

Másodlagos adatforrások használata a statisztikában



Központi Statisztikai Hivatal

**Másodlagos adatforrások használata
a statisztikában**
Általános ismeretek és a hazai gyakorlat

Budapest, 2017

© Központi Statisztikai Hivatal, 2017

ISBN 978-963-235-500-9

Felelős szerkesztő:

Vereczkei Zoltán

főosztályvezető

Készült a KSH Módszertani főosztályán.

Főosztályvezető:

Galántai Katalin, Kovács Marcell, Tokaji Károlyné, Tóth Szabolcs, Vereczkei Zoltán
és Waffenschmidt Jánosné

Összeállította:

Erdei Virág, Gárdos Éva, Györki Ildikó,
Harcza István, Horváth Beáta, Lakatos Miklós,
Nagy Beáta, Nagy Eszter, Szép Katalin,
Szőke Katalin, Vereczkei Zoltán és Víg Judit

Lektorálta:

Bagó Gergely Márk, Csányi Zoltán

Szerkesztő:

Gárdos Éva

Olvasószerkesztő:

Nagy Ildikó

A kéziratot tördelte:

Bada I. Csilla

Másodlagos publikálás csak a forrás megjelölésével történhet!

A tanulmány kutatói véleményeket tükröz, amelyek nem esnek szükségképpen egybe
a KSH hivatalos álláspontjával.

A kiadvány kialakítása egyedi, annak tördelési, grafikai, elrendezési és megjelenési megoldásai a KSH tulajdonát képezik.

Ezek átvétele, alkalmazása esetén a KSH engedélyét kell kérni.

Internet: <http://www.ksh.hu>

(+36-1) 345-6789 (telefon), (+36-1) 345-6788 (fax)

Borítókép: Fotolia

Nyomdai kivitelezés: Xerox Magyarország Kft. – 2017.067

Tartalom

Bevezetés	5
I. A másodlagos adatforrásokról általában	7
II. A nyilvántartások és regiszterek jellemzői és a másodlagos források hasznosításával összefüggő szerepük	28
III. A másodlagos adatok statisztikai felhasználási területei	48
IV. Az adminisztratív adatforrások használatának jogszabályi környezete	52
V. Adminisztratív adatok felhasználása a népszámlálásokban	58
VI. A másodlagos adatok minőségügyi kérdései	80
VII. Együttműködés az adatgazdákkal	92
VIII. A másodlagos adatok korszerű informatikai kezelése a KSH-ban	97
IX. Másodlagos adatforrásból származó adatok felfedés elleni védelme	110
Fogalomtár	118
Felhasznált irodalom	120
Melléletek	123
1. melléklet: A másodlagos adatforrások hasznosítási területei a KSH-ban	124
2. melléklet: A KSH által használt másodlagos adatforrások	137
3. melléklet: Együttműködési megállapodás	149
4. melléklet: A KSH metainformációs rendszere, különös tekintettel az adatátvételekre	160
5. melléklet: A KARÁT-rendszerben képezhető indikátorok	163

Bevezetés

Szerte a világon egyre nagyobb nyomás nehezedik a statisztikai szervezetekre¹, hogy javítsák a statisztikai adat-előállítási folyamat hatékonyságát, különös tekintettel a költségekre és a szakembergárda kihasználtságára. Ugyanakkor a közvélemény részéről fokozódó követelményként jelenik meg az adatszolgáltatói – különösképpen a vállalkozásokra nehezedő – terhek csökkentése. A bürokrácia visszaszorítása hozzájárulhat a cégek eredményesebb üzletmenetéhez, következésképpen az általános gazdasági fejlődéshez is.

Ezzel párhuzamosan az utóbbi két-három évtized ugrásszerű informatikai fejlődése nyomán digitális átalakulás zajlik az egész világban. Egyre nagyobb a rendelkezésre álló adatok mennyisége, és ezek hatékony újrahasznosításához az elektronikus adattovábbítás technikai feltételei is biztosítottak. Ezzel egyidejűleg drasztikusan csökken az információs termékek előállításának, megszerzésének költsége. Mindezek új lehetőségeket jelentenek a nemzeti statisztikai szolgálatok, kiemelten a nemzeti statisztikai hivatalok számára, és a lehetséges adatforrások folyamatosan bővülő köre nemcsak egyszerűen kiválthatja a közvetlen statisztikai adatgyűjtéseket, hanem befolyásolhatja az előállított hivatalos statisztikai termékek összességét és minőségét, valamint az adat-előállítási folyamatokat is.

A hatékonysággal és erőforrás-kímélő módszerekkel összefüggő elvárások a statisztikusokat egyre inkább arra készítetik, hogy a statisztikai adatgyűjtések helyett a már máshol létező és használható adatokat hasznosítsák újra. Sokféle adat gyűlik a közigazgatásban és azon kívül statisztikai és nem statisztikai céllal egyaránt, és bár ezek az adatok ritkán tudják közvetlenül pótolni egy célzott, közvetlen statisztikai adatgyűjtés eredményeit, mégis – esetenként több adatforrás kombinációja révén – lehetőséget kínálnak arra, hogy teljesen vagy részben kiváltsanak egy közvetlen statisztikai adatgyűjtést, illetve hozzájáruljanak más statisztikai tevékenységekhez. A külső adatforrások felhasználásának mértéke a statisztikai szervezetek adat-előállítási folyamatában országonként jelentősen eltér. Némelyek már szinte teljes mértékben nyilvántartás alapú statisztikai rendszert alakítottak ki, míg mások csak a közelmúltban kezdték rendszerszerűen alkalmazni ezt a fajta módszertant. Az Európai Unió is több mint egy évtizede fokozódó aktivitással szorgalmazza a közigazgatásban keletkező, ún. adminisztratív adatok statisztikai célú hasznosítását. Már a közösségi statisztikáról szóló 1997-es tanácsi rendelet is egy rövid bekezdést szentelt a közösségi statisztika előállítása érdekében az adminisztratív adatokhoz való hozzáférés biztosításának. 2005-ben az Európai Bizottság megjelentette „Az Európai Statisztika Gyakorlati Kódexe” című ajánlást, ami már kiterjedtebben foglalkozik a közigazgatási adatok statisztikai felhasználásával. Az adminisztratív adatok használatával kapcsolatban a hozzáféréseken túl további szabályokat fogalmaz meg a fogalmi rendszernek a statisztikai rendszerrel történő összehangolására, az adatszolgáltatói terhek csökkentésére és a költséghatékonyságra vonatkozóan. A 2011. szeptemberi felülvizsgálat során a kódex kibővült azokkal az elvárásokkal is, amiket az adatok hatékony felhasználása érdekében az adminisztratív adatok elsődleges tulajdonosaival szemben támaszthat a statisztikai szervezet.

¹ A továbbiakban statisztikai szervezet alatt a hivatalos statisztikát előállító, fejlesztő, közzevető szervezeteket értjük.

A hivatalos statisztika alapvetően kétféle adatforrásból dolgozik, közvetlen statisztikai adatgyűjtésekből és adminisztratív adatállományokból. A statisztikai célú adathasznosítás szempontjából azonban a legutóbbi évek egyik legnagyobb kihívása a big data, vagyis a rendkívül nagy sebességgel, folyamatosan keletkező digitális adattömeg megjelenése. Bár a hivatalos statisztikai rendszernek még kevés a tapasztalata ezekkel a potenciális adatforrásokkal, semmiképpen sem hagyhatjuk őket figyelmen kívül. Ennek nemcsak az az oka, hogy használatukkal több releváns és időszerű statisztika állítható elő, mint a hagyományos adatforrások alapján, hanem az is, hogy fel nem használásuk közvetett módon kockázatot jelent a nemzeti statisztikai hivatalok megbecsültsége szempontjából.

A statisztikai hivatalokat általában jogszabályok hatalmazzák fel arra, hogy a közvetlen statisztikai adatgyűjtések mellett adminisztratív adatokat is használhassanak. Ez a big data esetében már nem így van, hiszen azok többnyire bárki számára szabadon elérhetőek, vagy magáncégek kezelésében vannak. A privát szektor (vissza)élhet az adatgazdagságából adódó helyzetével egyre több olyan statisztikát állítva elő, amik időszerűség és relevancia tekintetében jobbak lehetnek, mint a hivatalos statisztikák. Attól azonban bizonyára nem kell tartani, hogy a statisztikai hivatalok ezzel elveszítenék a „hivatalos statisztika” védjegyét, hiszen számos olyan garanciát tudnak nyújtani az általuk előállított statisztikákra, amelyeket az előbb említett szervezetek nem (pl. nagy elemszámú mintavételes felvételek és a statisztikai termékek ezzel összefüggő nagyobb pontossága), de a hírnevüket és elismert fontosságukat kockáztatva hátrányos helyzetbe kerülhetnek, ha figyelmen kívül hagyják a big datát. A statisztikai hivatalok olyan infrastruktúrával és szakmai tudással rendelkeznek, aminek köszönhetően a felhasználók részére sokoldalú és sokféle előnyt nyújtó szolgáltatásokat tudnak biztosítani. A hivatalos statisztikák szerepét és jelentőségét a releváns big data típusú adatforrásoknak a hivatalos statisztika adat-előállítási folyamataiba való beépítésével lehet védeni.

Ily módon a hivatalos statisztikák pozíciója minden más adatkezelőhöz képest kedvezőbb lenne, mivel egyszerre tudnák biztosítani az adataik pontosságát, a hivatalos statisztikák komplex rendszerének konzisztenciáját, és megfelelően tudnák értelmezni a nyilvánosságra hozott adatokat is, miközben folyamatosan garantálnák az elvárt relevanciát és időszerűséget.

A KSH statisztikai adatforrásrendszere sokszínű, a 2016 júniusi állapot szerint a közel 160 közvetlen adatgyűjtés mellett mintegy 200 adminisztratív adatforrást vesz át, és 80 azoknak a statisztikai adatgyűjtéseknek a száma, amit más intézmény felügyel, de az adatait a hivatal is hasznosítja. A KSH részére 53 adatgazda, illetve adatgazdacsoport biztosít adminisztratív és egyéb, nem az OSAP keretében gyűjtött adatot, míg statisztikai adatot az OSAP keretében 9 szervezettől vesz át. Az átvett adatok szakstatisztikák szerinti hasznosítását és adatgazdaszervezetek szerinti felsorolását az 1. és 2. melléklet tartalmazza.

Ennek a kiadványnak hangsúlyozott célja, hogy tájékoztatást adjon valamennyi érdekelt szervezetnek és szakembernek a másodlagos adatforrások statisztikai célú hasznosításának módszertanáról, a hatékony megvalósítás feltételeiről és az eddig megtett hazai lépésekről. Az összeállítás kiemelten szól azokhoz a KSH-s munkatársakhoz, akik hivatali munkájuk elején állnak és azoknak a külső intézményeknek az illetékes vezetőihez és ügyintézőihez, amelyek a hivatal adatátadó partnerei. A módszertani leírások fókuszába az adminisztratív adatokat helyeztük mint a statisztikai adatgyűjtésekhez képest bizonyos szempontokból eltérő eljárásokat igénylő adatforrásokat, amiket ugyan évtizedek óta használ a KSH, de tervszerű integrálásukra

csak az utóbbi tíz évben került sor. A big datát – bár kétségkívül nagy lehetőségeket rejt – kevésbé részletesen tárgyaljuk, minthogy rendszerszerű felhasználásuk még széleskörű módszertani és informatikai tervezést és fejlesztést igényel, és a KSH-nak egyelőre csak kevés ide vonatkozó tapasztalata van.

Szándékaink szerint ezzel a kézikönyvvel hozzájárulunk a nemzetközi szakirodalom bővítéséhez is, bemutatva a magyar fejlődés tendenciáit és a továbblépésekhez kapcsolódó speciális polémiákat.

I. A másodlagos adatforrásokról általában

1. Mik azok a másodlagos adatforrások?

A másodlagos és az adminisztratív adatforrásokra a szakirodalomban több meghatározás létezik. A gyakorlatban és a jelen kiadványban is alkalmazkodunk a legelterjedtebb definícióhoz.

Az adminisztratív (igazgatási) adatforrás közfeladatot ellátó szervezet által fenntartott adatgyűjtemény, amelyen keresztül eleget tesznek az adókkal, szolgáltatásokkal, juttatásokkal összefüggő feladataiknak. Ezek a nyilvántartások, adatbázisok olyan adatokat tartalmaznak, amelyek jogszabályban előírt jogok érvényesítéséhez, kötelezettségek teljesítéséhez, gyakran költségvetési bevételekhez és kiadásokhoz kapcsolódnak. Többnyire országos érvényű közfeladatok végrehajtása keretében épülnek fel, tehát az adott igazgatási program szempontjából teljes körűek.

Brackstone² szerint az adminisztratív adatoknak a statisztikai hasznosítás szempontjából négy megkülönböztető jellemzője van:

1. a statisztikai célra adatokat szolgáltató szervezet nem azonos azzal az egységgel, amire az adatok vonatkoznak (ellentétben a legtöbb statisztikai felméréssel);
2. az adatokat eredetileg jól meghatározott, nem statisztikai célra gyűjtötték;
3. célja a célcsoport teljes lefedettsége;
4. az adatgyűjtés és a feldolgozás módszerét a statisztikai hasznosítás ellenére továbbra is a közigazgatási szervezet határozza meg.

A fenti leírás a másodlagos adatforrások hagyományos vagy szűkebb definíciója, ami tehát csak az ún. adminisztratív adatokat fedi le, vagyis azokat, amiket állami szervezetek gyűjtöttek nem statisztikai célra. Egy szélesebben értelmezett definíció magában foglalja a magánszektorban keletkező adatokat is.

Tovább általánosítva a másodlagos adatok fogalmát ide sorolhatjuk mindazon formalizált információkat, amiket más személy vagy szervezet gyűjtött a helyett, amely elemzi vagy felhasználja azokat. Ez különbözteti meg az elsődleges adatoktól, amelyek esetében a két funkciót végző szereplő azonos. A statisztika adatforrásai így elsődleges (statisztikai célra gyűjtött saját adatok) és másodlagos adatforrásokra (minden egyéb adat) oszthatók.

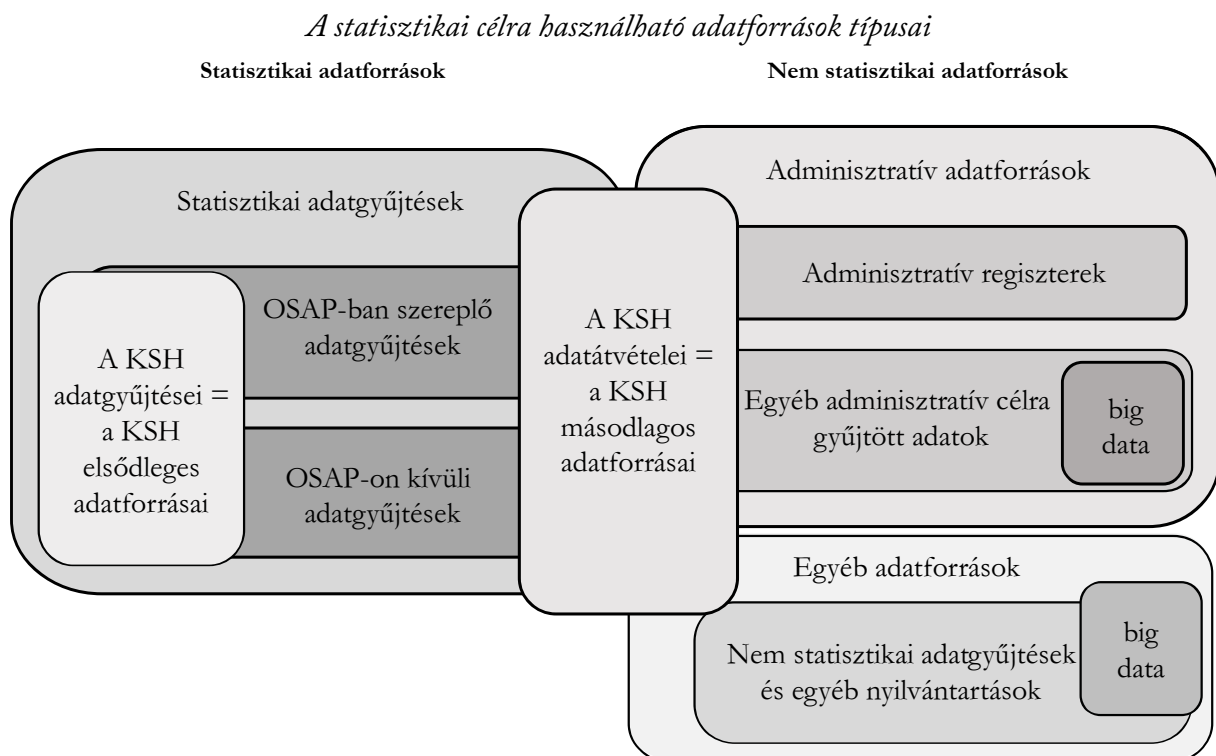
A hivatalos statisztikák Magyarországon a Hivatalos Statisztikai Szolgálat (HSSz) tagjainak felügyelete alatt állíthatók elő. E célból az adatszolgáltatók közvetlen bevonását

² Brackstone G J: "Statistical Issues of Administrative Data: Issues and Challenges", in "Statistical Uses of Administrative Data - An International Symposium", organised by Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988). <http://www.oecd.org/dataoecd/58/16/36237567.pdf>

igénylő, kötelező adatgyűjtéseket csak úgy lehet lebonyolítani, ha azokat törvény vagy az évente megújított és kormányrendeletben közzétett Országos Statisztikai Adatgyűjtési Program (OSAP) rendeli el.

Ha kifejezetten egy szervezet szemszögéből – esetünkben a KSH-éból – vizsgáljuk a másodlagos adat definícióját, már ide kell sorolnunk a hivatalon kívüli szervezetek által statisztikai célra gyűjtött adatokat is, amennyiben ezeket a hivatal további felhasználásra veszi át. Indokolt, hogy ezt a széles értelmezést tekintsük a másodlagos adatforrások definíciójának, hiszen a külső intézmények esetében gyakran összemosódik az adminisztratív és a statisztikai cél, és az ágazati statisztikai indikátorokat eredetileg igazgatási célra gyűjtött adatokból is származtatják.

1. ábra



Az is a szélesebb értelmű definíció használatát támasztja alá, hogy a magánszektorban is egyre több olyan adatbázist alakítanak ki és tartanak karban, amelyek statisztikai hasznosításra alkalmasak, és bár a jogszabályok egyelőre nem támogatják ezeknek a statisztikai adatforrások közé való besorolását, érdemes ezt a lehetőséget is figyelembe venni. Itt elsősorban azokra az adatgyűjteményekre kell gondolni, amiket jogszabályok tesznek a vállalkozások meghatározott csoportjai számára kötelezővé. A szolgáltatók adatbázisai (pl. közüzemi szolgáltatók, telefontársaságok, posta) is az országos, illetve regionális szintű adminisztratív adatokhoz hasonló tulajdonságokkal rendelkeznek.

Megjegyzendő, hogy ha nem egy statisztikai intézmény, hanem egy szakstatisztika szemszögéből vizsgáljuk a lehetséges adatforrásokat, a statisztikai adatforrástípusok köre a fentieknél bővebb. Ebből a szempontból adatforrás lehet bármelyik más szakstatisztika is, továbbá azok a statisztikai regiszterek, amiket nem az adott szakstatisztika állít elő.

A hivatalos statisztika felhasználói egyre több új adatot igényelnek. Statisztikai adatgyűjtések lebonyolítására viszont egyre kevésbé van lehetőség, ezért megnő az egyéb adatforrások jelentősége, közöttük a magánszférában előállítottaké is.

Az utóbbi években mind nyilvánvalóbbá vált, hogy az adatok piaci értékkel bírnak, és szerte a világban egyre több vállalkozás foglalkozik különféle adatbázisok létrehozásával és értékesítésével (pl. marketing célú speciális címlisták összeállítása). Emellett nő az a fajta piaci tevékenység is, melynek során az igazgatási és a magánszférában létrehozott adatok összekapcsolásával új adatállományokat hoznak létre, amiket ellenőriznek, javítanak, olykor validálnak is, és az így értéknövelt adatokat más szervezetek számára értékesítik. Ez a fajta adatforrás érdekes lehet a hivatalos statisztikák előállítói számára is, mivel fennáll annak a kockázata, hogy ezek a magánszolgáltatók olcsóbban és gyorsabban tudnak adatot előállítani, mint a statisztikai szervezetek, gyakran egyszerűen csak azért, mert a költségeket számos ügyfél között tudják szétteríteni, illetve azért, mert kevesebb figyelmet fordítanak a minőségre.

További lényeges szempont, hogy az ilyen adatforrásokat a hivatalos statisztika is eredményesen fel tudná használni, például a saját statisztikai regisztereinek validálására.

A statisztikában napjainkig használt adatok általános jellemzője a strukturáltság, függetlenül azok eredeti felhasználási céljától. Mind az adminisztratív, mind pedig a statisztikai adatokat előre tervezett módon, tartalommal és szerkezetben gyűjtik, dolgozzák fel és tárolják. Ugyanez igaz azokra a hasznosított másodlagos adatokra is, amelyek a magán- és a civil szférából származnak. Modern világunkban azonban egyre több adat keletkezik a weben és a körülöttünk lévő, növekvő számú elektronikai eszközök által működtetett szenzorokban. Ezeknek az adatoknak az óriási mennyisége és keletkezésük rendkívüli sebessége vezetett el a big data fogalmához. A big data olyan adatforrás, ami nagy mennyiségű, sebességű és változatos adatokat szolgáltat. Ezeket a jellemzőket szokás 3V-vel jelölni.^{3 4} A nagy mennyiség egyik jellemzője, hogy az interneten naponta 3 milliárd gigabájtnyi adat termelődik. A sebesség azt jelzi, mennyire gyorsan kell feldolgozni a rendelkezésre álló adatmennyiséget ahhoz, hogy abból értékes információt lehessen nyerni, a változatosság pedig a megjelenő adatok széles választékára utal, a számtalan fájl- és adattípusra, melyek az információt tárolják. A big data többnyire nem strukturált, azaz nincs hozzájuk előre meghatározott adatmodell, illetve ezek az adatok nem illeszkednek jól a hagyományos relációs adatbázisokhoz. Ennek ellenére a hivatalos statisztika szempontjából nagy lehetőségeket rejtenek akár önmagukban, akár hagyományos adatforrásokkal kombinálva. A hivatalos statisztika nyelvére lefordítva a 3V azt jelenti, hogy a nemzeti statisztikai hivataloknak növelni kell a technológiai kapacitást annak érdekében, hogy a big data típusú adatok tárolhatók, feldolgozhatók és elemezhetőek legyenek, azaz, hogy a statisztikai adat-előállítási folyamatban integrált adatforrásként legyenek felhasználhatóak.^{5 6}

³ 3V-vel szokás jellemezni a big data -t: volume=mennyiség, velocity=sebesség, variety=változatosság a big data háromdimenziós modelljének pillérei.

⁴ Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety" (PDF). Gartner. Retrieved 6 February 2001. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

⁵ What does "Big Data" mean for official statistics? UNECE, 2013.

<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170614>

⁶ Szőke Katalin: Big data a statisztikában, KSH, 2015 (munkaanyag).

