törvényben vagy az OSAP-ban lehet előírni.

Másodlagos adatok felhasználása

A másodlagos adatforrásokat eredetileg vagy nem statisztikai célra hozták létre, vagy legalábbis nem az újrahasznosító szervezet céljaira. Ezért az újrafelhasználásnak speciális feltételei vannak.

Jogalap

A jogszabályoknak biztosítani kell, hogy a statisztikai szervezet hozzáférjen a más szervezetek által gyűjtött, megfelelő részletezettségű adatokhoz. A statisztikai adatok esetében a Stattv. szerint¹⁷ a hivatalos statisztikai szolgálathoz tartozó szerv az OSAP végrehajtásából rendelkezésre álló adatokat egyedi adat kivételével a HSSZ másik szervének kérésére köteles továbbítani. Hozzá kell azonban tenni, hogy a KSH egyedi adatot is átvehet.

A Stattv. ugyanitt¹⁸ felhatalmazást ad a KSH-nak egyedi adminisztratív adatok statisztikai célú átvételére, „ha törvény eltérően nem rendelkezik.” Ez utóbbi kitétel miatt

¹⁶ A fejezet tartalmilag támaszkodik a szerző szerkesztésében a „Másodlagos adatforrások használata a statisztikában” címmel készülő KSH kiadvány kéziratára.

¹⁷ 21. §

¹⁸ A Stattv.-nek ez a rendelkezése összhangban van az Infotv. 13. § (1) bekezdésével.

az adminisztratív adatok szükséges mértékű eléréséhez a módosított európai statisztikai rendelet biztosítja ugyan a jogalapot, de a jogforrásként megjelenő törvényeket módosítani kell ahhoz, hogy megszűnjenek az európai statisztikai rendelettel jelenleg fennálló ellentmondások, és az információk egyedi azonosításra alkalmas módon, statisztikai felhasználás céljából átadhatók legyenek.

Az egyéb adatforrások tekintetében, ha azok személyes adatot tartalmaznak, az Infotv. korábban idézett 5. §-a a mérvadó, más esetben a Polgári Törvénykönyv magán, üzleti és egyéb titkokra vonatkozó rendelkezéseit kell figyelembe venni.

A közvélemény támogatása

A statisztikai munkában arra törekszünk, hogy a másodlagos adatok egyre nagyobb köre szolgálhasson adatforrásként, amihez a megfelelő jogi szabályozás mellett a közvélemény támogatása is szükséges. Tudatosítani kell a szakmai és civil adattulajdonosokkal, hogy a másodlagos adatforrások statisztikai célú felhasználásával elkerülhetők az átfedő adatgyűjtések, ami költséghatékony és csökkenti az adatszolgáltatási terheket. Ezenközben biztosítani kell a közvéleményt, hogy a statisztikai újrahaznosítás nem jelent megnövekedett adatvédelmi kockázatot.

Egységes azonosító kódrendszerek, adat-összekapcsolás

Mint erről már korábban is volt szó, a másodlagos adatforrások eredeti célja többnyire nem azonos azzal a statisztikai céllal, amire felhasználják, ezért gyakori, hogy egyetlen adatforrás nem biztosítja a szükséges információkat, hanem több adatforrást kell rekord szinten összekapcsolni, amihez elengedhetetlen a célpopuláció elemeinek egyértelmű azonosítása. Egy egységes azonosító rendszer megkönnyíti az adatállomány lefedettségének, teljességének vizsgálatát is. Statisztika szempontból egy ideális helyzetben a nyilvántartások egységes azonosító rendszereket használnak. Ha nincs ilyen, gyakran rendkívül nehéz és fáradságos, olykor lehetetlen hatékony módon összekapcsolni a különböző nyilvántartásokhoz tartozó adatokat.

A statisztikai alkalmazásban alapvetően három fő azonosító rendszer szükségességét szokták kiemelni: a személyeket azonosító kód, az épület-, lakásazonosító kód és a gazdasági szervezeteket azonosító kód, mert ezek a népszámlálási adatok előállításához nélkülözhetetlenek. Valójában azonban minden statisztikai egységnek (település, megye, gépjármű, termék stb.) kell rendelkezni ilyenekkel. Ezek közül a gazdasági szervezetek egységes azonosítására a törzsszám használható, ami összekapcsolja a különböző államigazgatási szervek adminisztratív nyilvántartásait. Az épületek és lakások esetében a legkézenfekvőbb a pontos cím, amit a 2015. január 1-jével létrejött címregiszter biztosít. A személyek azonosítása már lényegesen bonyolultabb, hiszen mint ismeretes, ebben a vonatkozásban több általános kódrendszer létezik (személyi szám, társadalombiztosítási azonosító jel, adóazonosító jel), de a különböző rendszerek elemei nem feleltethetők meg egymásnak, vagyis a különböző állományok rekordjai leggyakrabban a természetes azonosítók (név, nem, születési hely és idő stb.) mentén kapcsolhatók csak össze, amik írása, helyesírása nem mindig következetes.

Átfogó és megbízható igazgatási nyilvántartási rendszerek

Az igazgatási nyilvántartások, adatbázisok országos vagy regionális szinten hasonló tulajdonságokkal rendelkeznek, teljes lefedettségűek, továbbá módszertanilag és minősé-

gileg is általában egységes állományok. Az adminisztratív adatok természetük szerint többnyire pontosak és a célnak megfelelően naprakészek. A statisztikai célú újrashasznosítás éppen ezeket a tulajdonságokat használja ki, mert az ilyen típusú adatforrásokra lehet hatékony módon statisztikai rendszert építeni.