Annak megválaszolása, hogy mely big data típusú adatok miként hasznosíthatók a hivatalos statisztikában, illetve, hogy hogyan küzdhetők le a hasznosítással kapcsolatos problémák, rendkívül izgalmas feladat. Habár ezek a kihívások nagyrészt még megoldásra várnak, már meglehetősen sok gyakorlati tapasztalat⁷ gyűlt össze, jóllehet vegyes eredményekkel.

2. A másodlagos adatforrások típusai

A statisztikai célokra is felhasználható másodlagos adatforrások potenciális köre nagy, és egyre növekszik. Az alábbi lista az eredetileg nem statisztikai célra gyűjtött adatforrások főbb típusait és tartalmi sokszínűségét igyekszik általánosan megmutatni, azaz nem kötődik egyetlen ország adatrendszeréhez sem, ugyanakkor semmiképpen sem tekinthető teljes körűnek.

2.1. *Adminisztratív adatok*

- Alapnyilvántartások
 - Adó- és vámadatok
 - személyi jövedelemadó,
 - hozzáadottérték-adó (áfa),
 - társasági adók,
 - vagyoadók,
 - import-/exportvámok.
 - Szociális védelmi adatok
 - járulékok,
 - segélyek,
 - nyugdíj.
 - Egészségügyi és oktatási nyilvántartások
 - Személyek / vállalkozások / ingatlanok / járművek nyilvántartási rendszerei
 - népességre vonatkozó regiszterek,
 - vállalkozásokra vonatkozó regiszterek,
 - lakásokra, ingatlanokra vonatkozó nyilvántartások,
 - építési engedélyek,
 - gépjárművek nyilvántartása.
- Egyéb adminisztratív nyilvántartások
 - Személyazonosító igazolvány / útlevel / vezetői engedély,
 - Választói névjegyzék,
 - Mezőgazdasági regiszterek,
 - Önkormányzati nyilvántartások,
 - Engedélyhez kötött tevékenységek, speciális jogosultságok.

2.2. *Magánvállalkozások adatai*

- Hitelügynökségek,

⁷ Lásd: <http://unstats.un.org/unsd/bigdata/>

- Üzleti elemzők,
- Közművek,
- Telefonkönyvek,
- Telefonszolgáltatók és
- Hűségkártyarendszert működtető kiskereskedők stb. adatai.

2.3. Kutatóintézetektől származó adatok

- Szociológiai felmérések,
- Epidemiológiai felmérések,
- Közvélemény-kutatások stb.

2.4. Big data

A big data adatok hasonlóan a másodlagos adatok teljes köréhez az adatgazdák típusai szerint a fenti módon csoportosíthatók. Nem tartozik azonban ide valamennyi másodlagos adat, hanem csak azok, amik megfelelnek a 3V kritériumnak. Ebből a szempontból az alábbi kategóriák definiálhatók⁸:

- hivatalos, ügyviteli adatok, pl.: elektronikus egészségügyi nyilvántartások, kórházi betegforgalom, biztosítási nyilvántartások, banki adatok,
- kereskedelmi vagy tranzakciós adatok, pl.: bankkártya-tranzakciók, online tranzakciók,
- szenzoros berendezések által gyűjtött adatok, pl. műholdképek, forgalomfigyelő rendszerek, meteorológiai állomások adatai,
- nyomkövető eszközök, pl. útvonalkövetési adatok mobiltelefonoktól, GPS-ektől,
- viselkedési adatok, pl. termékekre, szolgáltatásokra vagy egyéb más jellegű információkra vonatkozó online keresések, honlapok látogatottsága,
- vélemények, pl. hozzászólások a közösségi médiában.

3. A másodlagos adatforrások használatának előnyei

Jelen részben jobbra csak a nem statisztikai célra gyűjtött adatok használatának előnyeit tárgyaljuk részletesen, hiszen ez jelenti minden statisztikai szervezet számára a módszertani újszerűséget, amivel kapcsolatban a tapasztalatok rendszerszerű gyűjtését az utóbbi időkben kezdték meg.

Az alig igényel részletes magyarázatot, hogy miért lehet hasznos egy külső szervezet által a statisztika szabályai szerint összegyűjtött és ellenőrzött adatállomány átvétele és továbbhasznosítása. Mindenképpen erőforrás-kímélő eljárás, ha több szervezet tudja használni ugyanazokat a statisztikai adatokat.

3.1. Alacsonyabb költségek

A statisztikai felmérés a legköltségesebb statisztikai tevékenység. Kérdőívet kell kidolgozni, mintát tervezni, mintavételi keretet létrehozni, a válaszadókat fel kell keresni, ösztönözni őket, hogy válaszoljanak, a válaszokat rögzíteni kell és feldolgozni, majd az adatokat ellenőrizni, és

⁸ What does Big Data mean for Official Statistics? (<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170622>)

végül a szükséges statisztikai mutatókat ki kell számítani. Bár az informatikai eszközök mentesítik a statisztikusokat a feldolgozási terhek jelentős része alól, egy statisztikai adatgyűjtés még mindig meglehetősen munkaigényes, különösen a válaszadók megtalálása és az interjúk lebonyolítása, mely munkafázisokat valószínűleg soha nem lehet majd teljesen automatizálni.

Egy hagyományosan lebonyolított népszámlálásnak a mintavételes adatgyűjtésekhez viszonyítva sokkal nagyobb az erőforrásigénye, mert a feladatokat lényegesen nagyobb léptékben kell elvégezni. A hagyományos népszámlálást tartó statisztikai hivatalok saját rendes költségvetésükből nem is tudják fedezni ezt a munkát, ezért a censusokhoz többletfinanszírozást igényelnek. Ez magyarázza azt is, hogy csak ritkán – általában tízévenként – lehet sort keríteni tradicionális módon végrehajtott népszámlálásra, aminek következtében bizonyos statisztikai mutatókat csak ilyen hosszú idő elmúltával lehet újra előállítani, miközben a közvélemény egyre több és frissebb adatra tart igényt. A hagyományos népszámlálások nagy költségigényük miatt ki vannak szolgáltatva a politikai prioritásoknak is.

Az a költség, amit a másodlagos adatforrásokra épített statisztikai rendszer kialakítása igényel, akár el is érheti azt az összeget, ami egy statisztikai adatgyűjtésekre alapozott rendszer létrehozásához szükséges. Ez különösen igaz lehet akkor, amikor big data újrahasznosításáról van szó, hiszen a nagy tömegű, gyorsan érkező adatok fogadásához és feldolgozásához speciális eszközökre van szükség. Nemcsak a megfelelő IT-felszereltség, hanem az adatok megszerzése is költségigényes lehet.

A másodlagos adatokra épülő statisztikai rendszer működési költségei azonban általában jelentősen alacsonyabbak, mint a statisztikai adatgyűjtéseken alapuló. A V. fejezetben bemutatott 1. tábla néhány európai országban 2011 körül végrehajtott népszámlálásokra vonatkozóan szemlélteti, hogy a két módszer a működési költségeket tekintve mekkora különbségeket okozhat. A legtöbb országban a népszámlálások során az adminisztratív adatforrásokat a hagyományos lakossági adatgyűjtéssel ötvözve alkalmazzák, ezért a fajlagos költségek széles tartományban mozognak. Azokban az országokban, ahol alapvetően statisztikai adatgyűjtéssel végezték a népszámlálást, az egy főre eső költség több tízszerese, sőt akár több százszorosa is a szlovéniaiak, ahol a népszámlálás teljes mértékben igazgatási forrásokon alapult, és a költség fejenként mindössze 0,03 USA-dollár volt. Ez tekinthető a legerősebb érvnek amellet, hogy a statisztika minél szélesebb körben használja az adminisztratív adatokat.

A statisztikai hivatalok, így a KSH számára is a másodlagos adatforrásokhoz való hozzáférés legtöbbször ingyenes, különösen akkor, ha az adatok az állami szektorból származnak. Az adatátvétellel járó költségek azonban a statisztikai felméréshez viszonyítva még akkor is lényegesen kisebbek, ha az adatgazda az adatfájl előállításáért díjat számol fel.

Azokban az esetekben, amikor valami ok miatt a statisztikai felmérések nem helyettesíthetők a rendelkezésre álló másodlagos adatokkal, még mindig érdemes lehet átvett adatforrásokat használni a statisztikai regiszterek, mintavételi keretek kialakításához és fenntartásához, mert ezzel is csökkenthetők az általános költségek.

3.2. Kisebb adatszolgáltatói terhek

Bármilyen külső adatforrást hasznosítunk, az elsődleges adatszolgáltatók egyúttal mentesülnek attól, hogy egy újabb adatszolgáltatási kötelezettségnek tegyenek eleget, így a statisztikai adat-előállításnak ez a módja nemcsak költségkímélő, hanem az adatszolgáltatói terhek

csökkentéséhez is hozzájárul. Számos országban – ahogy Magyarországon is – az adatszolgáltatói terhek csökkentése a kormányzat üzleti fejlődést és növekedést ösztönző szakpolitikai törekvése, különösen a vállalkozások vonatkozásában.

A vállalkozások általában megértik azokat az okokat, amiért adatot kell szolgáltatni igazgatási nyilvántartások részére, még akkor is, ha ez többletterheléssel jár. A statisztikai célú adatszolgáltatási kötelezettségeket azonban – mivel gyakran nem látják át azok okait és szükségességét – nem tekintik fontosnak, sőt kifejezetten felháborítóknak tartják, ha már más kormányzati szerv részére is szolgáltatott hasonló tartalmú adatokat.

Mind a kormánypolitikai, mind pedig az adatszolgáltatói megfontolások univerzális jellegét jól mutatja, hogy az adatszolgáltatói terhek csökkentése az uniós szintű szabályok közé is bekerült, és hogy a statisztikai szervezetekre háruló feladatokat az Európai Statisztika Gyakorlati Kódexében is megfogalmazták.

3.3. A statisztikai adatok gyakoribb előállítása

A másodlagos adatok közül elsősorban az adminisztratív adatok, sőt talán még inkább a big data felhasználásának további előnyei közé tartozik, hogy egyes esetekben lehetőség van a statisztikák gyakoribb előállítására anélkül, hogy ez a válaszadók terheit növelné, vagy a költségeket számottevő mértékben emelné. Erre jó példa a finnországi népszámlálási adatok előállítása, ahol igazgatási adatforrásokból évenként számítanak indikátorokat. Ezzel szemben azokban az országokban, ahol a hagyományos módszereket követik, ezt csak öt-tíz évenként tehetik meg.

Az adminisztratív adatokból előállított statisztikák gyakoriságát főként az határozza meg, hogy milyen gyakran frissítik a forrásul szolgáló adatállományt. Nem lehet havi érvényességű mutatókat számítani évenként karbantartott adminisztratív adatokból, míg ez megtehető, ha az aktualizálás folyamatos (mint általában a big data esetében is). Ekkor nem kell szezonális hatással sem számolni, vagy legalábbis elegendő információ áll rendelkezésünkre ahhoz, hogy kiküszöböljük.

Folyamatosan bekövetkező események nyilvántartásai (mint születés, építési engedély kiadása) statisztikai szempontból jelentős rugalmasságot kínálnak, hiszen az eseménnyel gyakorlatilag egy időben történő nyilvántartásba vétel lehetővé teszi, hogy az adatokat akár naponta kigyűjtsék, és a statisztikákat elkészítsék. Ebben a vonatkozásban még nagyobb potenciállal rendelkezik a big data, aminek egyik legfontosabb erénye, hogy az adatok valós idejű viselkedést tükröznek, azaz megfelelő informatikai felszereltség és a big data-ra épülő adat-előállítási folyamat teljes automatizáltsága mellett a megfigyelt jelenségkör változása után néhány percen belül akár elő is állhat a friss indikátor.

3.4. Jobb időszerűség

Az adminisztratív források használatával a statisztikai eredmények időszerűsége a statisztikai adatgyűjtésen alapuló becslésekhez képest kedvezőbb lehet, ha a szükséges változók tekintetében naprakész adatok állnak rendelkezésre. A statisztikai felmérések általában hosszú időt vesznek igénybe a koncepció megfogalmazásától az eredmények előállításáig, különösen az éves vagy az ad hoc adatgyűjtések esetében. Ezért az adminisztratív adatforrások használata különösen a rövid távú mutatók időszerűsége tekintetében jelenthet hatékonyabb megoldást.

Másik terület, ahol az adminisztratív adatforrások használata különösen pozitívan befolyásolhatja az időszerűséget, a mintavételi keretek kialakítása és a statisztikai regiszterek karbantartása. Az adminisztratív információk ugyanis gyakran sokkal naprakészebbek a célcsoportok változásairól (pl. születések és halálozások vagy cégek megalakulása és megszűnése), mint amilyen egy statisztikai felmérésből származó adat valaha is lehet. A valós idejűsége miatt ebből a szempontból kiemelkedő jelentősége lehet a big datának is, amennyiben megfelelően naprakész források azonosíthatók be, és big data eszközökkel statisztikai célra leválogathatók.

3.5. A célsokaság jobb lefedettsége

Az adminisztratív források gyakran teljes vagy majdnem teljes lefedettséget biztosítanak a célsokaság vonatkozásában, míg a mintavételes felmérés közvetlenül gyakran csak a sokaság egy viszonylag kis részére terjed ki. Ezért – ha az adminisztratív adatforrás és a statisztikai adatgyűjtés célsokasága teljes mértékben megegyezik – az adminisztratív adatforrás felhasználásával a teljes körű statisztikai adatgyűjtéshez hasonlóan lényegében nincs mintavételi hiba és nem válaszolás, továbbá pontosabb és részletesebb becslésekre van lehetőség a különböző részpopulációkra, például a kistérségekre, vagy más egyedi jellemzőkkel rendelkezőkre vonatkozóan.

A big data típusú adatforrások esetében az elemszám akár akkora is lehet, mint a vizsgált közösség egésze. Itt azonban az igazgatási adatokhoz képest még nagyobb körültekintéssel kell megvizsgálni, hogy az adatok milyen mértékben fedik le a statisztikai célsokaságot, hiszen a big data jó része irányítás nélkül, spontán módon keletkezik, illetve az értékelést követően bármikor beállhat a big data forrás alapjául szolgáló sokaságban olyan változás, amely torzítást vihet az azon alapuló statisztikai eredményekbe.

3.6. A statisztikai tevékenység kedvezőbb megítélése a közvélemény részéről

A közvélekedés az adatok megosztásáról országonként jelentősen változik különösen azokban az esetekben, amikor különböző kormányzati szervek egymás közötti adatcseréjéről van szó. Amennyiben a közvélemény általában elfogadja, illetve támogatja az adatok megosztását, a meglévő adatforrások jobb kihasználása növelheti a statisztikai hivatal presztízsét, mivel tevékenysége ezáltal hatékonyabbá és költségtakarékosabbá válik.

A nemzetközi tapasztalatok szerint az adatok megosztása a lakosság körében kelthet ugyan nyugtalanságot, de a tény, hogy mindez hozzájárul a kormányzat hatékonyságának növeléséhez, illetve a költségvetés csökkentéséhez, általában kiegyensúlyozza ezt az aggodalmat. Az olyan politikai jelszavak, mint a „hatékony közigazgatás” a lakosság számára gyakran vonzóak, és enyhítik a magánélet megsértésétől való félelmeket. A lakosság véleménye az adminisztratív adatforrások használatáról tehát nagymértékben függ az azzal kapcsolatos tapasztalataitól és attól, hogy azt a kormányzati szervek hogyan mutatják be.

3.7. Egyéb lehetséges előnyök

A big data adatforrásként való használatát további, más adatforrásoknál jellemzően nem tapasztalható előnyök is indokolják. A big data túlnyomó része valós viselkedést tükröz (például

bankkártya-tranzakciók, különböző szenzorok adatrögzítései), így nem kell az önbevallásra hagyatkozni, mint ahogy a statisztikai adatfelvételek esetében vagy a jogkövető magatartásra egyes adminisztratív nyilvántartások vonatkozásában. Adatforrástól függően azonban ezek az adatok is lehetnek torzák, illetve eltúlzottak, vagyis az adatok minőségéről ebből a szempontból is meg kell győződni, mielőtt statisztikai célra felhasználnánk.

A big data adatforrásoknak másik fontos előnye a nemzetköziség, ezáltal országok között is jól összehasonlíthatóak lehetnek.

4. Másodlagos adatforrások használatának feltételei, nehézségei

A fentiekben a másodlagos adatforrások használatának számos pozitívumát soroltuk fel, tudni kell azonban, hogy statisztikai felhasználásuk során felmerülhetnek problémák is. Néhány ezek közül csak egy-egy konkrét adatállományhoz kapcsolódik, de többségük általános jellegű. Az alábbiakban ez utóbbiakról illetve néhány olyan módszerről lesz szó, amiket a nemzetközi vagy a hazai gyakorlat alkalmasnak talált e problémák kezelésére, hatásaik minimalizálására.

Összességében igaz az, hogy a legtöbb probléma, ami a másodlagos adatok használata kapcsán felmerül – hasonlóan a statisztika számos egyéb területén jelentkező nehézségekhez – hatékony tervezéssel és szervezéssel, az adatforrások alapos ismeretével, kreatív gondolkodással, a partnerekkel, valamint más statisztikai szervezetekkel való aktív együttműködéssel, illetve a tapasztalatok kicserélésével megoldható, vagy legalábbis jelentősen csökkenthető.

Azt azonban le kell szögeznünk, hogy a másodlagos adatok használatával járó nehézségek ellenére az előnyök sokkal jelentősebbek, mint a hátrányok.

4.1. Hozzáférés az adatokhoz

Az átvett adatforrások széleskörű statisztikai célú felhasználásának egyik legfontosabb akadály a lehet, ha nem férünk hozzá az adatokhoz. A statisztikai célú hozzáférésnek jellemzően több vetülete van: jogi, politikai, szervezeti és technikai. Ezek közül a jogi feltételek biztosítása a legfontosabb, de a zavartalan adatáramlás, valamint az adatgazda és a statisztikai szervezet között felmerülő lehetséges problémák vagy félreértések kockázatának minimalizálása érdekében a többivel is foglalkozni kell.

Az egyes országok szakembereinek érdemes tanulniuk egymás tapasztalataiból, de nem szabad elfelejtenünk, hogy az egyedi helyzetek és problémák gyakran egyedi megoldásokat igényelnek.

4.1.1. Jogszabályi feltételek

A más szervezettől átvett adatforrások statisztikai célokra történő felhasználásához jogszabályi felhatalmazás szükséges. Azokat a feltételeket, amik mellett a hivatalos statisztikákat előállító intézmények adatot gyűjthetnek – ideértve a külső adatokhoz való hozzáférést és azok felhasználását is – a statisztikai hivatalok szerepét és feladatait meghatározó jogszabályban, jellemzően a statisztikai törvényben fogalmazzák meg.

Az adminisztratív adatokat tekintve a nemzetközi gyakorlatban a jogszabályi feltételeknek két alaptípusa ismert:

1. A statisztikai szervezet jogosult használni a más intézményeknél keletkező adatokat, amennyiben ezt más jogszabály nem tiltja.

2. A statisztikai szervezet általános felhatalmazást kap valamennyi nem minősített adat hasznosítására a szükséges aggregátsági szinten, akár azonosításra alkalmas módon is.

Ha a statisztikai szervezetnek nincs általános felhatalmazása arra, hogy jogszabályban kitűzött feladatainak ellátása érdekében kérésére megkapjon bármilyen titoknak nem minősülő adatállományt, törvények, rendeletek állapítják meg azokat a feltételeket, amelyek mellett statisztikai célra felhasználhatja az egyes adatforrásokat. Emellett az Európai Unió statisztikai rendelete⁹ is előírja a tagállamokra nézve kötelező jelleggel az adminisztratív adatok hozzáférhetőségét.

Statisztikai célú felhasználás esetén az egyes országok jogi keretei különböző mértékű hozzáférést biztosítanak az adminisztratív adatokhoz. A történelmi, politikai és kulturális tényezők, hagyományok jelentősen befolyásolják a kialakult gyakorlatot, aminek eredményeként eltérő társadalmi környezetben különbözőképpen ítélik meg a kormányzat szervei és hivatalai közötti adatmegosztást.

Az európai statisztikai kötelezettségek teljesítése érdekében az Európai Unió statisztikai rendelete a 2015-ös módosítás óta biztosítja a tagállamok nemzeti statisztikai szervezetei számára az ingyenes és azonnali hozzáférési jogot az egyedi azonosításra alkalmas adminisztratív adatokhoz. Az Európai Unió egyes konkrét szakstatisztikai területeket szabályozó rendeleteiben azonban már ezt megelőzően is továbbment, és korlátlan hozzáférést biztosított minden szükséges adminisztratív forráshoz. Erre példa a gazdasági szervezetek regiszterére vonatkozó rendelet.

Bár az adminisztratív adatokhoz való hozzáférés az európai uniós szabályozás alapján az EGT-régió országaiban a közeljövőben megoldódik, a vállalkozási és civil szféra adatainak felhasználása a legtöbb országban – köztük Magyarországon – továbbra is problémás marad. Vannak ugyan jó példák is e tekintetben. Néhány országban (pl. Szlovénia, Észtország, Kanada) a statisztikai törvény nemcsak a közigazgatási célú adatokhoz biztosít hozzáférést, hanem a vállalkozások adminisztratív adataihoz is. Általában azonban adatvédelmi szempontok nehezítik az ezekhez az adatokhoz való hozzájutást. Mindez egyúttal azt is jelenti, hogy a big data típusú adatok nagy részének hasznosítását más problémák mellett a jogi háttér sem segíti.

Az adminisztratív adatok statisztikai célú felhasználásának jogszabályi környezetéről a kiadvány IV. fejezete szól.

4.1.2. Elvi feltételek

A jogszabályi feltételek mellett sok országban létezik olyan, a közfeladatot ellátó szervezetek közötti adatmegosztás alapelveit meghatározó kormányzati stratégia, ami befolyásolhatja a statisztikai szervezet nem statisztikai adatokhoz való hozzáféréseinek körülményeit is. Fontos, hogy a statisztikai hivatal a számára kedvező feltételek biztosítása érdekében részt vegyen minden olyan tárgyaláson, egyeztetésben, amely ezt a stratégiát érinti.

Nemzetközi szinten vizsgálva a másodlagos, különösen az adminisztratív adatok statisztikai célú felhasználásával összefüggő elvi feltételeket két dokumentumot kell mindenképpen megemlíteni. Az ENSZ 1992-ben elfogadott „A hivatalos statisztika alapelvei” és az EU 2005-ben kiadott „Az Európai Statisztika Gyakorlati Kódexe” tartalmazza a hivatalos

⁹ Az Európai Parlament és a Tanács 223/2009/EK rendelete az európai statisztikákról.

statisztikával szemben támasztott elvárásokat. Mindkét dokumentum kimondja, hogy a statisztikai hatóság jogosult arra, hogy hozzájusson minden olyan adathoz, amely feladatainak elvégzéséhez szükséges. Legfontosabb indokként a dokumentumok a költséghatékony működést és a válaszadói terhek csökkentését nevezik meg.

A statisztikai törekvéseket szolgálja, ha a hivatalos statisztikák előállításáért felelős szervezetek nyilvánossá teszik a tevékenységüket meghatározó alapelveket, amelyekben többek között garanciát vállalnak arra, hogy a személyek és a vállalkozások adatait csak meghatározott célokra használják. A KSH minőségi irányelvei tartalmazzák ezeket a biztosítékokat. A dokumentum a KSH honlapján elérhető.¹⁰

4.1.3. Szervezeti feltételek

Miután a jogi és elvi feltételek kialakításával lehetővé válik valamely másodlagos adatforrás használata, olyan szervezeti kereteket kell biztosítani, amelyek elősegítik az adatgazda és a statisztikai szervezet közötti adatáramlást. Ennek eszköze jellemzően egy dokumentum lehet, egy szerződés, vagy egy megállapodás. Az előbbi jogilag kötelező erővel bír, míg az utóbbi inkább a két fél őszinte jó szándékára alapoz.

Bármelyik formát is választjuk, vannak bizonyos alapvető szempontok, amelyeket a dokumentumban rögzíteni kell. A megállapodás egyéb részletei mellett meg kell nevezni azokat a metaadatokat és lehetőség szerint az adatok minőségére vonatkozó információkat is, amik az adatok átvételéhez, értelmezéséhez és feldolgozásához szükségesek. A KSH gyakorlatában szokásos együttműködési megállapodást a jelen kiadvány VII. fejezete ismerteti.

4.1.4. Technikai feltételek

A technikai feltételek alatt azokat az eljárásokat értjük, melyek segítségével az adatok, valamint minden, az adatok felhasználásához, értelmezéséhez szükséges információ továbbítása zajlik. Az adatátviteli mechanizmus többféle lehet a postán küldött papíralapú iratoktól egészen a biztonságos elektronikus kapcsolaton keresztül elérhető valós idejű adatfrissítésekig. A használandó módszer kiválasztása során figyelembe kell venni azokat a műszaki lehetőségeket, amik mind a küldő, mind pedig a fogadó szervezet számára elérhetők. Ezek mérlegelésének eredménye gyakran csak kompromisszumot tükröző, legalább az egyik szervezet szempontjából az optimálist csak megközelítő megoldás lehet.

Számos nemzetközi szabvány létezik az adatok és metaadatok továbbítására. Fontos tehát, hogy a felek megállapodjanak arról, hogy milyen szabványt fognak használni.

Az IT-eszközök kialakításának szempontjait, a lehetséges és nemzetközileg alkalmazott módszereket, valamint a KSH-ban kifejlesztett eszközt a VIII. fejezet mutatja be.

4.2. Átfogó és megbízható igazgatási nyilvántartási rendszerek

A statisztikai felhasználás szempontjából átfogó, megbízható, pontos és a célnak megfelelően naprakész igazgatási nyilvántartásokra van szükség, mert ezek egyrészt lefedik az ország egészét, másrészt módszertanilag és minőségileg is egységes állományoknak tekinthetők. Széttöredezett,

¹⁰ http://www.ksh.hu/docs/bemutakozas/hun/minosegi_iranyelvek_2014.pdf

részenként jelentősen eltérő módszerekkel előállított adatforrásokra nem lehet hatékony statisztikai rendszert építeni.

Bármilyen pontosnak is tekinthető egy nyilvántartás, az ott rögzített adatok bizonyos mértékig mégis eltérhetnek attól, amit egy statisztikai adatfelvétellel gyűjtenénk. Ez azonban nem kérdőjelezi meg a másodlagos adatforrások statisztikai célú hasznosítását, hiszen lényegében minden statisztika, bármely adatgyűjtésből is származzon, csak közelíti a valóságot, és a torzítás mértéke a vizsgált jelenségtől, valamint a felhasznált adatforrás jellegétől függ.

A gazdálkodó szervezetek esetében az adminisztratív és a statisztikai adatok közötti differencia oka egyre kevésbé az, hogy az adatszolgáltató másként nyilatkozik a statisztikai adatgyűjtés részére, mint más hatóságnak, hiszen a nagyobb szervezetek esetén a könyvelési szoftver szolgáltatja az adatokat mindkét irányban. Nagyobb eltérésre a lakossági adatok esetén lehet számítani, amiket számos, a nyilatkozás pillanatában fennálló szubjektív tényező is befolyásolhat.

4.3. Megfelelő koncepció és tartalom

Mielőtt egy adatforrásról el lehetne dönteni, hogy statisztikai hasznosításra egyáltalán alkalmas-e, illetve hogy milyen feltételek mellett, milyen célra használhatók az adatok, nagyon alapos vizsgálódás szükséges, amit lehetőleg az adatgazda bevonásával kell elvégezni. Részletesen meg kell ismerni az adatok tartalmát, a változók közötti összefüggéseket, már csak azért is, mert az átvett bonyolult adatrendszerekből sokszor nehéz kinyerni a statisztika igényeinek megfelelő információkat.

4.3.1. A célsokaság egysége

Ha másodlagos adatforrást használunk, az egyik gyakori probléma lehet, hogy az ott használt egységek nem felelnek meg a statisztikai célsokaság felhasználó által megfogalmazott definíciójának. Az átvett adatállomány egységeit (jogi egységek, adózási egységek, szolgáltatások igénylői stb.) statisztikai egységekké (vállalkozások, emberek, háztartások stb.) konvertálni fogalmi szempontból sokszor meglehetősen nehéz, és modellezési feladatokkal jár együtt. A vállalkozásstatisztika ezt az eljárást profilképzésnek¹¹ nevezi, ami lényegileg a gazdasági szervezetek statisztikai regiszterén alapul.

A profilképzést más szakstatisztikák is alkalmazzák, és annak ellenére, hogy általában nem egyszerű, vannak olyan területek, ahol a folyamat nagy részét automatizálni lehet. Ilyen például a háztartás meghatározása az egy lakcímen élők közötti kapcsolatok alapján. A skandináv országokban ezt a módszert sikerrel alkalmazzák az adminisztratív adatokon alapuló népszámlálásokban is.

A statisztikai egységek kialakítására a profilalkotás mellett bizonyos esetekben léteznek más, egyszerűbb módszerek is. Ezek között érdemes megemlíteni azt, amikor az adminisztratív és a statisztikai definíciók közötti különbséget statisztikai kiigazítással, vagyis további információk bevonásán alapuló becsléssel, súlyozással korrigáljuk.

¹¹ Business Registers – Recommendations Manual, Eurostat Methodologies and working papers, 2010.
<http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf>

Itt fel kell hívnunk a figyelmet arra, hogy a big data használatával összefüggésben az egyik legnagyobb kihívás, hogy az ebbe a körbe tartozó adatok jelentős része spontán keletkezik, esetükben nincs adatmodell, nincs vagy nem megfelelő bizonyossággal definiálható a célsokaság, és így nem beszélhetünk lefedettségről sem, reprezentatív mintát sem választhatunk ezekből az adatokból.

4.3.2. *A változók definíciója*

A másodlagos adatforrások használata esetén nagy valószínűséggel számolnunk kell azzal is, hogy a statisztikai definícióhoz képest eltérő változók szerepelnek az állományban. A strukturált másodlagos adatokat eredetileg egyedi igazgatási, üzleti vagy egyéb, az újrahasznosításával nem egyező statisztikai célra gyűjtik, és az ehhez kapcsolódó igények és prioritások eltérhetnek a fogadó rendszer számára szükségessé válóktól. Klasszikus példa erre a munkanélküliség meghatározása, melynek standard statisztikai definíciója az alábbi¹²:

Munkanélkülinek tekintendő az a személy, akire egyidejűleg érvényesek a következő feltételek:

- az adott héten nem dolgozott (és nincs olyan munkája, amelytől átmenetileg távol volt),
- aktívan keresett munkát a kikérdezést megelőző négy hét folyamán,
- rendelkezésre áll, azaz két héten belül munkába tudna állni, ha találna megfelelő állást vagy talált már munkát, ahol 90 napon belül dolgozni kezd.

Az adminisztratív adatforrásokban azonban a munkanélküliségre használt meghatározás a munkanélküli ellátás igénylésén, illetve a munkanélküli nyilvántartásba vételen alapul. Vannak emberek, akiknek ugyan nincs munkájuk, de nem regisztráltatják magukat munkanélküliként. Másrészt a munkanélküli ellátást igénylők egy része nem tudna munkát vállalni, vagy nem is keres aktívan munkát, ezért annak ellenére, hogy regisztráltatják magukat, statisztikailag nem tekinthetők munkanélkülinek.

Az eltérő definíciókból fakadó probléma megoldásának első lépése, hogy megértsük a különbségeket, és a hatásaikat számszerűsítsük. Néhány esetben ezeknek az eltéréseknek gyakorlatilag nincs befolyásuk a statisztikai eredményre, ezért figyelmen kívül lehet hagyni őket, máskor viszont szisztematikusnak tekinthető a különbség, ami statisztikai kiigazítással korrigálható. Bizonyos esetekben van mód arra is, hogy az eredeti meghatározást módosítsák úgy, hogy az már az új statisztikai felhasználás igényeit is figyelembe vegye.

4.3.3. *Osztályozási rendszerek*

Ahogy a másodlagos adatforrásokban a változók definíciója eltérhet az átvevő szervezet által használt definíciótól, az osztályozási rendszerek is különbözőek lehetnek. Mivel a két adatkezelő célja az adatokkal nem azonos, elképzelhető – még abban az esetben is, ha a két adatrendszer meghatározásai megegyeznek –, hogy a külső adatforrásban az osztályozási rendszert másképpen alkalmazzák, mint ahogy azt az adatokat újrahasznosító szervezet teszi. Például egy vállalkozás gazdasági tevékenységét tekintve az adminisztratív nyilvántartást vezető szervezet számára inkább az a fontos, hogy megvan-e a szükséges engedélye, eleget tesz-e a munka-, vagy

¹² KSH, Módszertani dokumentáció / fogalmak.

http://www.ksh.hu/apps/meta.objektum?p_lang=HU&p_menu_id=110&p_almenu_id=201&p_ot_id=200&p_level=1&p_session_id=75633295&p_obj_id=4282

a környezetvédelmi kötelezettségeinek, míg statisztikai szempontból a vállalkozás fő gazdasági tevékenysége a lényeges.

Más esetekben a másodlagos adatforrásban ugyanazt az osztályozást kisebb részletezettséggel alkalmazzák, mint ami az újrahasznosítási célokra megfelelne, és az is előfordul, hogy maga az osztályozás egyszerűen nem fontos a nyilvántartó szervezet számára, ezért nem fordít az adat pontosságára kellő figyelmet, ami pedig rontja a minőséget.

Amennyiben az adatgazda és az adatátvevő statisztikai szervezet más osztályozási rendszert, vagy annak eltérő változatát használja, az adatforrás és a statisztikai besorolás elemeinek megfeleltetésére a rendelkezésre álló adatoktól és erőforrástól függően több módszer is létezik. Első lépésként azonban mindenképpen fontos részleteiben megismerni és megérteni, hogy az eredeti adatállomány részére a besorolandó adatokat hogyan gyűjtik és dolgozzák fel, és milyen céllal, illetve feladathoz használják.

Ha az újrahasznosító statisztikai rendszerben a kódolási minőség rosszabb, mint az eredeti, fennáll annak a veszélye, hogy az adatokban torzulás jelenik meg. Ahogy több ország sikeres gyakorlata is bizonyította, ezt a torzítást közös automatikus kódoló szoftver használatával lehet kiküszöbölni. Ez a technika nagyfokú konzisztenciát biztosít a kódolásban, és erőteljesen ösztönzi az érintett szervezeteket egy közös osztályozási rendszer kialakítására, bár ez nem mindig valósítható meg.

A közös kódolási eszköz használata mellett még az is növelheti az adatok konzisztenciáját, ha az adatgazdák részére az újrahasznosító statisztikai szervezet szakmai támogatást és képzést biztosít. Ugyanakkor hangsúlyozni kell, hogy az mindig hasznos lehet, ha az adat-előállítók és -felhasználók egyformán ismerik egy közös osztályozási rendszer használatának előnyeit, és az is segít, ha a statisztikai osztályozási rendszer minden változásáról időben értesítik az adatgazdákat, akik számára a változások bevezetése idején a lehető legtöbb segítséget biztosítanak.

4.3.4. Időszerűség

Ahogy erről korábban szó volt, az adminisztratív adatok egyik lehetséges előnye a statisztikai adatokhoz képest a jobb időszerűségük. Mégis, az időszerűséggel kapcsolatban előfordulhat három olyan probléma, ami megkérdőjelezheti a másodlagos adatforrások használhatóságát:

- az adatátvevő igényeinek megfelelő időben nem áll rendelkezésre az adat,
- a másodlagos adat vonatkozási ideje nem felel meg az adatátvevő elvárásának,
- az egyik rendszer adata tartamspecifikus, míg a másiké időpont-specifikus.

Ezek háttérben sokkal inkább az áll, hogy a másodlagos adatokat eredetileg nem a statisztikai felhasználás céljára gyűjtötték, és kevésbé van jelentősége annak, hogy adminisztratív, statisztikai vagy más fajta adatállományról van szó.

Adminisztratív adatokkal összefüggésben az első esetet tekintve gyakran előfordul, hogy egy esemény rögzítése a nyilvántartásban a valós bekövetkezésénél később történik meg, és utána még eltelik bizonyos idő addig, amíg a statisztikai szervezet hozzájuthat az adathoz. Mint ahogy már többször is utaltunk erre, itt is hangsúlyozni kell, hogy az időeltolódás elemzése a legfontosabb eszköz ahhoz, hogy el lehessen dönteni, hogy az adatok közvetlenül használhatóké, vagy szükséges valamilyen kiigazítást végezni.

Az időbeli csúszás különösképpen a rövidtávú statisztikák összeállításában okozhat problémákat, hiszen például egy havonta számított indikátor esetében hathónapos időeltolódás nagy valószínűséggel nem fogadható el, de egy éves mutató esetén már megfelelő lehet.

Meg kell vizsgálni, hogy mi okozza az időeltolódást, mert esetleg van lehetőség arra, hogy az adatgazda szervezet megváltoztassa az eljárást, lerövidítse az esemény és az adatátadás közötti időt úgy, hogy az mindkét szervezet szempontjából elfogadható legyen.

Az egyéb problémákhoz hasonlóan sok esetben a vonatkozási idők eltérése is kezelhető egyedi megoldással. Általános lehetőségként a súlyozás módszere javasolható.

Az időszerűséget befolyásoló három helyzet közül talán a tartamra és időpontra vonatkozó adatok közötti eltérésekből adódó problémák megoldása a legbonyolultabb. Kiigazítást kell végezni, ami egy matematikai modellen vagy egy viszonylag kis mintán történő statisztikai adatfelvétel eredményein alapulhat.

4.3.5. Konzisztencia az adatforrások között

Egyetlen másodlagos adatforrás adattartalma az esetek nagy részében nem felel meg teljes mértékben egy adott statisztikai cél igényeinek. Hatékony felhasználásuknak az a feltétele, hogy egyszerre több adatforrást is tudjunk használni, melyek szükség esetén összekapcsolhatók. Sokszor azonban a különböző adatforrások egymásnak ellentmondónak tűnő adatokat tartalmaznak, ami abból adódhat, hogy más a célsokaságuk, a vonatkozási idejük, a változók definíciói, vagy az osztályozási rendszerük, de akár az is előfordulhat, hogy valamelyikük hibás.

Ilyen helyzetben megoldás lehet, ha változónként prioritásokat állapítunk meg az adatforrások között attól függően, hogy melyikük bizonyul a megbízhatóbbnak. Ha adott változóhoz megtörtént egy ilyen sorrend felállítása, magasabb prioritású adatforrásból származó adatot nem írhat felül egy alacsonyabb prioritású. A prioritások megállapításánál figyelembe kell venni az adat érvényességi idejét, hiszen bármilyen megbízható legyen is egy adatforrás, ha nagyon régi adatokat tartalmaz, valószínű, hogy egy frissebbet kell a sorrendben előrébb helyezni, még akkor is, ha ennek az adata a saját vonatkozási idejét tekintve kevésbé pontos. Az adatforrások közötti prioritások eldöntéséhez változónként meg kell vizsgálni azok viszonylagos minőségét is. Minél több adatforrást használunk, az értékelési folyamat annál bonyolultabb, de több adatforrást felhasználva nagyobb biztonsággal lehet az adatokat validálni. Ennek eredményeként esetenként születhet olyan döntés, hogy bizonyos adatforrásokat statisztika előállítására közvetlenül nem szabad felhasználni. Azonban az ilyen adatforrások is felhasználhatók lehetnek a statisztikai minőségbiztosítás részeként a statisztikai felhasználásra alkalmasnak ígérkező adatok értékelésére. A különböző adatforrások minőségéről ily módon összegyűjtött ismereteket érdemes megosztani az adatgazdákkal is, mert ezáltal javulhat az adatforrás minősége is. Az információ közlése azonban általában csak a hiba természetére vonatkozhat, vagy aggregált szinten történhet, hogy az adatvédelmi szabályokat ne sértsük meg.

4.3.6. Azonosító kódrendszerek

Több adatforrás egyidejű és azonos célú használata esetén gyakran szükségessé válik azok rekord szintű összekapcsolása, amihez elengedhetetlen a célpopuláció elemeinek egyértelmű azonosítása. Egy egységes azonosító rendszer ezen felül az adatállomány lefedettségének, teljességének vizsgálatát is megkönnyíti. A statisztika szempontjából ideális helyzetben minden

másodlagos nyilvántartás egységes azonosító rendszereket használ. Ha nincs ilyen, gyakran rendkívül erőforrásigényes, fáradságos és nehéz, sőt lehetetlen a különböző nyilvántartások adatait összekapcsolni.

A statisztikai alkalmazásban – ezek közül is kiemelten a népszámlálások szempontjából – alapvetően három fő azonosító rendszer szükséges:

- gazdasági szervezeteket azonosító kódok,
- épület-, lakásazonosító kódok és
- személyeket azonosító kódok.

Azokban az országokban, ahol a gazdasági szervezetekre létezik közös azonosító rendszer, vagy ahol különböző hivatalos azonosító rendszerek működnek ugyan, de azok könnyen összekapcsolhatók, az adminisztratív és statisztikai adatforrások integrálhatók. A szakemberek szerint¹³ az lenne előnyös, ha minden országban a statisztikai hivatalok határoznák meg a gazdasági szervezetek egyedi azonosító számát, és ezt az azonosítót használnák a teljes közigazgatási rendszerben. A magyarországi gyakorlatban az ún. törzsszám a közös azonosító, ami eleme mind az adóhatóság által kiadott adószámnak, mind pedig a KSH által kiadott statisztikai számjelnek is.

Az épületek és lakások azonosítására a legkézenfekvőbb a pontos cím, illetve az utóbbi időben az ún. geokódok, azaz olyan adatok összessége, amik alapján egy cím helye meghatározható a térképen. A címmel történő azonosításhoz elengedhetetlen egy egységes címszabvány alkalmazása, ami egyértelműen definiálja, hogy egy-egy címnek milyen elemeket kell kötelezően tartalmaznia. Magyarországon 2015. január 1. óta létezik¹⁴ országos címszabvány, de beépítése valamennyi, címet tartalmazó nyilvántartásba még bizonyára hosszú időt vesz igénybe, és számos intézménynél a nyilvántartó rendszer fejlesztését is igényli.

A legtöbb problémát a személyek egységes azonosítása jelenti, illetve az, hogy ha létezik erre alkalmas kódrendszer, azt a hivatalos statisztika jogosult-e használni. A nemzetközi gyakorlatban sokféle megoldás létezik. Az európai országok többségében ugyan kialakítottak egységes személyazonosító rendszert, de ezeknek csak kis részében építik be minden személyi adatot tartalmazó adminisztratív adatállományba. Az észak-európai országokban és Szlovéniában az egységes személyazonosító minden releváns igazgatási adatbázisnak eleme, és a statisztikai hivatalok felhasználhatják a külső szervtől átvett adatállományok összekapcsolásához, Észtországban egy ilyen eljáráshoz eseti engedély szükséges egy speciális hatóságtól, míg Bulgáriában a statisztikai hivatal nem jogosult használni az egységes személyazonosító rendszert.

Magyarországon a 70-es évek második felétől minden lakos rendelkezik egy egységes szabályok szerint képzett személyi számmal, de 1991-től kizárólag a népesség-nyilvántartás alkalmazhatja, ugyanakkor a társadalombiztosítási és az adózási célokra új azonosítási kódszámrendszereket dolgoztak ki. Alapszabály, hogy a különböző azonosító rendszerek elemei nem feleltethetők meg egymásnak, vagyis az adóadatok közvetlenül nem kapcsolhatók össze a társadalombiztosítás adataival. Emiatt mindenkit legalább három különböző kódszám azonosít, sőt számos további személyazonosító rendszer is létezik (pl. az orvosi pecsétszám, az oktatási

¹³ Use of Administrative Sources for Business Statistics Purposes - Handbook on Good Practices, Eurostat, 2000.

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/CA-24-99-897/EN/CA-24-99-897-EN.PDF

¹⁴ Lásd: 345/2014. (XII. 23.) Korm. rendelet a központi címregiszterről és a címkezelésről.

rendszerben a tanulók, illetve a pedagógusok személyazonosító kódja, a bankoknál, illetve a biztosítóknál használt kliensazonosító).

A személyazonosító kódrendszereknek ez a „sokszínűsége” jelentős akadályt gördít a lakossági adatállományok statisztikai célú összekapcsolása elé. Sok esetben pusztán a személyi adatok szolgálnak kapcsolati kódként. Informatikai eszközökkel azonban az elírások, az állományonként eltérő írásmódok nehezen kezelhetők.

Egyes big data jellegű adatok, mint például a közösségi hálók adatai, szándékosan nem azonosíthatók be, ezért ezek összekapcsolhatósága más adatokkal, és így a statisztikai hasznosításuk is erősen korlátozott a többihez képest.

4.3.7. Hiányzó adatok

Az adatforrásokban az adathiány oka a válaszadás teljes vagy részleges elmaradása, de az is előfordulhat, hogy az adatkezelés során törölték az adatokat. Az adminisztratív, illetve ügyviteli források esetén az adathiány szisztematikusabb problémát jelezhet, mint a statisztikai adatgyűjtésekben. Bizonyos adatokat például egy adott nyilvántartásban egyáltalán nem szabad gyűjteni, vagy csak olyan egységekre engedélyezett, amelyek speciális jogi feltételeknek tesznek eleget. Az is elképzelhető, hogy egy változónak az adatgazda szempontjából nincs nagy jelentősége, ezért az adatok hiánya nem jelent számára gondot, így kevesebb figyelmet is fordít rá.

Az az eset, amikor a másodlagos adatforrásnak csak egy-egy rekordjában hiányzik egy adat, gyakorlatilag ugyanazokkal a standard eljárásokkal kezelhető, amiket a statisztikai felmérésekben a nemválaszolások esetén is használnak, azaz az imputálás különféle módszereivel. Olyan esetben, amikor a szükséges adat egyáltalán nem található az állományban, de legalábbis a legtöbb rekordból hiányzik, modell alapján lehet egyedi szinten becslést végezni a szükséges változó értékeire.