A gazdálkodó szervezetek által az adminisztratív és a statisztikai célra szolgáltatott adatok egyre kevésbé térnek el egymástól, hiszen a nagyobb szervezetek adatszolgáltatása mindkét irányban a könyvelési szoftverrel történik. Jelentősebb eltérésre a lakossági adatok esetén lehet számítani, amiket számos, a nyilatkozás pillanatában fennálló szubjektív tényező is befolyásolhat.

A közigazgatási hatóságok együttműködése

A közigazgatási szervek közös erőfeszítése egy olyan statisztikai rendszer kiépítésére, amely a lehető legnagyobb mértékben használja fel forrásként a nem-statisztikai céllal gyűjtött adatokat, csak akkor lehet valóban eredményes, ha azt egy kormányzati szintű határozott és egyértelmű kötelezettségvállalás is támogat, valamint szoros együttműködés van az érintett hatóságok között. Különösképpen a másodlagos adatforrásokra épülő átfogó, nagy statisztikai adatrendszerek kiépítéséhez (pl. népszámlálás) van szükség arra, hogy a statisztikai hivatal megfelelő alkupozícióval rendelkezzen a többi hatósággal szemben.

Az elmúlt nyolc évben két kormányhatározat¹⁹ szorgalmazta az adminisztratív adatok statisztikai célú hasznosítását. Ezek ráirányították a HSSZ tagszervezeteinek figyelmét az adminisztratív adatok jelentőségére a statisztikai munkában, bővítették a témához kapcsolódó közös ismereteiket, növelték a szakmai egyetértésük mértékét, és hozzájárultak az adatgazda és a felhasználó szervezetek közötti párbeszéd elindulásához, illetve felerősödéséhez.

Statisztikai tájékoztatás és adatvédelem

Stattv. által az adatok publikálására vonatkozóan megfogalmazott alapszabály szerint a hivatalos statisztikai szolgálathoz tartozó szervek által végrehajtott adatgyűjtések eredményei alapvetően nyilvánosak. Nem lehet azonban nyilvánosságra hozni minősített adatot, valamint az adatszolgáltatóval kapcsolatba hozható adatot (egyedi adat).

A statisztikai adatokat hagyományosan statikus táblázatok formájában publikálják. Ennél korszerűbb mód, amikor a felhasználó egy adatkör változóiból önállóan válogatja össze a tábla dimenzióit. Manapság azonban egyre inkább fokozódik a kutatók igénye az ún. mikroadatok – a közvetlen azonosítóitól megfosztott egyedi rekordok – iránt. Míg a táblázatos megjelenítés esetében az adatkiadás előtt lehet biztosítani a felfedés elleni védelmet, a mikroadatokhoz való hozzáférés már bonyolultabb adminisztrációs eljárást tesz szükségessé, illetve biztonságos környezetben előállított kutatási eredmények utólagos adatvédelmi szempontú ellenőrzését.

A különböző felfedés elleni védelmi módszerek eltérő hatást gyakorolhatnak a védendő adatállományokra, az adatok minőségére. A felfedés elleni nagyon szigorú eljárások csökkentik az adatok pontosságát, részletezettségét. A megfelelő módszer megválasztá-

¹⁹ 2212/2007. (XI.14.) kormányhatározat az államigazgatási nyilvántartások adatainak statisztikai célra történő felhasználásáról (hatályon kívül helyezve); 1133/2011. (V. 2.) kormányhatározat a vállalkozói adminisztratív terhek csökkentésére irányuló rövid- és középtávú kormányzati programról.

sakor arra kell törekedni, hogy az felfedési kockázat mértéke és az adatminőségben bekövetkező veszteség egyensúlyban maradjon.²⁰

IRODALOM

- KSH (2014): Módszertani dokumentáció/Fogalmak, definíciók. http://www.ksh.hu/apps/meta.menu?p_lang=HU&p_menu_id=220&p_param=S&p_session_id=82019363
- OECD (2008) OECD Glossary of Statistical Terms. <https://stats.oecd.org/glossary/about.asp>
- ENSZ (2014): ENSZ-alapelvek a hivatalos statisztikára http://www.ksh.hu/nemzetkozi_ajanlasok
- BRACKSTONE, G. J. (1987): Statistical Issues of Administrative Data: Issues and Challenges. In: Statistical Uses of Administrative Data - An International Symposium. Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988).
- ENSZ (2011): Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices. UNITED NATIONS. New York and Geneva. <http://unstats.un.org/unsd/EconStatKB/Attachment442.aspx>
- STATISTICS FINLAND(2004): Use of Registers and Administrative Data Sources for Statistical Purposes Best Practices of Statistics Finland http://www.stat.fi/tup/julkaisut/kasikirjoja_45_en.pdf
- ESR (2011): Az Európai Statisztika Gyakorlati Kódexe. Az Európai Statisztikai Rendszer Bizottsága által 2011. szeptember 28-án elfogadott változat https://www.ksh.hu/az_europai_statisztika_gyakorlati_kodexe

²⁰ A KSH az adataihoz különböző hozzáférési módokat biztosít, hogy a különböző felhasználói igényeket kielégítse pl. előre megadott szerkezetű táblázat (STADAT), felhasználó által definiált tábla (tájékoztató adatbázis és egyéni kérések), nyilvános mikroadat, szerződéssel kiadott anonimizált mikroadat, továbbá kutatószoba, távoli hozzáférés, illetve távoli végrehajtás. A kutatószoba és a távoli hozzáférés esetén a KSH biztonságos környezetben ad hozzáférést kutatás céljából előkészített mikroadat állományokhoz kutatók részére. Ezekben az esetekben mikroadat állományt nem, csak a belőlük származó kutatási eredményt kapja meg a kutató a kutatási eredmények utólagos adatvédelmi szempontú ellenőrzését követően.