4.3.8. Koordináció az adatgyűjtési célok összehangolására

A másodlagos adatforrások tartalmát és a használt definíciókat érintő problémák megoldási lehetőségeinek felvázolása során többször felmerült annak szükségessége, hogy az adatgazdákkal rendszeres, szoros szakmai kapcsolatot kell fenntartani annak érdekében, hogy minél hatékonyabban lehessen kielégíteni a hivatalos statisztika igényeit.

A statisztikák külső adatforrásokon alapuló előállításában nem csak a statisztikai szervezet és az adatgazdák közötti kétoldalú együttműködések teszik szükségessé, hanem az adminisztratív adatok tekintetében valamennyi érintett hatóság között egy olyan szoros kooperációt is, amit egy kormányzati szinten kinyilvánított, határozott és egyértelmű kötelezettségvállalás erősít meg. 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.

A hatékony koordináció kialakítása hosszú folyamat. Az adatgazdák kezdetben minden bizonnyal idegenkedve fogadják, hogy mások is bele akarnak látni az általuk kezelt adatokba, tartanak attól, hogy kiderülhetnek a nyilvántartások hiányosságai és hibái, ami ronthatja a szervezetről kialakított véleményt. A folyamat elején, aminek során majd a másodlagos adatforrások felhasználása a hivatalos statisztikában rendszerszerűvé válik, az adatgazdák

sokszor nehezen tudják még megkülönböztetni a statisztikai és a nem statisztikai célt, és ezért különösképpen aggodnak, hogy a statisztikai szervezetnél egy-egy személyre vagy vállalkozásra vonatkozóan nagyon sok információ gyűlik össze, és az majd veszélyezteti azok személyiségi jogait, illetve üzleti érdekeit.

A statisztikai szervezetnek aktív szerepet kell vállalni a potenciális partnerek tájékoztatásában, hogy pontos ismereteik legyenek a nem statisztikai adatok statisztikai célú felhasználásának feltételeiről, azokról a korlátokról, amelyek garantálják, hogy a statisztikai hasznosítás ne sértse meg az adatvédelmi szabályokat. Meg kell győzni az adatgazdákat arról is, hogy szakmai, politikai elismertségük sem csorbul azzal, ha kiadják az általuk kezelt adatokat a hivatalos statisztikai indikátorok előállítására. Tudatosítani kell a módszer előnyeit kiemelve, hogy az újabb célokra történő felhasználás pozitívan hat az adatok minőségére, ami a hatóságnak is érdeke.

A közigazgatási szervek együttműködéséről – hangsúlyozva a hazai igényeket és tapasztalatokat – a VII. fejezet szól részletesebben.

4.4. Editálás, validálás

A statisztikai adat-előállítási folyamat két fázisában, az adat-előkészítés, illetve a validálás során külső adatforrások felhasználására is szükség van. Az egyedi adatok ellenőrzése, javítása történhet a vizsgálttól eltérő állomány adatai alapján, sőt az is előfordul, hogy a hiányzó adatokat egy külső adatforrásból pótoljuk. A validálás munkafázisában azt vizsgáljuk, hogy az előállított statisztikai adatok koherens egységben vannak-e az adott szakstatisztikával kapcsolatos információk összességével.

Ha másodlagos adatforrást használunk a statisztikák előállítására, esetenként összetettebb, bonyolultabb munka lehet a megfelelő validálás, hiszen a külső szervezet adatgyűjtését nem, vagy csak alig befolyásolhatjuk, azzal kapcsolatban sokkal kevesebb előzetes információval rendelkezünk, mint egy általunk tervezett, irányított és ellenőrzött felmérés esetében.

4.5. Új irányítási feladatok

Az a tény, hogy sokféle és esetenként nagy tömegű, külső adat kerül be a statisztikai szervezethez, maga után vonja, hogy fokozott figyelmet kell fordítani az adatvédelemre.

A big data típusú adatok hatékony felhasználásához újfajta képzettséggel rendelkező szakemberekre is szükség van. Ezeknek az ún. adattudósoknak a matematikai és statisztikai tudáson túl programozói ismeretekkel és valós tapasztalatokkal is kell rendelkezniük az adott területen, továbbá képesnek kell lenniük mindezek kreatív kombinálására, együttes alkalmazására is. Az adattudomány a hivatalos statisztikában új szakterület, és ezért a statisztikai szervezeteknek fel kell kutatni ehhez értő szakembereket, illetve a régi munkatársakat tovább kell képezni.

4.6. A közvélemény egyetértésének megszerzése

Rendkívül fontos, hogy a szakmai és civil közvélemény megértse és elfogadja, hogy a másodlagos adatforrások statisztikai célú felhasználása milyen előnyökkel jár, illetve a közvélemény

széleskörű jóváhagyása mellett kerüljön sor a nem statisztikai adatoknak a statisztikák előállításához szükséges felhasználására.

Egyes országokban ezt viszonylag könnyű elérni. Másutt előfordulhat, hogy általános nyugtalanságot kelt a gondolat, hogy az adatokat a kormányzati szervek megosztják. Az ilyen aggodalmakat csak kitartó, jól megtervezett felvilágosító politikával lehet enyhíteni. Olyan, egyértelműen és világosan megfogalmazott tájékoztatókat kell megjelentetni az adatok használatára vonatkozó korlátokról és szabályokról, hogy a lakosság és a vállalkozások megértsék, a statisztikai célokra megszerzett érzékeny adatok semmiképpen sem kerülhetnek át más kormányzati szervekhez.

Az ENSZ által kiadott „A hivatalos statisztika alapelvei” című dokumentum 5. alapelve kimondja, hogy statisztikai célokra bármilyen forrásból fel lehet használni adatokat, beleértve a statisztikai felméréseket és az adminisztratív nyilvántartásokat is. A statisztikai hivatalok a minőséget, az időszerűséget, a költségeket és a válaszadók terheit figyelembe véve választhatják ki a megfelelő adatforrást. A 6. alapelv úgy fogalmaz, hogy a statisztikai szervezetek által statisztikai célra gyűjtött egyedi adatok – függetlenül attól, hogy természetes vagy jogi személyre vonatkoznak – szigorúan bizalmasak, és kizárólag statisztikai célokra használhatók föl, azaz az adminisztratív adatok csak egy irányban továbbíthatók. E két alapelv tartalmazza a felvilágosító tevékenység legfontosabb üzeneteit.

A költségvetés és a válaszadó szempontjából végzett költség-haszon elemzések közzététele is hozzájárulhat ahhoz, hogy visszaszorítsuk a közvélemény aggodalmaskodását. Azt is fel lehet hozni érvként, hogy az egyedi adatok nagyobb biztonságban vannak, ha a statisztika másodlagos adatforrásokat használ elsődleges helyett. Nem kell kitöltött kérdőíveket postai úton továbbítani, nincs összeíró az adatgyűjtési folyamatban, aki papíron vagy a saját számítógépén tárolja az egyedi adatokat, kevesebben vesznek részt az adatok feldolgozásában, vagyis az előkészítés és feldolgozás során összességében kevesebb ember férhet hozzá az adatokhoz.

Az adminisztratív adatokon kívüli big data újrahasznosítása a közvélemény szempontjából újfajta problémát vet fel. Miközben mind az egyének, mind pedig a vállalkozások számára jogszabály biztosítja, hogy ellenőrzésük alatt tarthassák, illetve szükség esetén korlátozzák felfedhető adataik körét, a big data típusú adatokat generáló szolgáltatások és eszközök felhasználói gyakran nincsenek tudatában annak, hogy információkat hoznak nyilvánosságra magukról, illetve, hogy az így keletkező adatokat mire lehet használni. Az adatok összekapcsolásával, ha azok között big data típusú adat is szerepel, az adatvédelmi kockázatok tovább növekednek.¹⁵

Figyelemmel a költséghatékonyságra, a lakosság válaszadói terheinek csökkentésére és az adatgyűjtések átfedéseinek elkerülésére Magyarországon is törekedni kell arra, hogy a közvélemény minél szélesebb körben egyetértsen az igazgatási és egyéb nyilvántartások, adatbázisok statisztikai célú felhasználásával. Fontos, hogy maguk az adattulajdonosok is tisztában legyenek azzal, hogy adataik megfelelő védelemben részesülnek, és indokolatlanul érzik magukat fenyegetve vagy kiszolgáltatva. Folyamatosan szem előtt kell tartani, hogy a közvélemény bizalmát nagyon könnyű elveszíteni, de óriási erőfeszítéseket igényel újra megszerezni.

¹⁵ What does “Big Data” mean for official statistics? UNECE, 2013.

<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170614>

A jövőben meg kell ragadni azokat az alkalmakat, amikor a KSH széles körű adatok integrálásával kíván nagy statisztikai adatgyűjtéseket (mint a népszámlálás) kiváltani, hogy különböző fórumokon tájékoztassuk a lakosságot adataik újrahasznosításáról, hangsúlyozva, hogy a statisztikai hivatal mindig is szigorúan garantálja mind a magánéletükkel, mind pedig a vállalkozásaikkal összefüggő adataik védelmét.

4.7. Közkapcsolatok fenntartása

A statisztikai felmérések révén a statisztikai szervezetnek közvetlen kapcsolata van a nagyközönséggel. A másodlagos adatok felhasználásával a statisztika állandó jelenléte a magánszemélyek és a cégek életében visszaszorul, így előfordulhat, hogy a lakosság a statisztikai munkáról egyre kevesebb információhoz jut, ami csökkenti az átláthatóságot és a statisztikával kapcsolatos bizalmat is. Ilyen esetben a legkézenfekvőbb megoldás az, ha javítjuk a szervezet marketingjét, növeljük a publikált adatok és egyéb megjelenések számát. Az ehhez szükséges költségek a másodlagos adatok felhasználásával hosszú távon együtt járó megtakarításokból fedezhetők.

A statisztikai hivatal tevékenységét és eredményeit bemutató rövidtávú eljárások közül talán az a leghatékonyabb, ha az ezekről szóló információk gyakran megjelennek a médiumokban, közép és hosszú távon pedig az, ha az oktatásban minél nagyobb teret kapnak, és eljutnak a statisztikai adatok kiemelt felhasználóihoz is. Felhasználói csoportok létrehozása és a velük folytatott folyamatos párbeszéd hozzájárul a hivatalos statisztika minőségének javításához, valamint a közvéleménynek a statisztikai szervezetet és a statisztikai munkát támogató megítéléséhez.

4.8. A statisztikusok ellenállása

A másodlagos adatforrások hivatalos statisztikákban való hatékony felhasználásának egyik akadálya éppen a statisztikai szervezeten belül jelentkezhetsz. A statisztikusok gyakran tartanak az olyan adatok felhasználásától, amit nem maguk gyűjtöttek, mivel nem bíznak azok minőségében, illetve félnek attól, hogy az adatgyűjtési feladatok csökkenésével munkájuk feleslegessé válik.

A minőségre vonatkozó ítélet sokszor csak szakmai elfogultságon alapul, és nem az ellenőrzött valóságon. Javíthat ezen a helyzeten, ha a statisztika oktatásában, illetve a statisztikusok képzésében nagyobb szerepet kap a másodlagos adatforrásokban rejlő lehetőségek megismertetése. Az oktatás során fontos megismerni azokat a szempontokat és eszközöket, amik lehetővé teszik a statisztikai felmérés és másodlagos adatok közötti tényleges minőségi különbségek megismerését. Gyakran feltételezik például, hogy az igazgatási adatforrások definíciói nem egyeznek a statisztikai fogalmakkal, míg a statisztikai adatgyűjtések adatai annak megfelelőek. Gyakorlatilag azonban nem biztos, hogy ez így van, mert az adatszolgáltató jobbra ugyanazt az adatot adja meg mindkét irányban, nem veszi a fáradságot, hogy a statisztikai kérdőív utasítását végigolvassa, értelmezze, hanem a kért adatot egyszerűen átmásolja a közelmúltban valamely hatóságnak szolgáltatottak közül.

Egy további lehetőség arra, hogy a belső ellenállás gyengüljön, ha bebizonyosodik, hogy az erőforrásigénynek a másodlagos adatforrások használatával járó csökkenése nem vezet

létszámleépítéshez, mert a megtakarított források más fontos feladatokra – például a minőség javítására vagy egyes statisztikák gyakoribb közlésére – használhatók fel.

4.9. A másodlagos adatokat érintő változások kezelése

Az igazgatási nyilvántartásokat általában azért hozzák létre, hogy beszedjék az adókat, vagy ellenőrizzék a kormányzati programok végrehajtását, és ez maga után vonja azt is, hogy ezek létezésük, tartalmuk, fontosságuk és a kezelésükért felelős szervezet szempontjából ki vannak téve a szakpolitikai, illetve stratégiai változásoknak. Mindez kihathat az adminisztratív adatállományok lefedettségére, a definíciókra stb., sőt az is előfordulhat, hogy bizonyos nyilvántartásokat feleslegesnek ítélnék, és ezért megszüntetnek. Az adminisztratív adatok tárolása és feldolgozása során alkalmazott számítástechnikai rendszerek változása is befolyásolhatja a statisztikai célokra felhasználható adatok kínálatát. Még a magánszektorban keletkező adatforrások és a külső szervezet által gyűjtött statisztikai adatok sem védettek az ilyen fajta változásoktól, bár esetükben az adatrendszerben bekövetkező módosulásokat másféle okok generálják.

A másodlagos adatforrások tartalmi változása sokszor hirtelen, előzetes figyelmeztetés nélkül történik. Az efféle kockázat adminisztratív adatok esetében különösen közvetlenül kormányváltást, miniszterek cserélődését, vagy jogszabály-módosítást, magánvállalkozások ügyviteli nyilvántartásai esetében az üzleti feltételek változását, míg a statisztikai adatok esetében szervezeti vagy módszertani változásokat követően magas. Óvintézkedés lehet az olyan, jogszabályok által támogatott eljárások bevezetése, melyek a hirtelen változásoknak a statisztika számára fontos adatokra gyakorolt hatását minimálisra csökkentik. Ilyen lehet az, ha a statisztikai szervezetnek törvény szerint egyetértési joga van minden olyan jogszabályt tekintve, ami létrehoz, módosít vagy megszüntet adminisztratív és statisztikai adatállományokat.

Ha egy statisztikai feladat megoldását egyetlen külső adatforrásra alapozzuk, az mindig hordozni fog bizonyos fokú kockázatot. Ez a veszély megfelelő jogszabály vagy szerződés, illetve együttműködési megállapodás útján csökkenthető. A gyakorlatban azonban az a legjobb módja a problémák elkerülésének, ha rendszeres kapcsolatot tartunk az adatgazdákkal, annak érdekében, hogy megismerjék a statisztikai igényeket, és hogy időben kapjunk tájékoztatást az esetleges változásokról. Még abban az esetben is, amikor egy adatforrás szilárdnak, stabilnak bizonyul, és ezért bízunk benne, hogy az adatai a kívánt tartalommal és minőségben hosszú távon rendelkezésre állnak, érdemes egy vészforgatókönyvet kidolgozni, ami meghatározza, mit lehet, illetve mit kell tenni, ha az adatforrás elérhetetlenné válik. Ez biztosan jobb megoldás, mintha utólag kell reagálni az eseményre.

II. A nyilvántartások és regiszterek jellemzői és a másodlagos források hasznosításával összefüggő szerepük

1. A regiszter fogalma

A másodlagos adatok között kiemelt szerepe van a más szervezetek által vezetett nyilvántartások, regiszterek felhasználásának. A regiszter fogalomnak azonban nincs nemzetközileg általánosan elfogadott definíciója, sokszor bármilyen nyilvántartást vagy másodlagos adatállományokat is regiszternek neveznek, ezért szükséges a fogalom tisztázása.

Az általános szóhasználat szerint a nyilvántartás egy, az adott témában meghatározott céllal rendszeresen vezetett feljegyzés. A regiszter pedig olyan nyilvántartás, amit egységes elvek szerint szervezve vezetnek, és a nyilvántartott adatok visszakereshetők.

Statisztikai célra ennél pontosabb definícióra van szükség.

Az Eurostat fogalomtárában (Eurostat's Concepts and Definitions Database: CODED) hat definíció található rá (lásd felhasznált irodalom), ám minden a témával foglalkozó szakmai anyag saját definíciót fogalmaz meg.

Az SDMX közös metaadat szótára, az SDMX Common Metadata Vocabulary a következő meghatározást alkalmazza:

„A regiszter olyan adattár, amely jegyzékbe vett tételeket rögzít és kezel.”

Magyarázatként hozzáfűzi, hogy „a regiszter olyan írott és teljes nyilvántartás, amely rendszeres bejegyzéseket tartalmaz objektumok meghatározott halmazához tartozó tételekről és részletezésükről”.

A szűken vett definíció (forrása: ISO/IEC 11179–1999)¹⁶ túl általános értelmezési lehetőséget ad, nem tér ki sem a célra, sem a tárgyra, sem a kezelési módra. Bármely adatbázis, adattár, nyilvántartás belefér.

A magyarázó rész pontosítása (forrása: UNECE–2000)¹⁷ szűkíti csak a regiszter fogalmát. Egy regiszter nem általában tartalmaz adatokat valamiről, hanem azzal a céllal, hogy valamilyen objektumtípus egyedeit teljeskörűen és rendszeresen nyilvántartsa.

Ennél a definíciónál közelebb hoz bennünket a gyakorlathoz a svéd Anders és Britt Wallgren¹⁸ regiszter definíciója, amely szerint: „A regiszter célja adott objektumcsoportba vagy sokaságba tartozó objektumok teljes listájának létrehozása. A minőség hiányossága esetén hiányozhatnak bizonyos adatok egyes objektumokra vonatkozóan, de az objektum azonosítójának mindig rendelkezésre kell állni, mivel ez biztosítja a regiszter aktualizálhatóságát és új változóra vonatkozó értékekkel való bővíthetőségét. A teljesség és az ismert azonosítóval rendelkezés ezért fontos jellemzője a regiszternek”.

¹⁶ ISO/IEC International Standard 11179, Part 1, Framework for the specification and standardization of data elements, 1999.

¹⁷ Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000.
<http://www.unece.org/stats/publications/53metadaterminology.pdf>

¹⁸ WALLGREN, Anders and Britt (Statistics Sweden), "Register-based Statistics. Administrative Data for Statistical Purposes", John WILEY & Sons, Ltd, England, 2007.

A regiszter célja tehát valamely sokaság egyedeinek teljes körű nyilvántartása. A sokaság lehet valamely statisztikai megfigyelés sokasága, de lehet bármely jogszabállyal előírt tevékenység objektumainak teljes körű adminisztratív nyilvántartása is.

A regiszterrel leírt sokaság egységét, objektumtípusát a továbbiakban nyilvántartási egységnek nevezzük. A definíció szerint az objektumtípus, azaz a nyilvántartási egység előfordulásaihoz olyan azonosítót kell rendelni, amely *egyedileg azonosítja* az objektumokat. Fontos, hogy az azonosítót a nyilvántartás ne csak a regiszter működtetésére, belső azonosítóként használja, hanem *állandó*, mások által is *ismert* és alkalmazott legyen, mert ez biztosítja a más forrásokból származó adatokkal való aktualizálhatóságot, összekapcsolhatóságot.

A regiszternek alapvető tulajdonsága a karbantarthatóság. Valamely objektumok statikus listáját nem tekintjük regiszternek. A regiszternek követnie kell a nyilvántartott objektumokban történt változásokat a regiszter céljának megfelelő aktualitással.

A változások ellenőrizhetősége, korábbi állapotok előkereshetősége azt igényli, hogy a regiszter ne csak az objektumok aktuális állapotát írja le, hanem a változások oka, az objektumok korábbi állapota, története is visszakereshető legyen.

A mai kor követelményeinek megfelelően a nyilvántartások számítógéppel kezelhető, adatbázisban tárolt és kezelt nyilvántartások, ahol a nyilvántartás tárolását, karbantartását, elérhetőségét alkalmazási rendszer támogatja.

Ezek alapján a regiszter általunk használt definíciója:

„A regiszter valamely sokaság egyedeinek azonosítására, elérhetőségére, a regiszterre épülő eljárásokat támogató jellemzőkre vonatkozó rendszeresen karbantartott nyilvántartás. A regiszterből elérhetők a sokaság aktuális és történeti állapotai, a változások oka, hatálya, forrása. A regiszter a sokaság egyedeire (egységeire) vonatkozó adatokat strukturáltan, adatbázisban tárolja, a regiszter karbantartását, elérhetőségét erre a célra létrehozott informatikai alkalmazásrendszer támogatja.”

Ha valamely sokaság egyedeinek kezelése nem felel meg a fenti definíciónak, például a történeti adatokat, változásokat nem visszakereshető módon tárolja, akkor azt nem regiszternek, hanem *nyilvántartásnak* nevezzük.

Előfordul, hogy a regiszterek a nyilvántartási egységre vonatkozó egyéb adatokat is tárolnak az egységet jellemző ismérvként. Mégis élesen el kell választani a nyilvántartásoktól, regiszterektől az ezekhez kapcsolódó adatállományokat. A regiszter elsődleges célja a nyilvántartási egységek tárolása kezelése, naprakész karbantartása, az adatállományoké pedig a nyilvántartási egységekre vonatkozó, meghatározott felhasználási célú és témájú adatok gyűjtése, kezelése, feldolgozása. A regiszterek az adott információs rendszer törzsadatairaiként szolgálnak, karbantartásuk az egységekben bekövetkezett egyedi változásokhoz kapcsolódik. Az adatállományok pedig időközönkénti megfigyeléseket tartalmaznak a regiszterben szereplő összes élő egységre, vagy azok részhalmazának meghatározott tulajdonságaira vonatkozóan. Az adatállományokban minden megfigyelést azonosít a nyilvántartási egység, objektum azonosítója és az idő, amire a megfigyelés vonatkozik.

2. A regiszter nyilvántartási egysége és sokasága

A regisztersokaság leírása pontos szabályokkal körülhatárolt leírás kell, hogy legyen egyrészt a nyilvántartás tárgyát, egységét, másrészt a leírandó nyilvántartási egységek körét, sokaságát illetően.

A regiszter leírhat egy vagy több egymással kapcsolatban álló (alá-fölérendeltségi viszonyban levő) objektumot, nyilvántartási egységet (pl. gazdasági szervezet és telepei, tevékenységei).

Egy nyilvántartási egység definiálásához, azonosítójának képzéséhez ismerni kell, hogy milyen változások alakítják a nyilvántartási egységet. Mely változások idézik elő egy új egység keletkezését, módosulását, megszűnését. Mikor kell egy változásnál új egységet (új azonosítót) létrehozni a régi megszüntetésével, és mikor csak módosításként kezelni a régi azonosító megtartásával. Ez nem mindig egyértelmű feladat, például egy gazdasági szervezet átalakulásakor az egységek képzési szabályát a regiszterdokumentációnak kell tartalmazni.

A sokaság az adott nyilvántartási egység, objektumtípus előfordulásainak összessége. A sokaságot leíró szabálynak tartalmazni kell, hogy

- mely objektumok esnek bele kötelezően a regiszter sokaságába,
- mely objektumokat zárjuk ki a sokaságból,
- esetleges önkéntesség alapján mely objektumokra terjedhet ki.

Például a polgárok személyi adatainak és lakcímének nyilvántartásáról szóló 1992. évi LXVI. törvény 4. §-a írja le a törvény alá tartozó személyek körét, ezzel a nyilvántartás sokaságát. A nyilvántartás módszertani leírása pedig ki is emeli a sokaságba tartozó és a sokaságba nem tartozó egyedeket.¹⁹

„A személyiadat- és lakcímnyilvántartás tartalmazza a

- 1., magyar állampolgárok,
- 2., a bevándorlási, letelepedési engedéllyel itt élő, továbbá a menekültként elismert, vagy ennek hiányában az oltalmazott külföldi állampolgárok,
- 3., 2004. május 1-je óta az Európai Gazdasági Térséghez (EGT) tartozó országok állampolgárai közül az EGT tartózkodási engedéllyel Magyarországon élő külföldi állampolgárok,
- 4., 2007. július 1-je óta a szabad mozgás és tartózkodás jogával rendelkező külföldi állampolgárok, ha legalább 3 hónapot Magyarországon töltenek, és regisztrációs igazolást vagy tartózkodási kártyát kaptak,
- 5., a külföldön élő magyar állampolgárok adatait. (Ez utóbbiakat abban az esetben, ha kérik adataik nyilvántartását.)

Fentiekén kívül a nyilvántartásban (történeti állományban) szerepelnek az 1975-ben vagy később elhalt, vagy véglegesen külföldre távozott személyek adatai. *Nem terjed ki a személyiadat- és lakcímnyilvántartás a nem EGT-országokból érkezett, átmenetileg – beutazási vagy tartózkodási engedéllyel – Magyarországon élő külföldiekre még akkor sem, ha itt kívánnak letelepedni, de letelepedési engedélyük még nincs. Nem tartalmazza a*

¹⁹ Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala (KEK KH), Erdész Tiborné: A személyiadat- és lakcímnyilvántartás, okmánykiadás összesített adatai, Népszámság-nyilvántartási füzetek (szerk.: Vetési Iván), 2010.
http://www.nyilvantarto.hu/letoltes/kiadvanyok_2010_jan_szoveg.doc

nyilvántartás a diplomatak, a külföldi diákok, a hivatalos vagy üzleti ügyben beutazók, a turisták és más Magyarországra látogatók adatait sem.”

Amennyiben a regiszter több nyilvántartási egység kapcsolatát is leírja, az egyes nyilvántartási egységek sokaságai eltérhetnek egymástól, például csak bizonyos tulajdonsággal rendelkező egységekre vonatkozóan tartjuk nyilván a másik nyilvántartási egységet. Ilyen esetekben nyilvántartási egységtípusonként kell megadni a sokaságok közötti eltérést.

A regiszter sokaságának lehetnek olyan részsokaságai, amelyek akár a karbantartás módja, akár az objektum eltérő jellemzői miatt más kezelést igényelnek. Mind az objektum leírásának, mind a részsokaságok leírásának ki kell térni ezekre a sajátosságokra. Például a szociális intézményeket leíró regiszterben az intézmények egy része adószámmal rendelkezik, más része csak részben önálló egység, ezért adószámmal nem rendelkezik. A két részhalmaz kezelése más szabályokat, másféle azonosítást igényel.

3. A nyilvántartási egység ismérvei

A regiszterek a nyilvántartási egységet különböző ismérvekkel jellemzik. A nyilvántartási egységre vonatkozó ismérveket a következő csoportokba sorolhatjuk.

3.1. Azonosító ismérvek

Egy nyilvántartási egységnek lehet mesterséges és természetes azonosítója. A mesterséges azonosító általában valamely képzési szabály alapján létrehozott nem beszédes azonosító. Előfordul, hogy az egyedi azonosítás csak több azonosító együttes alkalmazásával lehetséges, amikor a nyilvántartási egység függ egy másik nyilvántartási egységtől, vagy ha a nyilvántartás a nyilvántartási egység különböző állapotait, történetét ugyanabban az állományban kezeli.

A természetes azonosítók közé soroljuk a nyilvántartási egység nevét, megnevezését és más, az egyediség megállapításához szükséges ismérveket (személyeknél például a születési idő, hely, anyja neve). Ezeket a természetes azonosítókat a regiszterek vonatkozásában nem tekintjük azonosító ismérveknek, hanem az ismérvek más csoportjaihoz (elérés, csoportosító stb.) soroljuk.

3.2. A nyilvántartási egység elérését segítő ismérvek

A nyilvántartási egységgel vagy annak tulajdonosával való kapcsolattartást elősegítő ismérvek az elérést segítő ismérvek. Gazdasági szervezeteknél például a szervezet neve, székhely és levelezési címe, a képviselő neve, telefon, fax, e-mail elérhetősége. Személyeknél is a személy neve, lakóhelyének és tartózkodási helyének a címe.

3.3. A nyilvántartási egység demográfiai ismérvei

Ide soroljuk azokat az ismérveket, amelyek a nyilvántartási egység állapotváltozásait írják le, keletkezés, megszűnés idejét, az egység állapotát és az állapot hatálya. Például: egy szervezet alapításának éve, működési állapota vagy egy személy születési, halálozási időpontja.

3.4. A nyilvántartási egység leíró ismérvei

A leíró ismérvek a nyilvántartási egység tulajdonságait jellemző ismérvek, amelyek alapján a nyilvántartási egységek csoportokba rendezhetők. Ilyen például egy lakás rendeltetése, egy szervezet főtevékenysége, területi elhelyezkedése stb. A leíró, csoportosító ismérvek általában kódolt ismérvek, amelyeknél a lehetséges tulajdonságértékeket, értékészletet osztályozás, nómenklatúra írja le.

3.5. A nyilvántartási egységek kapcsolatát leíró ismérvek

Egy adott nyilvántartási egység kapcsolatban állhat

- az adott nyilvántartási egység más előfordulásaival (pl. szervezetek tulajdonosi, ellenőrzési kapcsolata, szülő-gyerek, valamint férj-feleség kapcsolata),
- adott regiszteren belül más nyilvántartási egységgel, (pl. a szervezet és telepei közötti kapcsolat),
- más regiszterbeli nyilvántartási egységgel (pl. személyek kapcsolata a személy- és lakcímnnyilvántartással, valamint a társadalombiztosítási regiszterrel).

A kapcsolat létrehozását segíti, ha a kapcsolódó nyilvántartási egység azonosítója is jellemzi az adott nyilvántartási egységet.

3.6. A regiszterfelhasználáshoz kapcsolódó folyamatot támogató ismérvek

A regiszterek vezetése nem öncélú, hanem különböző folyamatok épülnek rájuk. A folyamatok működése a fentiekén kívül saját ismérveket igényel. Ilyen folyamat lehet a személy- és lakcímnnyilvántartás esetében például a választási körzetek meghatározása, vagy a statisztikai regisztereknél a mintaválasztás, adatgyűjtés-szervezés. A folyamatokat vezérlő ismérvek lehetnek általános leíró ismérvek, de lehetnek nem általános érdeklődésre számot tartó vagy technikai ismérvek is. Az utóbbiakat soroljuk ebbe a csoportba.

3.7. A nyilvántartás vezetését támogató egyéb ismérvek

A regiszterek aktualizálása általában több forrásból történik. A nyilvántartási egység változása érintheti annak egészét vagy csak valamely tulajdonságát, ismérvét. Lehetőség szerint valamennyi, de legalább a fontos ismérvekre vonatkozóan vezetni kell a változás forrását, a változás hatályát, rendszerbe kerülésének dátumát, a módosító felhasználó azonosítóját. A változásokat leíró ismérveket soroljuk a nyilvántartás vezetését támogató ismérvek közé.

4. A regiszter változáskezelése, karbantartása

A regiszterek tartalma a forrásokból érkező információk alapján állandóan változhat. Új nyilvántartási egységek keletkeznek, régiek megszűnnek, átalakulnak, más nyilvántartási egységekkel, más regiszterekkel való kapcsolataik módosulnak, a tulajdonságaik megváltoznak. A regiszterek tervezésénél az egyik fő feladat a karbantartás lehetséges forrásainak és a nyilvántartási egység változástípusainak felmérése. A változástípusok, események jegyzékbe, nómenklatúrába foglalhatók.

A változástípusok, események lehetnek egyszerűek, mint például a nyilvántartási egység egy tulajdonságának megváltozása, de lehetnek összetettek is, amikor módosítások sorozatát kell végrehajtani egy esemény regiszteren való végigvezetéséhez. Ez általában akkor igaz, amikor egy regiszter több egymással kapcsolatban álló nyilvántartási egységet kezel. Ilyen összetett esemény például a címregiszter esetén a települések közötti területátadás, amikor a módosítás településrészeket, közterületeket érint, esetleg házszámok újra sorszámozásával is jár. A regiszter-karbantartó rendszer szofisztikáltságának függvénye, hogy egy változástípushoz tartozó módosítássorozatot együtt kezel le, vagy a részekre bontást és azok logikai sorba rendezését a felhasználóra bízta. A változástípushoz, eseményekhez kapcsolódó feladatokat a dokumentációnak kell rögzítenie.

A változások az esemény felmerüléséhez képest kisebb-nagyobb késéssel kerülnek a regiszterbe, vagy azért, mert később jelentik, vagy, mert az esemény ellenőrzése, javítása, elfogadása időt vesz igénybe. Amennyiben az esemény elsődleges regisztrálási, felvételi helyéről nem online módon módosul az adott regiszter, hanem a regiszternek való adatátadás kötegelten, később történik, akkor a változás regiszterben való megjelenése függ az adatátadások-átvételek gyakoriságától is.

Alapelv a regisztereknél, hogy az összes releváns adat akkor és csak akkor legyen gyűjtve, amikor az objektumhoz kapcsolódó esemény felmerül. Gazdaságossági okból is így célszerű, és ez megelőzi az adatok redundanciáját is. Csak azokat az egységeket és attribútumokat kell módosítani, amelyekben változás történt.

A regiszter a nyilvántartási egységek állapotváltozásainak tárolására, kezelésére többféle módszert alkalmazhat. Ez függ a regiszterrel, nyilvántartással kapcsolatos elvárásoktól, illetve attól függően, hogy mennyire tekinthető a regiszter céljának a régi információk előkereshetősége, a különböző időponti állapotok meghatározása és a karbantartás nyomon követhetősége.

A nyilvántartások, regiszterek az alkalmazott változáskezelés szempontjai szerint is eltérhetnek.

- A legegyszerűbb nyilvántartás az, amelyben csak a nyilvántartási egység utolsó állapotát tárolják, annak történetét nem. Ilyen típusú nyilvántartás-kezelés csak olyan címlistáknál, egyéb objektumlistáknál célszerű, ahol az egységek korábbi állapotaira nincs szükség, mindig csak az aktuálisra. Itt is célszerű az utolsó karbantartás forrását, időpontját megőrizni. Az ilyen nyilvántartást nem tekintjük regiszternek.
- A nyilvántartásnak lehet több, a karbantartás, illetve a felhasználás gyakoriságának megfelelő időpontokra vonatkozó állapota. A történeti adatok a korábbi időszakok adataiból kereshetők vissza. Itt sem őrzik meg a változások sorozatát, amelyek egy adott nyilvántartási egységnél az új időszak új állapotához vezetnek.
- Egyszerű regiszter esetében, a változások hatályát a nyilvántartási egységben jelzik. Minden egyes változási esemény a nyilvántartási egység egy új állapotát hozza létre. A hatály kezdete az egység azonosítójának a részévé válik. A regiszterből bármely időponti állapot előkereshető a változás hatályának segítségével. Erre a regisztertípusra példa a KSH településregisztere.
- Azoknál a regisztereknél, ahol több forrásból történik az aktualizálás, különböző változási események vannak és a nyomon követhetőség elvárás, ott a fenti típusok nem elegendők. A

regiszternek rögzíteni kell minden egyes változási eseményt a változás forrásával, hatályával, a rendszerbe kerülés dátumával, a módosító esemény azonosítójával, esetleg a módosítás alapjául szolgáló dokumentumok tárolásával. A módosítás előtti és utáni állapot is része a regiszternek, azaz a történeti adatok előkereshetők. A nyilvántartási egységek ismérvei között különbség lehet a történetiség kezelésében; vannak fontos ismérvek, amelyeknek a történetét, állapotváltozásait minden esetben meg kell őrizni, és lehetnek kevésbé fontos ismérvek, amelyeknél mindez nem elvárás, elegendő az utolsó állapot megőrzése. Ezen az elven működik a KSH-ban például a gazdasági szervezetek regisztere (GSZR) és a statisztikai címregiszter is.

A statisztikai folyamatok (adatgyűjtés, feldolgozás, elemzés stb.) általában nem közvetlenül a regiszterekre épülnek, éppen azért, mert a regiszterek változhatnak. A regiszterekből a statisztikai felvételek vonatkozási idejének megfelelően ún. befagyasztott állapotok készülnek (pl. december 31-i vagy havi állapotok), amelyek a statisztikai felvételek felvételi keretének alapját képezik. A befagyasztott állapotokon, és a ráépülő felvételi kereteken legfeljebb a regiszterrel párhuzamos hibajavítás történhet, de változáskezelés nem.

5. Adminisztratív és statisztikai regiszterek

A regisztereket megkülönböztetjük aszerint, hogy adminisztratív vagy statisztikai információs rendszer részét képezik. Wallgren²⁰ szerint „az adminisztratív információs rendszerekben adminisztratív célra használt regisztereket *adminisztratív regiszternek* nevezzük. Az adminisztratív regiszter tartalmazza az összes kezelt objektumot, és az adminisztratív célra használt változókat.” Az északi országok statisztikai hivatalai által a regiszter alapú statisztikáról készített jelentés²¹ szerint az előző információk kiegészülnek a következőkkel: „Ezeket a regisztereket használják a köz- és magánintézmények és vállalkozások áruk és szolgáltatások előállításában, illetve ennek a tevékenységnek az eredménye a regiszterinformáció. A statisztikai célra használt adminisztratív regisztereket állami vagy kapcsolódó helyi hatóságok működtetik, bár a magánszervezetek által működtetett regiszterek szintén használatosak.”

A magánszervezetek által működtetett regiszter példájául szolgálhatnak a magánbiztosítók által működtetett regiszterek. Az előzőket még kiegészíthetjük azzal, hogy az adminisztratív regiszterek vezetése általában valamely jogszabályra vagy azok csoportjára épül. Az adminisztratív regiszterekben a változásokat általában okirattal támasztják alá, meglevő információ módosítása csak újabb okirattal alátámasztva kerülhet be a regiszterbe. Így az alábbi definíció használatát javasoljuk:

„Adminisztratív regiszter egy adminisztratív információs rendszerben alkalmazott, jogszabály(ok)ra épülő regiszter, amely nyilvántartja az információs rendszerben kezelt, adott típusú objektumokat és azok változóit (tulajdonságait). Az adminisztratív regisztert általában állami vagy helyi hatóságok működtetik. Adminisztratív regisztert működtethet vállalkozás vagy intézmény is.”

²⁰ Wallgren, Anders and Britt (Statistics Sweden): Register-based Statistics. Administrative Data for Statistical Purposes. Wiley Series in Survey Methodology, John WILEY & Sons, Ltd, England, 2007.

²¹ Economic Commission for Europe of the United Nations (UNECE), „Register-based statistics in the Nordic countries – A review of best practices with focus on population and social statistics”, United Nations, New York and Geneva, 2007.
<http://essnet.admindata.eu/WikiEntity?objectId=4901>

Statisztikai regiszterek Wallgren szerint „a statisztikai célra működtetett regiszterek. A statisztikai regiszterek képzése adminisztratív regiszterek felhasználására épül, oly módon, hogy az objektumok halmaza, az objektumok és a változók megfeleljenek a statisztikai elvárásoknak.”

A statisztikai regiszterek célja valamely vizsgált sokaság egyedeinek teljes körű nyilvántartása, azonosítása és elérhetőségének, fontosabb jellemzőinek regisztrálása, karbantartása. A nyilvántartási egységek a statisztikai felvételek és elemzések statisztikai egységei. A statisztikai regiszterek adott időponti állapotai felvételi keretek alapjául szolgálnak. Reprezentatív felvételek esetén a regiszterek időponti állapotai adják a mintavételi keretet, a mintavétel és becslés rétegeképző ismérveit. A statisztikai feldolgozás és elemzés során a regiszterekre épül az információk csoportosítása, aggregálása, új statisztikai egységek képzése.

A regiszter általános definíciójára építve *statisztikai regiszternek* valamely sokaság egyedeinek azonosítására, elérhetőségére és az adatfelvételi folyamatot támogató jellemzőkre vonatkozó rendszeresen karbantartott nyilvántartást tekintjük. A regiszterből elérhetők a sokaság aktuális és történeti állapotai, a változások oka, hatálya, forrása. A regiszter a sokaság egyedeire (egységeire) vonatkozó adatokat strukturáltan, adatbázisban tárolja.

A statisztikai regiszter forrása általában egy vagy több adminisztratív regiszter. Egyetlen adminisztratív regiszterből nem mindig képezhető közvetlenül a statisztikai regiszter, hanem csak több forrás felhasználásával, az információk felülvizsgálatával vagy statisztikai felvételek segítségével való pontosítással. Ilyen okok lehetnek a következők:

- Egy-egy adminisztratív regiszter gyakran nem fedi le a statisztika által *elvárt sokaságot*, ezért több adminisztratív regiszterre kell építeni a statisztikai regisztert. Ez a helyzet például a gazdasági szervezetek regiszterénél, ahol a gazdasági tevékenységgel foglalkozó szervezeteket, az adószámmal rendelkező szervezetekkel közelíti a statisztikai regiszter. Az adószámmal rendelkező szervezetek egyes részhalmazai (cégbejegyzésre kötelezettek, egyéni vállalkozók és egyéni cégek, költségvetési szervek) különböző adminisztratív regiszterekből származnak.
- A statisztikai regiszter nyilvántartási egysége eltérhet az adminisztratív regiszterekben kezelt nyilvántartási egységtől, és ezekben az esetekben az adminisztratív regiszterre épülve új nyilvántartási egységet kell képezni. Ennek példája a KSH kereskedelmi szálláshely regisztere, ahol az önkormányzatok *szálláshelyengedély*-nyilvántartására építve készül el a működési engedéllyel rendelkező kereskedelmi *szálláshelyek* regisztere.
- Előfordul, hogy a *nyilvántartási egység* regisztrálandó *ismérveinek* egy része is újabb regiszterek bevonásával tölthető csak fel (pl. a GSZR-nél a csoportképző kódok változása (TEÁOR, GFO) az adóhatóságtól). Ha egy ismérv egyik adminisztratív regiszterben sem áll rendelkezésre, akkor statisztikai felvételek felhasználásával pótolható.
- A nyilvántartási egységek megfigyelt ismérveinél fontos, hogy az *ismérveket leíró fogalmak*, nomenklatúrák, *osztályozások* összhangban legyenek az adminisztratív és a statisztikai regiszterben. Ha az elsődleges felvétel helyén, az adminisztratív regiszterben más értelmezés, más értékészlet szerint, az ismérvek közötti összefüggéseket figyelmen kívül hagyva veszik fel az egységre vonatkozó információkat, akkor az utólag már csak többletforrások felhasználásával, plusz munkaráfordítással javíthatók.
- Különbség lehet a nyilvántartás jellemzőinek kezelésében és felülvizsgálatában. Az adminisztratív regisztereknél, amint láttuk, az adatok módosítása hiba esetén is csak akkor

lehetséges, ha az okirattal alátámasztott. A statisztikai regiszterekben ezzel szemben a statisztikai felvételek jobb minősége érdekében a felderített hibákat akkor is javítani kell, ha az adminisztratív regiszterben még nem történt meg a javítás. Ennek érdekében a statisztikai regiszterek ismérveinél megkülönböztetjük az *adminisztratív ismérveket* és a *statisztikai ismérveket*. A legfontosabb adminisztratív forrásból származó ismérveket változatlan formában megőrzi a statisztikai regiszter, de mellette bevezeti az adott ismérv statisztikai változatát, amely az adminisztratív ismérv változásától függetlenül is módosítható. A GSZR-ben például adminisztratív ismérv a szervezet neve, székhely címe, adminisztratív főtevékenysége, amely nem módosulhat, csak adminisztratív forrásból. Emellett léteznek statisztikai ismérvek, mint például statisztikai főtevékenység, amely az adminisztratív főtevékenységből indul ki, de a szervezet ismert hozzáadott érték adatai alapján módosulhat. A szervezet székhelyén kívüli egyéb levelezési, kapcsolattartási címei szintén módosíthatók az elérhetőség érdekében.

Amíg az adminisztratív regisztereknél általános alapelv, hogy minden változás akkor kerüljön be a regiszterbe, amikor az esemény történik (illetve minél közelebb az esemény felmerülésének idejéhez), a statisztikai regisztereknél ez nem mindig szempont. A regiszterek aktualizálása az adminisztratív forrásokból periodikus adatátvételekkel történik. Az átvétel gyakorisága függ attól, hogy milyen gyakoriságú adatfelvételek épülnek az adott statisztikai regiszterre. Ahol csak évenkénti felvételek épülnek rá, ott a karbantartása is lehet, hogy csak évente valósul meg, ahol ennél gyakoribb, vagy a regiszter nemcsak az adatfelvételeket szolgálja, ott lehet ennél gyakoribb is. Például a GSZR-nél, amely számos felvétel alapjául szolgál és a KSH-n kívüli felhasználása is elterjedt, a legfontosabb adatforrásokból való adatátvétel napi aktualizálást tesz lehetővé.

A statisztikai regiszterek minőségének ellenőrzése érdekében az adminisztratív forrásokon kívül egyéb források is felhasználhatók. Például regiszter-karbantartó kérdőívek segítségével az új egységekre vagy bizonyos részsokaságokra vonatkozóan új információk szerezhetők be, vagy a régi információk ellenőrizhetők.

A statisztikai regiszterekben felismert hibák visszacsatolásának az adminisztratív regiszter felé gyakran nincs elfogadott rendje (ellenőrzés, javítás). Az adminisztratív regiszterek az okiratra épített karbantartásra hivatkozással nem hasznosítják a statisztikai regiszterekben talált hibákat. Ebből következően az adminisztratív regiszterből való következő átvételek is ezekkel a hibákkal terheltek, ha más forrásból nem javulnak. A statisztikai regiszter karbantartójának erre a hatályok és módosítási dátumok megfelelő kezelésével kell felkészülnie. Az ismérvek összefüggésrendszerének hibáinál javíthatna a helyzeten, ha az adminisztratív regiszterek is alkalmaznák a statisztikai regiszterek ellenőrzési szabályait. Az erre való készséget a regiszterek közötti együttműködési megállapodások erősíthetik.

6. Alap- és specializált regiszterek

A statisztikában használatos adminisztratív regisztereknek két fő csoportját különböztetjük meg, az alapregisztereket és a specializált regisztereket.

Alapregiszterek azok a regiszterek, amelyek a gazdaság és a társadalom legfontosabb objektumait írják le. Három vagy négy alapregisztert különböztet meg a nemzetközi irodalom. A három alapregiszter a személyek, címek és ingatlanok, valamint a szervezetek, intézmények

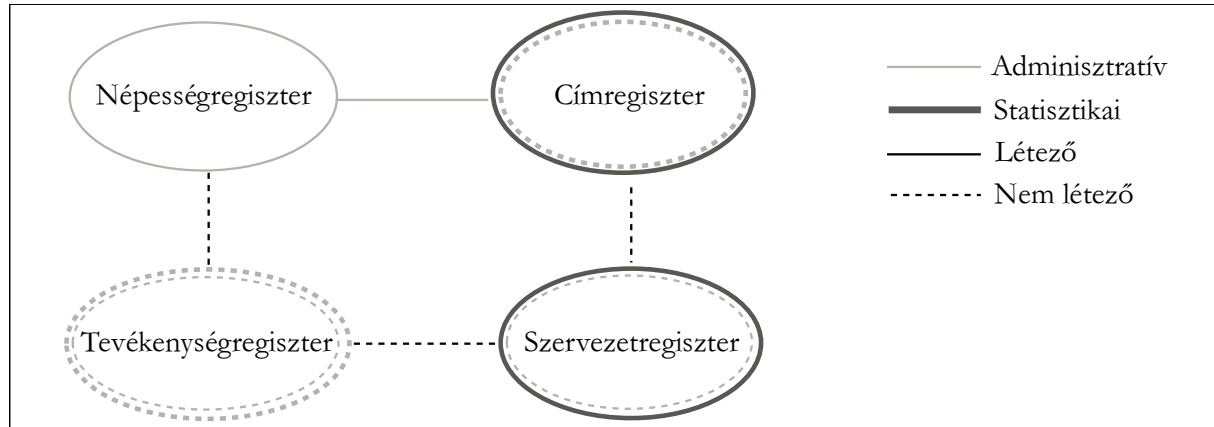
regisztere. Negyedikként társulhat még ehhez egyes országok szakembereinek véleménye szerint a tevékenységek regisztere. Az alapregiszterek feladata ezen objektumok teljes körének nyilvántartása, egységes azonosítása, más regiszterrel való kapcsolatuk leírása. Az elvárt alapfunkciók a következők:

- A személyek nyilvántartása az adott ország lakosságára vonatkozó alapinformációkat tartalmazza (név, születési idő, hely, anyja neve, állampolgársága stb.) a személyeket egyedi azonosítóval ellátva. A személynyilvántartás kapcsolatot létesít a címregiszterrel a lakóhely(ek) azonosításán keresztül.
- A címregiszter a címeket numerikusan azonosítja, közigazgatási területi egységhez (azok hierarchiájához) és földrajzi helyhez (geokódhoz) rendeli. A címregiszter sokaságába bármely rendeltetésű cím beletartozik, nemcsak a lakóhely rendeltetésű címek.
- A szervezetregiszter célja a gazdasági, társadalmi tevékenységet folytató szervezetek, vállalkozások, intézmények és telephelyeik lehető legszélesebb körű nyilvántartása és egyedi azonosítása.
- A tevékenységi regisztert csak a svédek tekintik statisztikai alapregiszternek, a munkával (foglalkozással) összefüggő tevékenységek regiszterét értik alatta, de ott sincs erre egy közös adminisztratív regiszter.

Az ideális állapotot és a Magyarországon jelenleg rendelkezésre álló regisztereket az alábbi ábra mutatja.

2. ábra

Alapregiszterek



Egységes adminisztratív regiszter a fenti három közül Magyarországon csak a népeségnyilvántartásra van (a személyadat- és lakcímnilyvántartás). Elfogadott címregiszter nincs. A szervezetekre, intézményekre, azok telepeire sem áll rendelkezésre egységes adminisztratív regiszter, csak szervezettípusonkénti részsokaságokra. Tevékenységregiszternek pedig még kezdeménye sincs.

A statisztikai alapregiszterek az adminisztratív regiszterekre épülnek. A népeségregiszternek nincs statisztikai megfelelője, a KSH igény esetén felhasználója az adminisztratív regiszternek. Az adminisztratív címregiszter mint alapnyilvántartás kialakítása folyamatban van, de még nem elérhető a statisztika számára. A statisztikai címregiszter pillanatnyilag a népszámlálásra és az építési engedélyekre épül. Nincs rendszeres, elfogadott kapcsolata sem a személyadat- és lakcímnilyvántartás címregiszterével, sem az ingatlan-nyilvántartással. A

szervezetek, intézmények statisztikai regiszterének szerepét több forrásból összeállítva a KSH gazdasági szervezetek regisztere tölti be.

Specializált regiszterek adott felhasználási célra, adott csoport részére létrehozott regiszterek. Ilyenek lehetnek:

- a személyekhez kapcsolódó nyilvántartások, a társadalombiztosítási, személyi adóalany, oktatási, foglalkoztatási, munkanélküli, bűnügyi, gépjármű-nyilvántartás stb.,
- a címekhez tartozó épület- és lakásnyilvántartás, az ingatlanok tulajdonosi nyilvántartása,
- a szervezetekhez tartozóan a szervezeti adóalany, oktatási intézmények, kereskedelmi egységek, szociális, nonprofit szervezetek regisztere, szervezetek által üzemeltetett gépjárművek stb.

Ezek a regiszterek általában az alapregiszterek sokaságának csak egy részét fedik le, az alapregiszter nyilvántartási egységeit új tulajdonságokkal bővítik, illetve új nyilvántartási egységeket hoznak létre a saját céljaiknak megfelelően. A specializált regiszterek is lehetnek adminisztratív vagy statisztikai regiszterek.

A statisztikai regisztereknél gyakori, hogy az adminisztratívól eltérő új nyilvántartási egység képzésére van szükség. Az alapregiszter feladata, hogy azokat az ismerveket kezelje, amelyek az ismert új nyilvántartási egység képzéséhez szükségesek. A svéd népesség-nyilvántartás például a szülők és a házastárs nevét is tartalmazza, hogy a családfogalomnak megfelelő egység, ami adminisztratív nyilvántartásban nem szerepel, egyszerűbben legyen képezhető. Hasonló probléma a háztartás nyilvántartási egység képzése, amire sehol nincs adminisztratív nyilvántartás, csak a személyek és lakások megfelelő algoritmus szerinti egymáshoz rendelésével képezhető.

Az alap- és specializált regiszterek szoros kapcsolatban állnak egymással. A specializált regiszterek az alapregiszterekre épülnek, közös azonosítók teszik lehetővé a kapcsolódó nyilvántartási egységek ismerveinek feldolgozását. A specializált regiszterekből azokat, amelyek egy adott alapregiszterhez kapcsolódnak úgy, hogy egy adott részsokaságra vonatkozóan új ismervekkel egészítik ki az alapregisztert, *szatellitregisztereknek* nevezzük. Ilyen szatellitregisztere a szervezetregiszternek például a működési engedéllyel rendelkező szociális szolgáltatást nyújtók regisztere.

7. A KSH regiszterei, nyilvántartásai

A KSH számos nyilvántartást kezel. Ezeknek csak egy része elégíti ki a regiszterfogalommal kapcsolatos elvárásokat. A nyilvántartások, regiszterek nem függetlenek egymástól. Az utóbbi években nagyon sok lépés történt az összehangolásukra, közös rendszerben kezelésükre. Két fő alapregiszterről beszélhetünk, amelyek köré a többiek csoportosulnak, illetve azokkal kapcsolatban állnak vagy kiegészítik őket:

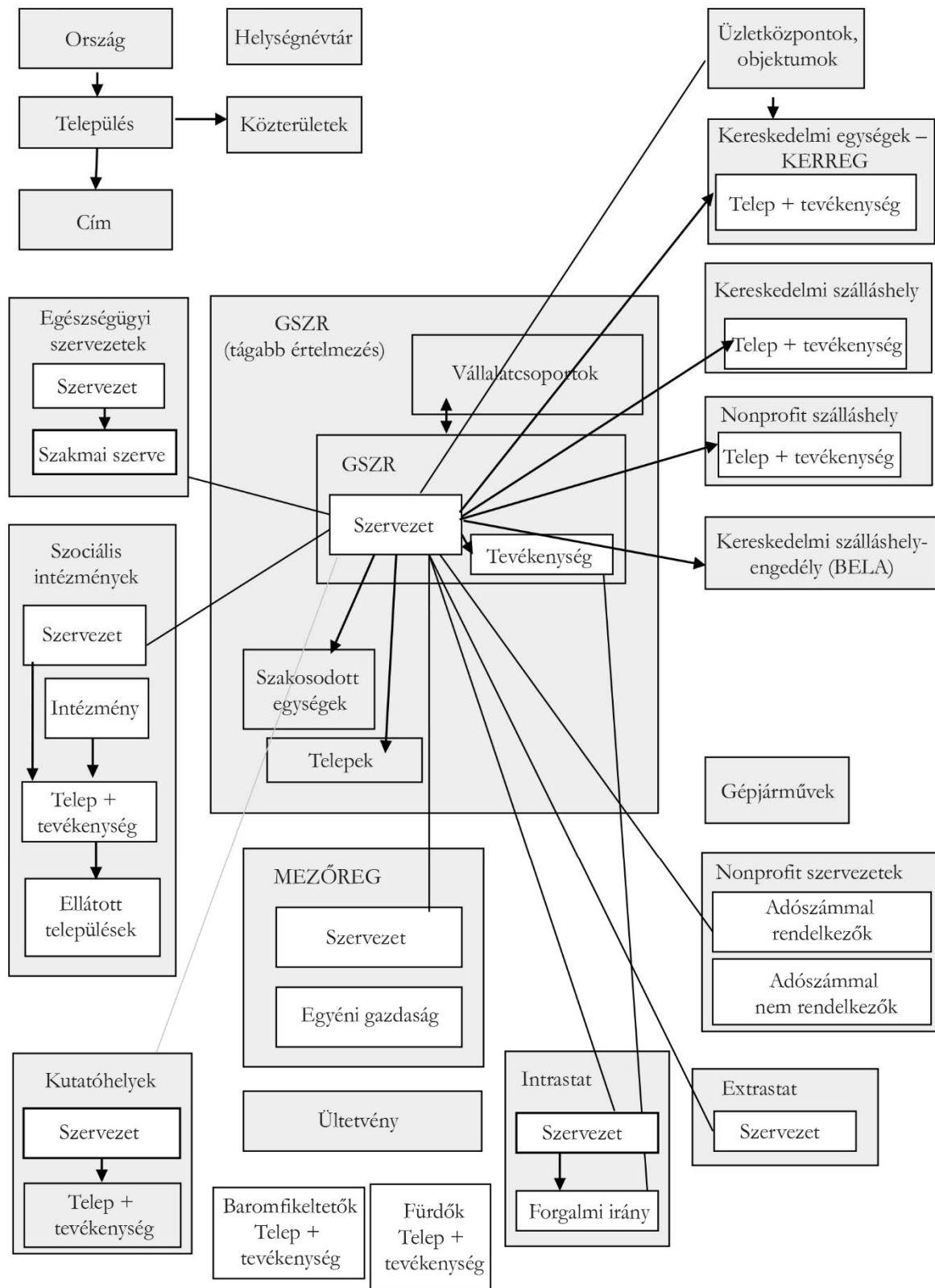
- a településregiszter és a kapcsolódó címregiszter,
- valamint a GSZR.

A statisztikai alapregiszterek által meghatározott két csoportba 2011-ben az alábbi regiszterek, nyilvántartások tartoztak:

- A közigazgatási és földrajzi egységeket leíró regiszterek:
 - országregiszter,
 - településregiszter,
 - címregiszter,
 - helységnévtár,

- közterület-nyilvántartás.
- A gazdasági, társadalmi egységeket, intézményeket és tevékenységüket leíró regiszterek
 - A GSZR és kapcsolódó regiszterei:
 - gazdasági szervezetek regisztere,
 - szakosodott egységek nyilvántartása,
 - telepek nyilvántartása,
 - vállalatcsoportok regisztere.
 - Kereskedelem, szolgáltatás nyilvántartásai, amelyek adószámmal rendelkező egységek és telepek, így szoros kapcsolatban vannak a GSZR-rel:
 - KERREG, kereskedelmi egységek és tevékenységük nyilvántartása,
 - kereskedelmi szálláshelyek nyilvántartása,
 - kereskedelmiszálláshely-engedélyek nyilvántartása, ami a szálláshely-nyilvántartás alapját képező nyilvántartás,
 - nem üzleti céllal üzemeltetett szálláshelyek üzemeltetőinek nyilvántartása,
 - utazási irodák nyilvántartása,
 - üzletközpontok nyilvántartása, ami a GSZR-hez nem kapcsolódó, inkább területi egységet képező nyilvántartás,
 - gépjárművek nyilvántartása.
 - Mezőgazdasági és környezetvédelmi egységek nyilvántartása, amelyek az adószámmal nem rendelkező egyéni gazdaságok és ültetvények kivételével kapcsolódnak a GSZR-hez:
 - mezőgazdasági szervezetek és egyéni gazdaságok regisztere,
 - szőlő- és gyümölcsültetvények regisztere,
 - baromfikeltetők nyilvántartása,
 - fürdőszolgáltatást nyújtó szervezetek nyilvántartása.
 - Szociális és egészségügyi intézmények nyilvántartásai, amelyek adószámmal rendelkező része szoros kapcsolatban vannak a GSZR-rel:
 - szociális intézmények (adószámmal rendelkező és nem rendelkező intézmények) és telepeik regisztere,
 - egészségügyi szervezetek és tevékenységük regisztere.
 - Nonprofit szervezetek regisztere. Az adószámmal rendelkező szervezetek nagyobb része kapcsolódik a GSZR-hez.
 - Kutatóhelyek nyilvántartása (GSZR-ben levő szervezetek és telephelyeik).
 - Külkereskedelmi tevékenység nyilvántartása (GSZR-ben levő szervezetek tevékenysége):
 - Európai Unión belüli külkereskedelmi tevékenység nyilvántartása,
 - Európai Unión kívüli külkereskedelmi tevékenység nyilvántartása.
- Közösségi elhelyezést szolgáló intézmények, intézetek nyilvántartása, ami a címregiszterhez kapcsolódó független regiszter.

KSH regiszterei, nyilvántartásai és kapcsolataik



A KSH nyilvántartásait, regisztereit és azok kapcsolatát a 3. ábra foglalja össze. Az ábrán vonallal megjelölt kapcsolatok a regiszterek között létező, ellenőrzött kapcsolatokat jelölnék. A regiszterek téglalapjain belül jelölt fehér egységek a regiszteren belül kezelt nyilvántartási egységek.

A fenti statisztikai regiszterek karbantartása több forrásra épül:

- elsődlegesen adminisztratív regiszterekre (NAV, cégbíróság, kincstár, okmányirodák (KEKKH), OITH, önkormányzati nyilvántartások, építési hatóságoktól származó információk stb.);
- jelentős szerepet töltenek be a censusok egyes regiszterek karbantartásában (mint a címregiszter, a mezőgazdasági és ültetvényregiszter);
- regiszter-karbantartó adatgyűjtések is segítik a nyilvántartási egységek karbantartását vagy ellenőrzését (nonprofit, szociális regiszter, GSZR stb.);
- bizonyos ismérvek karbantartása statisztikai adatfelvételtől származó információk alapján történik (pl. a szervezetek statisztikai főtevékenysége, létszám- és árbevétel nagyságkategóriái);
- hivatalos közlönyök és más források is alapul szolgálhatnak a regiszterek karbantartásához (pl. településregiszter).

Az adminisztratív nyilvántartási egységek és a statisztikai nyilvántartási egységek eltéréseinek jó példája a GSZR. A gazdaságstatisztika alapját az ESR²² rendszerben az ún. statisztikai egységek képezik, mint vállalkozás, telep, tevékenységi egység, helyi tevékenységi egység. Ezek adminisztratív forrásokból a teljes elvárt sokaságra vonatkozóan közvetlenül nem állíthatók elő. A vállalkozás helyett, közelítésképpen a jogi egységet, az adószámmal rendelkező szervezeteket alkalmazza a gazdasági szervezetek regisztere. A jövőben az Eurostat elvárásai miatt valószínű, hogy ez nem tartható, és az adminisztratív egységekre alapozva, de a vállalkozásokkal való közvetlen kapcsolat alapján (ún. profilképző módszerrel) kell meghatározni a „vállalkozás” típusú statisztikai egységeket. A tevékenységre vonatkozó egységek is csak a szervezetekkel való megállapodások illetve a statisztikai felvételek alapján képezhetők.

Mindazonáltal a cél az, hogy minél inkább az adminisztratív regiszterek váljanak a regiszter-karbantartás alapjává. Ennek nemcsak a gazdaságosabb üzemeltetés az oka, hanem az adminisztratív és statisztikai regiszterek közötti nagyobb összhang megteremtése is.

8. Az adminisztratív regiszterek felhasználása a statisztikában

Amint láttuk, az adminisztratív regiszterek legfontosabb statisztikai felhasználási területe az információ biztosítása a statisztikai regiszterek számára. A statisztikai regiszterek létrehozása az adminisztratív regiszterekből három módon lehetséges:

- Több forrás felhasználásával, ha nem áll rendelkezésre egy olyan adminisztratív regiszter, amely lefedné a statisztikai regiszter sokaságát, illetve nem kezeli valamennyi statisztikai regiszter által igényelt ismérvet. Erre példa a GSZR, aminek bizonyos részsokaságaira, ismérveire vonatkozó információk a cégbíróságtól, másokra az okmányirodából, adóhatóságtól, kincstártól, postától stb. szerezhetők be.

²² Európai Statisztikai Rendszer (rövidítve: ESR), European Statistical System (rövidítve: ESS).

- Az optimális megoldás mind a statisztikai, mind a kormányzati hatékonyság szempontjából az, ha az adott területre központi adminisztratív regiszter áll rendelkezésre, és minden változás egy rendszerben módosul. A felhasználási területek ebből az egy regiszterből kapják az információt.
- Az állampolgárok számára elfogadható megoldás az ún. egyablakos rendszer, amikor az információkat több helyen kezelik, de bárhol bejelenthető a változás, amely minden kapcsolódó rendszerben átvezetésre kerül. A rendszerek szinkronizáltak maradnak. Ennek példája több országban megtalálható, amikor egy személy bejelenti a lakcímváltozását, a bejelentést követően minden olyan rendszerben módosul a cím, ahol a személy lakcímét kezelik.
- Problémát jelentenek azok az információk, amelyek adminisztratív szempontból a statisztikai igénynél kevésbé fontosak, ezért az adatgyűjtő szervezetek ezek minőségére nem fordítanak megfelelő figyelmet. Az ausztrál statisztikai hivatal például az adóhivatalok munkáját a szervezet tevékenységváltozásának bejelentésénél szakértővel támogatja. Ezzel biztosítja, hogy a statisztikai szempontból kiemelt jelentőségű főtevékenység meghatározása a statisztikai szempontok alapján történjen.
- Nálunk kevésbé elterjedt, de a jövőben várhatóan gyakoribb lesz a nemzetközi környezetben már ismert megoldás, amikor több cég kezeli a sokaság elemeit egyenként meghatározott részsokaságokra specializálódva. Egy közös alkalmazás (data-sharing hub) segíti a közös elérést, felhasználást. Ilyen lesz statisztikai területen az Eurogroup regiszter a nemzetközi gazdasági szervezetekről. Természetesen ez nemcsak nemzetközi vonatkozásban, hanem a különböző kormányzati szervezetek közötti adatmegosztásban is szerepet játszhat.

Az így létrejött statisztikai regiszterek, az alap- és a kapcsolódó szatellitregiszterek (lásd 6. pont) elsődlegesen a statisztikai adatfelvételek alapjául szolgálnak. A regiszterek adott időponti állapotai, illetve azok meghatározott részhalmazai szolgálnak felvételi keretként a sokaság meghatározásához, a sokaság elemeivel, az adatszolgáltatókkal való kapcsolatfelvételhez. Reprezentatív felvételek esetén ezek szolgáltatják a mintavételi keretet, illetve a mintarétegek meghatározásához a rétegeképző ismerveket. A statisztikai felvételek végrehajtása, a kérdőívek begyűjtése, az adatok ellenőrzése, becslése, feldolgozása, elemzése is a felvételi keretekre, illetve a regiszterekre épül.

A regiszterek másik nagyon fontos szerepe a különböző forrásból származó *adatok integrálása, összekapcsolása* (lásd 4. ábra). A legegyszerűbb a helyzet, ha a különböző forrásból származó adatok ugyanarra a regiszterre épülnek, ugyanazt az azonosítót használják.

4. ábra

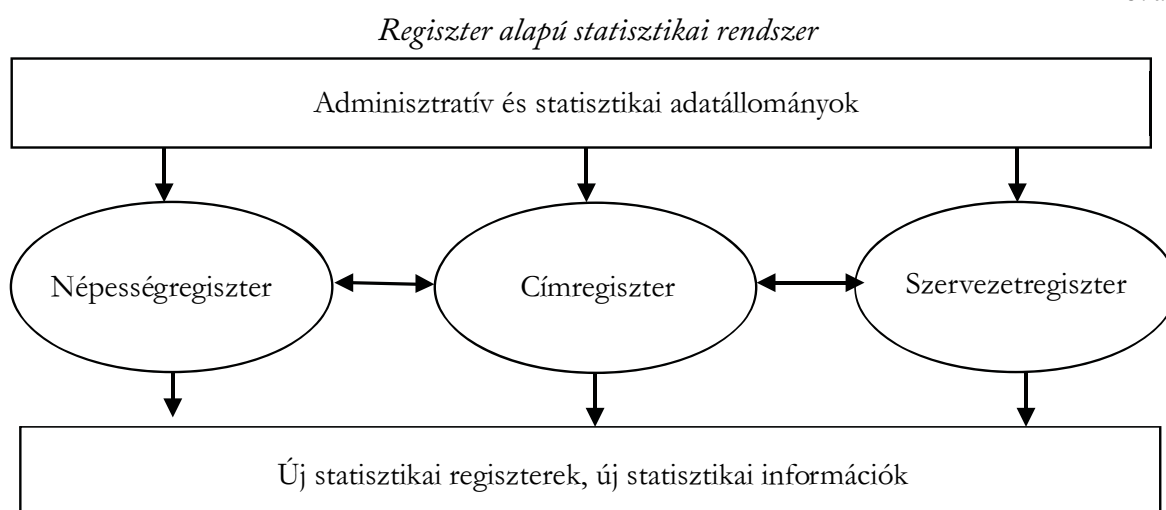


Amennyiben nem áll rendelkezésre közös azonosító, közösen alkalmazható regiszter, akkor az összekapcsolás csak természetes azonosítók vagy egyéb ismérvek alapján lehetséges, esetleg valószínűségi eljárások alapján, amelyek eredményessége, hatékonysága és pontossága messze elmarad az azonosítók alapján történő adatkapcsolásétól.

A következő, egyre jelentősebb felhasználása az adminisztratív regisztereknek, hogy közvetlenül vagy statisztikai regiszter előállításán keresztül a statisztikai adat forrásává válnak, statisztikai output állítható elő belőlük. Részből az egységekre vonatkozó ismérvek a sokaság egészére illetve a részsokaságokra vonatkozóan adhatnak információt. Az egységek demográfiája, változása képet adhat egy adott társadalmi jelenség alakulásáról.

Több regiszter összekapcsolásával új nyilvántartási egységek, új változók, új információk is előállíthatók. Erre épül a regiszter alapú statisztikai rendszerek fogalma is, ahol az összekapcsolt regiszterek és a rajtuk keresztül kapcsolható egyéb adminisztratív állományok adatfelvételek nélkül teszik lehetővé, hogy teljes körű adatok álljanak rendelkezésre a megfigyelendő jelenségekre vonatkozóan. Az adminisztratív forrásokból származó elsődleges sokaság és gyakran a megfigyelt ismérvek is eltérnek a statisztikai megfigyelés céljától, a statisztikai feldolgozás és elemzés feladata ezeknek az eltéréseknek a kezelése (lásd 5. ábra).

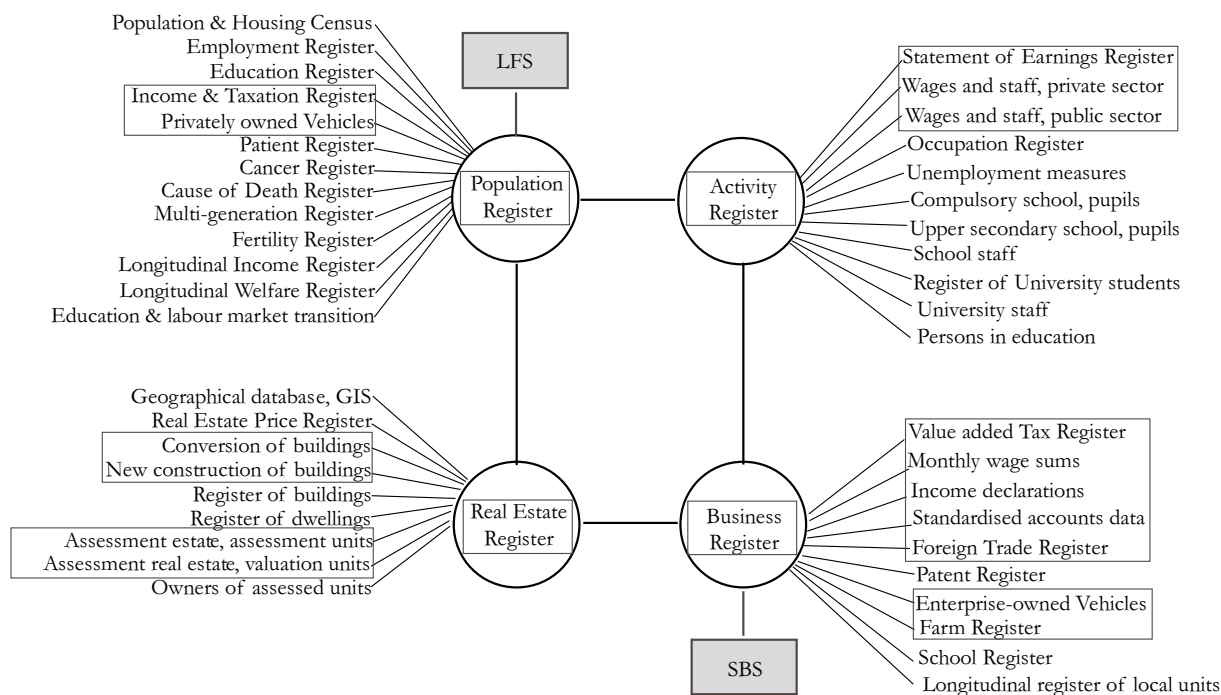
5. ábra



Ennek példája az északi országok statisztikai rendszere, amelyben a nemzeti számlák adatait az adminisztratív alap- és specializált regiszterek és statisztikai adatfelvételek, mint éves szerkezeti statisztika (SBS) és a munkaerő-felmérés (LFS) segítségével állítják elő. A 6. ábrában a bekeretezett regiszterek és adatállományok biztosítják a nemzeti számlák inputját²³.

²³ Wallgren, Anders and Britt: To understand the Possibilities of Administrative Data you must change your Statistical Paradigm! Joint Statistical Meetings (JSM) - Section on Survey Research Methods, Miami Beach, Florida, 2011.
http://www.amstat.org/sections/srms/proceedings/y2011/Files/300347_64422.pdf

Nemzeti számlák előállítása az északi országok statisztikai rendszerében



Az északi országokban a népszámlálás adatai is képezhetők a fenti módon. Az adminisztratív adatforrásokon alapuló népszámlálás témáját a V. fejezet részletezi.

Külön kiemelendő a regiszter alapú statisztikában a földrajzi információs rendszerek (GIS) egyre növekvő szerepe. Korábban az adminisztratív területi egységek képezték a területi statisztika alapját. Mivel egy földrajzi terület adminisztratív határai területszervezési döntések következtében változhatnak, illetve mivel nő az igény az adminisztratív határoktól független és az adminisztratív területi egységeknél kisebb bontású adatok iránt is, új eszközöket kell használni. Ezt szolgálja, hogy a címregiszter tartalmazza a címek, épületek földrajzi helyét, koordinátáit (geokódját) is. A regiszterek összekapcsolása révén lehetővé válik a más címmel ellátott nyilvántartási egységek földrajzi helyének megállapítása is. Erre épít az ún. pont alapú statisztikai rendszer, amely a regiszterekből és a kapcsolódó adatállományokból tetszőlegesen rugalmas területi statisztika képzésére alkalmas.

9. A regiszterek leírása – metaadatok és minőség

Mivel az adminisztratív regiszterek a statisztikai regiszterek előállításában és az adatok integrálásában nagyon fontos szerepet töltenek be, a dokumentálásuk, egyértelmű leírásuk a pontos felhasználás érdekében különös jelentőséggel bír. A KSH-ban kidolgoztuk a statisztikai regiszterek dokumentálásának sémáját.

A leírások részben szöveges dokumentációk (módszertani dokumentáció, minőségjelentés), részben pedig strukturált metaadatok (fogalmak, osztályozások, adatforrások, jogi alapok, nyilvántartási egységek szerkezete stb.). A szöveges leírások a felhasználói igényeknek megfelelően kétféleképpen lehetnek: vannak rövid leírások és hosszabb ismertetések a szakértő felhasználók számára. Minden szöveges leírás mind magyar, mind angol nyelven elkészül azokra az információkra, amelyek közzétehetőek. A szöveges leírások a változásokat egy

dokumentumban összefogó történeti leírások. A strukturált metaadatok évenkénti pontos dokumentációt adnak az adott évben élő metaadatelemekről.

A regiszterekre vonatkozóan a dokumentációnak a következő szerkezeti elemei vannak:

- A. A regiszter mint szakstatisztika dokumentációja.** A regiszterek nemcsak az adatfelvételek forrásai, hanem önálló szakstatisztikák is, az objektumok demográfiája készíthető el belőlük. Ezért a szakstatisztikák dokumentációs sémája rájuk is alkalmazható. Ennek keretében az alábbi információk megadását várjuk el:
- a. Azonosító adatok: a regiszter megnevezése, rövid neve, a regiszterért felelős főosztály, személy neve, azonosítója.
 - b. A regiszter létrehozásának, karbantartásának hazai és nemzetközi jogi alapja.
 - c. A regiszter célja.
 - d. A regiszter rövid tartalmi leírása, beleértve a regiszter sokaságát, nyilvántartási egységét (egységeit), megfigyelt ismérveit, történeti adatok elérhetőségét.
 - e. A nyilvántartási ismérveknél alkalmazott fontosabb osztályozások felsorolása.
 - f. A regiszter forrásainak, adatgyűjtési módjainak rövid ismertetése.
 - g. A regiszteradatok és a regiszterből készült elemzések közzétételi módjai, adatátadások.
 - h. A regiszter időszerűsége és revíziós politikája, beleértve a karbantartás gyakoriságát, az adminisztratív források kezelési módját, javíthatóságát.
 - i. A regiszter története összefoglalja a regiszter tartalmában, forrásaiban, módszerében bekövetkezett jelentősebb változásokat.
 - j. A regiszter rövid (és esetleg hosszú) módszertani leírása ismerteti a regiszter forrásainak beérkezési és felhasználási módját. Több forrás esetén a prioritások, a változások, a történeti adatok kezelését, a karbantartási periódusok szerepét, a fontosabb ellenőrzéseket, a hiányzó, hibás adatok kezelését, a regiszteradatok feldolgozását, a felvételi keretek kialakításához kapcsolódó adatátadást, a demográfiai előkészítést, más rendszerekkel való kapcsolatokat stb.
- B. A regiszter tartalmának részletes strukturált leírása** évente pontos leírást ad a regiszter tartalmáról, szerkezetéről, kapcsolatairól:
- a. A regiszter leírása megadja a regiszter azonosítóit, sokaságát, érvényességét, változáskezelésének típusát, a karbantartás gyakoriságát, közhitelességét, a regiszterkarbantartás felelősségi rendjét.
 - b. A regiszter nyilvántartási egységeinek leírása a nyilvántartási egység típusát, megnevezését, érvényességét, sokaságának leírását, az adatbázistábla nevét és tulajdonosát adja meg.
 - c. A nyilvántartási egység ismérveinek leírása ismerteti a regiszter szerkezetét, megadja az ismérvek azonosítóját, megnevezését, az értékészletét leíró nomenklatúrát és annak adott évre érvényes változatának azonosítóját. Az ismérvet statisztikai vagy adminisztratív jellegével, a közhitelességgel és a szerepével (azonosító, elérést támogató, csoportosító, demográfiai, kapcsolatot leíró, karbantartást szolgáló vagy felhasználási folyamatot támogató ismérv) jellemzi.

- d. A nyilvántartási egység kapcsolatainak leírása megadja a nyilvántartási egység regiszteren belüli, valamint más regiszterek nyilvántartási egységeivel való kapcsolatait.
- e. A regiszter befagyasztott állapotainak leírása megadja azokat a regiszterállapotokat, amelyeket elmentünk, és amelyeket a felvételek keretének meghatározásához felhasználunk.

C. Más metaadatokkal való kapcsolat

- a. A regiszter-adatgyűjtés adatforrásai évente. Az adatgyűjtés leírása megadja az adatgyűjtés nevét, elrendelőjét, adatszolgáltatói körét, témáját, egyéb, az OSAP-ban meghatározott ismérveit, ezek mellett az adatgyűjtéshez tartozó kérdőívet is tartalmazza.
- b. Adminisztratív regiszterek, adatok évenkénti átvétele karbantartás céljából. Az adminisztratív forrás leírása tartalmazza az átvett adatok körét és az átadó szervezetet.
- c. A regiszter adatforrásaként felhasznált más regisztereknél elérhetőek a forrásregisztereknek e dokumentumban leírt struktúraadatai.
- d. Szakstatisztikai adatok átvételének évenkénti leírása tartalmazza az átvett adatok körét.
- e. „A regiszterek felhasználása más regiszterekben és szakstatisztikákban” jelöli meg a felhasználás módját (felvételi, mintavételi keret, ellenőrzés, ismérvek átvétele).
- f. A regiszterekben használt fogalmak és kapcsolataik leírása során minden olyan fogalomhoz pontos definíciót kell adni, amely egzakt leírást igényel, így a nyilvántartási egységekre, azok fontosabb részalmazaira, és az egyes ismérvekhez is.
- g. A regiszterben használt osztályozások és kapcsolataik leírása során ismertetni kell az osztályozás célját, jogi alapját, módszerét, más osztályozásokkal való kapcsolatát. Egyúttal le kell írni az osztályozás elemeit, illetve azok felhasznált változatait is.

Ez utóbbi kettő különös jelentőséggel bír a regiszterek felhasználásánál és a statisztikai és adminisztratív regiszterek kapcsolatának elemzésénél.

D. A regiszter minőségének leírása is részét képezi a regiszterek dokumentációjának. A regiszterek minőségének van néhány speciális jellemzője statisztikai adatok minőségéhez képest, ezeket külön kiemeljük.

- a. A regiszter relevanciája kérdéskörben kell ismertetni, hogy kik a regiszter felhasználói, és mik az adatigényei; mi a regiszter sokasága és nyilvántartási egységei; valamint azt, hogy a regiszter milyen mértékben elégíti ki a regiszterrel szemben támasztott igényeket, melyek azok az elvárások, amiknek eleget tesz, és melyek azok, amiknek nem.
- b. A regiszter időszerűsége fontos szempont a statisztikai regiszter minőségének értékelésénél, mivel meghatározza a felvételi keretek aktualitását, pontosságát is. Az időszerűséget az események bekövetkeztétől a regiszterbe kerülésig eltelt idővel mérhetjük. Mivel ez nem könnyű feladat, érdemes annak vizsgálatával közelítenünk a problémához, hogy egy karbantartási periódus alatt mennyi változás történik (új, megszűnt, módosult egységek száma). Kevésbé gyakran módosuló regisztereknél a

ritkább karbantartás nem okoz olyan mértékű minőségromlást, mint a gyakrabban módosuló regisztereknél.

- c. A regiszter pontossága arra vonatkozik, hogy a regiszter nyilvántartási egységeinek halmaza, a nyilvántartási egységek és ismérvek mennyire felelnek meg a valóságnak. Több szempont szerint vizsgáljuk a pontosságot.
- i. A lefedettségi többlet és hiány mutatja, hogy a regiszterben nyilvántartott egységek mennyire fedik le a célsokaságot. A lefedettségi többletbe tartoznak azok az egységek, amik soha nem tartoztak vagy már nem tartoznak a célsokasághoz. Gazdasági szervezetek esetén ide sorolhatók a duplán regisztrált egységek vagy azok, amik működésének megszűntét nem jelzik a regiszterben. A lefedettségi hiányba tartoznak azok az egységek, amiket nem vagy nem időben regisztráltak. Lefedettségi hibát jelez az is, ha az egység maga létezik a regiszterben, de hibás állapotjelzése, érvényességi ideje alapján mégsem tekintjük egy adott időszakban élő egységnek. A lefedettségi hiány és többlet a sokaság egészéhez viszonyított aránnyal mérhető.
 - ii. A besorolási hibákhoz tartoznak a nyilvántartási egység csoportosító ismérveinél jelentkező hibák, vagyis azok az esetek, amikor nem a valóságnak megfelelő ismérvértéket vesz fel a mező. Például a szervezetek hibás főtevékenysége, nagyságkategóriája, egy cím hibás rendeltetése.
 - iii. Az elérési hibák a kapcsolattartásra szolgáló információk hibáját jelzik. Érvénytelen, nem létező vagy nem a valóságnak megfelelő címek, telefonok, kapcsolattartó személyek tartoznak ebbe a kategóriába.
 - iv. Egyéb hiba lehet például, ha a nyilvántartási egységek más egységekhez való kapcsolatai hibásak.

A regiszter pontossága meghatározza a regiszterre épülő felvételi keretek és ezáltal a felvételek pontosságát is, ezért a mérésüknek nagy szerepe van a minőségkezelésben.

- d. A regiszter hozzáférhetősége, érthetősége alatt egyrészt a regiszterben levő információk elérhetőségét, az adatokhoz való hozzájutás lehetőségét (elérési csatornák, alkalmazások, jogosultság rendszerét), másrészt az információk értelmezéséhez adott segítséget értjük. Ide tartozik, hogy a különböző felhasználói igényeknek megfelelő dokumentáció, metaadatok rendelkezésre állnak-e.
- e. A regiszter összehasonlíthatósága, koherenciája a regiszteren belüli és a regiszterek közötti módszertani, definícióbeli, osztályozásbeli differenciák hatásának vizsgálatát jelenti időben, térben és a regiszterek között.
- i. Az időbeni összehasonlíthatóság a módszertani változások miatt az adatokban jelentkező inkonzisztenciát jelenti, amikor a korábbi adatokon nincsenek a változások visszavezetve, ezért csak bizonyos töréspontokig hasonlíthatók össze az adatok.
 - ii. Térbeli összehasonlíthatósági problémák jelentkezhetnek területi rendszereknél, ha bizonyos módszertanokat egyes vidékeken másképp értelmeznek, pontosítanak. De ennél fontosabb a nemzetközi

összehasonlíthatóság, azaz, hogy mennyire felel meg a regiszter a nemzetközi elvárásoknak.

- iii. A tárgykörök szerinti összehasonlíthatóság keretében kell vizsgálni az egymásra épülő, együttműködő adminisztratív és statisztikai regiszterek koncepcionális eltéréseit (sokaságát, nyilvántartási egységeit, ismérveit, fogalmait, osztályozásait), a megvalósítás gyakorlati problémáit, az ellenőrzések, összefüggés-vizsgálatok, besorolási hibák elvi okait, karbantartási, adatátadási időkből adódó minőségi problémákat. Az eltérések okainak feltárásával, elemzésével és a regisztergazdák közötti egyeztetéssel, a módszertan koordinációjával jelentős minőségjavulást lehet elérni.

III. A másodlagos adatok statisztikai felhasználási területei

Másodlagos adatok gyakorlatilag az összes statisztikai munkafázisban alkalmazhatók. Az alábbi megállapítások jó része helytálló abban az esetben is, ha más szervezet statisztikai célra gyűjtött adatainak hasznosításáról van szó. Lényeges különbség akkor mutatkozik, amikor a statisztikai szervezet mintavételen alapuló statisztikai adatokat vesz át további felhasználásra, hiszen – ahogy erről már korábban volt szó – a nyilvántartások célsokaságukat tekintve teljes körűek, aminek a statisztikai felhasználás lehetőségeit illetően kulcsszerepe van. Mintavételen alapuló statisztikai felvétel adatai nem használhatók olyan célra, ahol éppen a teljességet kellene kihasználni, például statisztikai regiszter alapadatforrásaként.

Más szempontból korlátozott a big data típusú adatok nagy részének statisztikai hasznosítása. A tranzakciós adatok, a közösségi hálók adatai, mobiltelefonok és GPS-ek nyomkövetési adatai anélkül gyűlnek, hogy meg lehetne határozni a célpopulációt, tehát esetükben nem beszélhetünk sem teljességről, sem reprezentativitásról. Ennek következtében ezek az adatok a hivatalos statisztikában jobbra csak kiegészítő információkat szolgáltathatnak, közvetlen statisztikai adatgyűjtést nem váltanak ki.

A másodlagos adatok statisztikai célú hasznosítását alapvetően azok minősége határozza meg (lásd VI. fejezet), melyet minden felhasználás előtt értékelni kell. A külső forrásból származó adatok minősége általánosan is jellemezhető az adatgazda, az előállítási folyamat, valamint a statisztikai termékeket jellemző minőségi összetevők szempontjából. Az egyes minőségi jellemzők fontossága azonban más és más lehet, aszerint, hogy mire akarjuk az adatot felhasználni. Ezért az általános kritériumok szerinti minőségértékelés mellett az adott felhasználási cél szempontjából is szükséges elvégezni az értékelést. A felhasználást legtöbbször az alábbi tényezők együttese határozza meg:

- az adatgyűjtő rendszerben használt fogalmak, definíciók;
- az adatok vonatkozási köre, lefedettsége;
- az adatgyűjtő rendszerre vonatkozó minőség;
- az adatok rendelkezésre állása.

Amennyiben valamelyik tényezőt tekintve eltérés mutatkozik a statisztikai célban meghatározott elvárásoktól, az jelentősen befolyásolhatja az átvett adatok felhasználhatóságát.

1. Felvételi keret, statisztikai regiszterek építése

Az adminisztratív nyilvántartások – teljeskörűségüknek köszönhetően – gyakran szolgálnak mintavételi keretként. Ha a mintaválasztás kifejezetten személyekre irányul, a felvételi keret rendszerint a népesség-nyilvántartás, ami tartalmazza a lakossági minták kiválasztásához szükséges legfontosabb demográfiai adatokat (nem, életkor, családi állapot, lakóhely, állampolgárság). Egyes országokban maguk a statisztikai hivatalok üzemeltetik a népesség-nyilvántartást, Magyarországon azonban nem ez a helyzet, nálunk jelenleg a Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala a felelős az adatállomány karbantartásáért. Frissítése folyamatosan történik, ezért bármely időpontra az aktuálisan érvényes állapot nyerhető ki belőle. Statisztikai szempontból kedvezőbb lenne egy olyan regiszter fenntartása, ami az egyénekre vonatkozóan a jelenleginél több információt tartalmaz, például elveszületések száma, egyenes ági rokoni kapcsolatok, iskolai végzettség, gazdasági aktivitás. A skandináv országokban egy ilyen bővebb regiszter a népszámlálások személyekre vonatkozó mutatóinak alapvető adatforrása. Ezek a jellemzők azonban lényegesen gyorsabban változnak, mint a hazai gyakorlatban kezeltek, ezért karbantartásuk számottevően több erőforrást igényelne, továbbá adatvédelmi szempontból növelheti a hatóságok és a közvélemény aggályait is. Ezért egy bővebb népesség-nyilvántartás Magyarországon még várat magára.

Amikor a végső mintavételi egység a háztartás, a KSH által karbantartott címregiszter szolgál a lakossági összeírások mintavételi keretének alapjául, mivel a háztartásokhoz lakcímek rendelhetők. A címregiszterben egy rekord tartalmazza a település, településrész azonosító adatait és a pontos címadatokat. A címregiszter karbantartására több forrás kerül felhasználásra. Az alapadatokat az önkormányzatok által szolgáltatott területváltozási adatok és az építési hatóságok által jelentett lakás-használatbavételi és megszűnési jelentések, továbbá a posta irányítószám-állománya szolgáltatja. Kiegészítő, pontosító információk nyerhetők a társregiszterek címadataiból is.

A KSH-ban működtetett gazdasági szervezetek regiszteréhez a magyar gyakorlat szerint az alapadatokat a cégbíróságoktól és az okmányirodáktól érkeznek, amik az adóhatóság rendszeres karbantartó állományával naponta frissülnek. A költségvetési szerveknél történő változások a Magyar Államkincstár nyilvántartásából havonta kerülnek be a regiszterbe.

A KSH kezelésében léteznek további, bizonyos speciális ágazatokra vonatkozó statisztikai regiszterek is, például egészségügyi szolgáltatók regisztere, szociális intézmények és szolgáltatók regisztere. Ezek a gazdasági szervezetek szatellitregisztereinek tekinthetők, azaz egyéb igazgatási és statisztikai adatforrások alapján egy-egy jól definiált csoportra vonatkozóan a statisztikai felhasználás által meghatározott igények szerinti részletesebb adatokat tartalmazzák.

2. Statisztikai adatgyűjtések helyettesítése

A nyilvántartásoknak az erőforrásokat leginkább kímélő felhasználási módja, amikor egy vagy több adminisztratív adatforrás adataival teljes statisztikai adatfelvételt helyettesítünk. Kevésbé hatékony, de az adatszolgáltató terheit mindenképpen csökkentő eljárás, ha a statisztikai adatgyűjtés információtartalmát külső adatforrás adataival egészítjük ki. Erre példa a kérdőív ún. megszemélyesítése, amikor az adatszolgáltató úgy kapja meg a statisztikai felmérés kérdőívét, hogy az azonosító adatokat a statisztikai szervezet már előre beírta, és azt a kitöltőnek

csak ellenőriznie kell, így hosszú szövegek, kódszámok beírásától mentesül, és ez csökkenti az adatszolgáltatásra fordítandó idejét, energiáját, miközben javítja az adatok pontosságát. Így a felmérésekben az adatszolgáltatás és a feldolgozás során is a legfontosabb kérdésekre lehet koncentrálni.

Ha egyetlen külső adatforrás alkalmas arra, hogy helyettesítsen egy statisztikai adatfelvételt, az adatok felhasználása technikailag hasonlít egy statisztikai adatfelvétel beérkezett eredményeinek feldolgozásához. Például logikai ellenőrzéseket kell végezni annak érdekében, hogy a hibáktól megtisztítsuk az állományt. Gyakori hiba, hogy a tizedesvessző rossz helyen van, vagy figyelmen kívül hagyták azt az utasítást, miszerint az összegeket ezer forintban kell megadni.

Itt érdemes szólni azokról a statisztikai és adminisztratív szervezetek közötti együttműködésekről, amelyek alapján olyan igazgatási nyilvántartásokat alakítanak ki, amelyek – a statisztikai igényeket figyelembe véve – önmagukban is biztosítják egy adott szakstatisztika bizonyos indikátorkörének adatforrását.

Teljes adatgyűjtés kiváltása egyetlen adminisztratív adatforrásból sokszor éppen azért nem lehetséges, mert az adminisztratív és a statisztikai cél nem azonos. Ha egy-egy becslést csak több adatforrás együttes felhasználásával lehet elvégezni, akkor több – nem feltétlenül adminisztratív – adatállomány összekapcsolása szükséges. Ha a mikroadatok pontossága is fontos, ennek feltétele a sokaság elemeinek egyértelmű azonosítása.

Több másodlagos adatforrás egyidejű felhasználása alapvetően négyféle módon történhet, de ezeknek különböző kombinációi is elképzelhetők.

1. Ha több olyan adatforrás is rendelkezésre áll, amely közvetlenül tartalmazza egy ismerv értékét, első lépésként ki kell választani azt az adatforrást, amelyik az adott szempontból a legmegbízhatóbb. A többi adatforrást csak akkor kell felhasználni, ha az elsődleges adat hiányzik vagy hibás. Ezt a módszert kevésbé érintik a nyilvántartások tartalmi változásai, mivel az egyedi adatok általában több forrásból is származtathatók.
2. Az adatforrások tartalmilag kiegészíthetik egymást. Ez esetben egy megfigyelési egységre vonatkozóan a szükséges adatok egy részét az egyik, a másikat egy másik állományból nyerhetjük ki.
3. Az egyik adatforrás egyedei a másikéihoz hozzáadódnak, vagyis a célpopuláció teljes népessége több adatforrásból adódik össze.
4. Az is lehetséges, hogy azért használunk több másodlagos adatforrást, mert az egyes állományokból származó különféle információk mérlegelésével a sokaság elemeire egy új változó értékét szeretnénk megállapítani. Ebben az esetben a korábbi információkból kiindulva fel kell állítani egy modellt, vagy kialakítani azt a döntési szabályrendszert, amelynek segítségével az adminisztratív adatok alapján a statisztikai fogalmaknak megfelelő lehető legjobb becslés adható.

Napjainkban – bár az adminisztratív adatforrások egyre nagyobb teret hódítanak – a nemzetközi gyakorlatban sok szakstatisztikai területen a kombinált adatforrások használata jellemző, azaz az adminisztratív adatforrásokból hiányzó adatokat statisztikai felmérések adataival pótolják. Bármilyen jó minőségű nyilvántartások álljanak is a statisztikai felhasználás rendelkezésére, bármilyen együttműködésre készek legyenek is a partnerek, mindig lesznek olyan információk, amik egy igazgatási vagy egy szolgáltatási regiszterben sosem kerülnek

rögzítésre. (Jó példák erre az attitűdöket, véleményeket megragadó változók, amik az utóbbi időben egyre nagyobb jelentőséggel bírnak. Ezen a területen mutatkozik a big data statisztikai célú felhasználásának legnagyobb potenciálja.)

Felhívjuk a figyelmet arra, hogy bármennyire is hasznos egy statisztikai adatgyűjtés kiváltása másodlagos adatforrással, bizonyos kockázatokat is jelenthet az adatminőség szempontjából. Ha ugyanis egy közvetlen statisztikai felvételt azzal a másodlagos adatgyűjtéssel tervezünk kiváltani, amit eddig valamelyik részfolyamatban ellenőrzéshez használtunk, gondolni kell arra is, hogy ezzel csorbítjuk az ellenőrzési, javítási lehetőségeket is. Mérlegelni kell, hogy ez a veszteség milyen hatással van az adatminőségre.

3. Editálás és imputálás

A másodlagos adatforrásokból nyert információk a mintavételes adatfelvételekben összegyűjtött adatok minőségének javítására is felhasználhatók. Az adatgyűjtési folyamat során bármilyen hiba előfordulhat, ami rontja az adatok minőségét. Az esetek egy részében az adatösszefüggések alapján ezek közvetlenül javíthatók. Máskor viszont más, statisztikai vagy adminisztratív, esetleg egyéb big data típusú adatállományokból származó kiegészítő információk bevonásával lehet csak megbízható becslést készíteni a hibás adatok pótlására. „Nemválaszolás” vagy hiányzó információk a statisztikai adatgyűjtésekben mindig előfordulnak. Bizonyos esetekben ezek másodlagos adatforrásokból közvetlenül vagy becslések útján pótolhatók.

4. Közvetett használat a becslésekben

4.1. Kiegészítő adatok

Komplex statisztikai mutatók esetén az indikátorok számos adatforrás együttes felhasználásával határozhatók meg. A külső információk olykor nem is a mutató közvetlen kiszámításához szükségesek, hanem inkább csak annak finomításához, figyelembe véve, hogy egy mutatórendszeren belül milyen feltételezett összefüggések állnak fenn. Másik, itt megemlítendő lehetőség, amikor egy, a teljes célcsoportra érvényes indikátorból az egyes részekre vonatkozó mutatókra szeretnénk következtetni (például kisterületi becslést akarunk készíteni), és ehhez olyan adminisztratív adatforrást használunk, amiből a megoszlási adatok kinyerhetők.

4.2. Kalibrálás

Egy reprezentatívnak tervezett statisztikai felvétel mintája az adatgyűjtés során kisebb-nagyobb mértékben elferdülhet, ezért szükség lehet arra, hogy az egyes megfigyelési egységekhez tartozó súlyokat korrekciós eljárással módosítsuk, azaz kalibrálást hajtsunk végre. A kalibrálásnak két célja van: egyrészt mérsékli a meghiúsulások torzító hatását, másrészt biztosítja, hogy a minta jól reprezentálja a célpopulációt a főbb jellemzők tekintetében. Az eljáráshoz az adatok egy része másodlagos adatforrásból származhat.

4.3. Statisztikai adatfelvételek értékelése, validálása

A validálás munkafázisában azt vizsgáljuk, hogy az előállított statisztikai adatok koherens egységben vannak-e az adott szakstatisztikával kapcsolatos információk összességével.

A másodlagos adatforrások statisztikai adatfelvételekből nyert becslések, aggregált adatok validálásához is alkalmazhatók. A makrovalidálás feltétele, hogy az összehasonlításhoz felhasználandó adat minősége (valamennyi összetevő mentén) legalább olyan jó legyen, mint a validálandó adat, és a két adatforrásból származó mutató megalapozott tartalmi és logikai kapcsolatban álljon egymással.

Az adminisztratív nyilvántartások adatainak egy alkalmazási lehetősége az ún. nemválaszolási vizsgálat, amikor a minta azon részét szeretnénk jellemezni, amelyben a megfigyelési egységekre vonatkozóan megtagadták az interjút, vagy amiket a statisztikai szervezet más okok miatt nem ért el. Kedvező esetben ezekre a mintaelemekre sok, külső adatforrásokból származó információ áll rendelkezésre. A reprezentatív adatfelvételek megbízhatóságát ilyen vizsgálattal lehet értékelni. A válassz megtagadók pótlására példa a KEKKH népeség-nyilvántartásában szereplő alapvető demográfiai és lakcím adatok használata a 2011-es népszámlálás során.

5. Tájékoztatás kiegészítése, gazdagítása

Gyakran előfordul, hogy a más szervezet által – akár adminisztratív, akár statisztikai vagy egyéb célra – gyűjtött adatokból az adatgazda maga is statisztikai mutatókat számol, amiket a statisztikai hivatal a saját tájékoztatási rendszere bővítésére vesz át. Ilyen esetben a statisztikusnak általában kevés információja van magáról az adatállományról és az elemi adatok minőségéről. Rendszeresen átvett indikátorok esetén az idősoron, illetve egy-egy aggregátumon belüli mutatók konzisztenciája vizsgálható. Az adat pontosságára vonatkozóan elsősorban ennek az idősoroknak, illetve konzisztenciának az alapján az indikátor alakulását befolyásoló tényezők ismerete mellett vonhatunk le következtetéseket.

Ezekben az esetekben is fontos ismerni az adatgyűjtés módját és az adatgazda adatkezelési eljárásait, hogy a publikált adatok minőségéért felelősséget vállalhassunk.

IV. Az adminisztratív adatforrások használatának jogszabályi környezete

1. Magyarországi helyzet

Magyarországon az elmúlt másfél évszázad alatt hét statisztikai törvény született. A legutóbbi a hivatalos statisztikáról szóló 2016. évi CLV. törvény (a továbbiakban: statisztikai törvény). Az eddigi statisztikai törvényekben általában volt utalás az adminisztratív adatok statisztikai célra történő felhasználására. A statisztikai szolgálat felállításának kezdetén is nyilvánvaló volt, hogy az adminisztratív adatok felhasználhatók a gazdasági és társadalmi folyamatok bemutatására, statisztikai célú elemzésére.

Mindegyik statisztikai törvény a hazai statisztikai tevékenységek szempontjából kiemelt szerepet szánt a Központi Statisztikai Hivatalnak, így érthető módon az adminisztratív adatok statisztikai célú felhasználásának legfőbb kedvezményezettje a KSH volt. Párhuzamosan a hagyományos statisztikai felvételek nehézségeinek fokozódásával és az adatszolgáltatói terhek növekedésével az adminisztratív adatok statisztikai célú felhasználásával kapcsolatos igények

egyre szaporodtak. Határozott igény mutatkozott arra, hogy a hagyományos statisztikai adatgyűjtéseket – ha ez lehetséges – kiváltsa az adminisztratív adatok statisztikai célú felhasználása. Az adminisztratív adatok statisztikai célú felhasználását az utóbbi évtizedekben az is segítette, hogy az informatika jelentős mértékű fejlődése megkönnyítette a nagy tömegű és bonyolult összetételű adatok átadását, átvételét.

Ennek tudatában már a jelenleg hatályos statisztikai törvényt megelőző, 1993. évi XLVI. törvény alkotói is több rendelkezést iktattak be a témával kapcsolatban.

A 2016 végén elfogadott statisztikai törvény az európai uniós jogi szabályozással összhangban – az annak való megfelelés érdekében – teljesen újraszabályozta az adminisztratív adatforrások felhasználásra vonatkozó rendelkezéseket.

A statisztikai törvény kimondja, hogy elsődleges adatforrást, azaz közvetlen adatgyűjtést csak akkor lehet igénybe venni, ha nincs statisztikai célra alkalmas, hozzáférhető másodlagos adatforrás, ezáltal kihangsúlyozva, hogy az adatszolgáltató magánszemélyeket és vállalkozásokat csak abban az esetben lehessen statisztikai kérdőívek kitöltésével terhelni, ha más adatbázisban nincs olyan felhasználható adat, amely alkalmas a statisztikai cél kielégítésére.

Ahhoz, hogy az adminisztratív adatforrásból átvett adatok alkalmasak legyenek statisztikai célra, a KSH-nak egyértelműen azonosítania kell a kizárólag egyetlen adminisztratív adatforrás felhasználásával előállított statisztikát, kivéve a nyilvántartásban szereplő egyedeket, hiszen csak ekkor tudja az adatokat más állományokkal összekapcsolni, illetve tudja megtalálni a pótolandó, javítandó adathelyeket. Mindezekhez elengedhetetlen ezen adatok adminisztratív adatforrás általi, a nyilvántartási egységek vonatkozásában használt azonosítóval ellátott átvétele (egyedi azonosításra alkalmas adatátvétel).

Mindezek érdekében a statisztikai törvény kötelezi az adminisztratív adatforrást kezelő szervezetet, hogy a KSH részére hivatalos statisztikai célra – a minősített adat kivételével –, a hivatalos statisztikai célnak megfelelően, egyedi azonosításra alkalmas módon, a kapcsolódó kiegészítő információkkal (metaadat) együtt az általuk jogszabály felhatalmazása alapján nyilvántartásban – mint adminisztratív adatforrásban – kezelt adatot átadják.

A fentieknek megfelelően az adminisztratív adatok statisztikai célú átvételéről a statisztikai törvény 28.§-a az alábbiak szerint rendelkezik:

„(4) Az adminisztratív adatforrást kezelő szerv a Hivatalos Statisztikai Szolgálat tagjai részére hivatalos statisztikai célra – a minősített adat kivételével – az (1) és (5) bekezdésben foglaltakra tekintettel a hivatalos statisztikai célnak megfelelő, egyedi azonosításra alkalmas módon, a kapcsolódó kiegészítő információkkal (metaadat) köteles az általa jogszabály felhatalmazása alapján nyilvántartásban mint adminisztratív adatforrásban kezelt adatot átadni.

(5) A – KSH-t kivéve – a Hivatalos Statisztikai Szolgálat többi tagja a (4) bekezdés alapján személyes adatokhoz való hozzáférésre akkor jogosult, ha azt külön törvény elrendeli.

(6) A (4) bekezdés szerinti adatátadás térítésmentes, amennyiben az átadandó adatok elektronikus formában rendelkezésre állnak. Egyéb esetben a térítés felszámítására a 46. § (2) bekezdésében foglaltakat kell alkalmazni.”

Az adminisztratív adatok átvétele tehát nem történhet megfelelő cél nélkül, a releváns, valós hivatalos statisztikai célt előzetesen igazolni szükséges, és az adatátadás részleteiről a statisztikai törvény rendelkezéseinek megfelelően, az ott meghatározott tartalommal együttműködési megállapodást kell kötni az adminisztratív adatforrást kezelő szervezettel. Annak érdekében, hogy az adatátvételek és az azokhoz kapcsolódó adatkezelés átlátható legyen az adatok alanyai számára is, ezeket az együttműködési megállapodásokat a KSH honlapján nyilvánosságra kell hozni.

Az adminisztratív adatok statisztikai célú felhasználásával kapcsolatban hangsúlyozni kell, hogy nem elég csak arról rendelkezni, hogy az adminisztratív adatok statisztikai célra átvehetőek legyenek, hanem azt is szabályozni kell, hogy az adminisztratív célú adatgyűjtések egy későbbi statisztikai célú újrahasznosítás érdekében kompatibilisek legyenek a statisztikai célú adatfelvételek módszertanával, fogalmi rendszerével, és erre már a tervezés időszakában gondolni kell, oly módon, hogy az adminisztratív adatforrások felelősei a KSH véleményét kikérjék.

Erről rendelkezik a statisztikai törvény 28. § (2) bekezdése:

„(2) A hivatalos statisztikai felhasználásra való alkalmasság szempontjából az adminisztratív adatforrások tartalmát, módszertanát, fogalmi és osztályozási rendszerének kialakítását érintő kérdésekben az adatforrás létrehozásakor, módosításakor, megszüntetésekor – az eredeti adatkezelési cél sérelme nélkül – a KSH véleményét figyelembe kell venni.”

Az adminisztratív adatok egyedi azonosításra alkalmas átvétele során a gazdálkodó szervezetek adatai mellett személyes adatok, azon belül bizonyos esetekben különleges adatok kezelésére is sor kerülhet. Annak érdekében, hogy ez ne jelentse az érintettek jogainak sérelmét, az adatkezelés során szigorú szabályoknak kell megfelelni.

A statisztikai törvény 40. §-a rendelkezik a személyazonosításra alkalmas adatok kezelésének szabályairól, amely alapján a személyazonosító adatokat a statisztikai feldolgozás befejezésekor vagy törölni kell, vagy a személyes adatokat tartalmazó adatállományok esetében a statisztikai egység statisztikai adatgyűjtéshez használt közvetlen azonosító adatait különválasztva kell tárolni a statisztikai egységre vonatkozó egyéb adatoktól. Az adatgyűjtéshez használt közvetlen azonosító helyett az egyes statisztikai egységeket technikai azonosítóval kell ellátni, aminek segítségével a gyűjtött adatok és a statisztikai egység adatgyűjtéshez használt közvetlen azonosító adatai közötti kapcsolat helyreállítható.

Az adatállomány és a statisztikai egység statisztikai adatfelvételhez használt közvetlen azonosító adatai közötti kapcsolatot hivatalos statisztikai célból, az adott cél eléréséig ideiglenesen lehet helyreállítani. A cél teljesülése esetén, illetve ha a célhoz kötöttség nem áll fenn, az adatok teljességellenőrzését követően az azonosító adatokat le kell választani vagy törölni kell.

A fentiek mellett a statisztikai adatokhoz való hozzáférés feltételeit is részletesen meghatározza a statisztikai törvény VI. fejezete, amely egyúttal azt is biztosítja, hogy azokat kizárólag statisztikai célra használják fel, egyedi adat pedig – a törvényben foglalt kivételektől eltekintve – ne legyen hozzáférhető, megismerhető. Az adminisztratív adatokat, amikor statisztikai célra átveszik, illetve átalakítják, attól kezdve statisztikai célú adatállománnyá válnak, így ezen adatforrások egyedi adataira a statisztikai célú adatkezelés szabályai vonatkoznak. A statisztikai

törvény 39. § (8) bekezdése külön is rögzíti, hogy a statisztikai célra kezelt adatok büntető-, polgári peres, nem peres eljárás, hatósági, adóigazgatási eljárás során történő felhasználása, illetve belföldi jogsegély keretében való felhasználása tiltott.

Összefoglalva megállapítható, hogy a statisztikai törvény körültekintően rendelkezik a statisztikai adatok átadás-átvételéről.

A fentiekén túl kiegészítésképpen meg kell említeni, hogy az adminisztratív adatok egységesítése és a nyilvántartások közötti hatékony együttműködés nem csupán a statisztika területén bír kiemelt jelentőséggel, hanem az egyes állami nyilvántartások egymás közötti kapcsolata tekintetében is. Ennek az együttműködésnek a szabályozását célozza meg az elektronikus ügyintézés és a bizalmi szolgáltatások általános szabályairól szóló 2015. évi CCXXII. törvény. Ez a törvény kiszélesíti az intézményi környezetet, egy közös törvényben szabályozza az elektronikus ügyintézés, a kormányzati elektronikus szolgáltatások szabályait, valamint az intézmények közötti együttműködést, kapcsolattartást (interoperabilitás) is. Hatálya alá a KSH mint államigazgatási szerv tartozik, ennek alapján a hivatal által nyújtott elektronikus ügyintézési szolgáltatásra, valamint a más szervekkel való informatikai együttműködésre és elektronikus kapcsolattartásra alkalmazni kell a törvényt. Mindezek azonban nem sérthetik a statisztikai törvénynek az egyedi statisztikai adatokra vonatkozó adatvédelmi szabályait. A másodlagos adatok átvételénél ugyanakkor már figyelembe kell majd venni az abban foglaltakat.

2. Nemzetközi kitekintés

Az adminisztratív adatok statisztikai célra történő felhasználása viszonylag hosszú múltra tekint vissza az Európai Unió országaiban. Az elmúlt évekig ugyanakkor uniós szinten nem voltak törekvések a nemzeti gyakorlatok egységesítésére, a regiszterek használatát minden tagállam a saját nemzeti jogrendszerén belül szabályozta.

A statisztikai hivatalok számára rendelkezésre álló források csökkenésével, az adatszolgáltatói terhek mérsékléséhez, valamint a megnövekedett felhasználói igények kielégítéséhez fűződő érdekek közötti összhang megteremtésének szükségességével elkerülhetlenné vált a nemzeti gyakorlatok összehangolása, az alapvető szabályok uniós szinten történő rögzítése.

A közösségi statisztikákról szóló 322/97/EK rendelet külön cikkben rendelkezett a közigazgatási nyilvántartásokhoz való hozzáférésről. Kimondta, hogy a válaszadók terheinek csökkentése érdekében a nemzeti statisztikai hivataloknak és más nemzeti hatóságoknak és a Bizottságnak (Eurostat) – saját közigazgatási rendszerükön belül – olyan mértékben kell hozzáférniük az adminisztratív adatforrásokhoz, amilyen mértékben ezek az adatok az európai statisztikák fejlesztéséhez, előállításához és terjesztéséhez szükségesek. A hatékony hozzáférés megvalósítására vonatkozó gyakorlati szabályokat, valamint a hozzáférés feltételeit, ha szükséges, az egyes tagállamok és a Bizottság állapítják meg, saját megfelelő hatáskörükben.

Ezt a fenti szöveget vette át az európai statisztikákról szóló 223/2009/EK európai parlamenti és tanácsi rendelet is, amit a 2015/759/EK európai parlamenti és tanácsi rendelet 2015. április 29-én módosított, amely egyik legfontosabb módosító rendelkezése az adminisztratív nyilvántartásokhoz való statisztikai célú hozzáférést érintette.

A szupranacionális szabályozás hatására egyre több ország statisztikai törvényébe került be az adminisztratív adatok felhasználására vonatkozó rendelkezés. Az ilyen tartalmú szabályozás beépítése a legtöbb esetben összefüggött a nemzeti statisztikai hivatalok státuszának függetlenítésével, illetve koordinációs jogköreinek szélesítésével. A statisztikai törvények túlnyomó részében a nemzeti statisztikai hivatalok általános jellegű felhatalmazást kaptak a közigazgatási nyilvántartásokhoz való hozzáférésre, ugyanakkor a hozzáférés részletes tartalmát, az átadandó adatok körét nem határozta meg jogszabály. Előfordul olyan ország, ahol ezt a szabályozást miniszteri rendelet szintjén határozzák meg. Németországban külön törvény született az adminisztratív adatokhoz való hozzáférésről, ugyanakkor az nem minden adatforrásra terjed ki, csupán az adóhatóság és a foglalkoztatási hivatal adataira, rögzítve az átadandó adatok körét, az átadás gyakoriságát.

A nemzetközi tapasztalatok ugyanakkor azt mutatják, hogy a túl részletező szabályozás nem jelent feltétlenül célravezető megoldást, így az általános felhatalmazó rendelkezést ki kell egészíteni a hatósági nyilvántartások adatgazdáival kötött kétoldalú megállapodásokkal.

A felhatalmazó rendelkezés megfogalmazása kapcsán tanulságos, hogy azokban az országokban, ahol a hozzáférést mint a nemzeti statisztikai hivatal jogosultságát fogalmazták meg, később problémát jelentett, hogy ezt a jogosultságot az adminisztratív adatgazdák nem értelmezték a saját oldalukon megjelenő kötelezettségként, így az együttműködés az intézmények között sok esetben akadozott.

A szervezetek közötti együttműködésnek minden, az adminisztratív adatok felhasználásában élen járó ország (Finnország, Kanada) rendkívüli jelentőséget tulajdonít.

Ezek az országok az együttműködés különböző szintjeit alakították ki, amelyek egymásra épülve biztosítják az adatokhoz való hozzáférés akadálymentességét. Erre a rendszerre jó példa Finnország, ahol a statisztikai hivatalon belül minden nyilvántartáshoz kijelöltek egy felelős személyt, aki az adatgazda hatóság által erre a célra delegált munkatárásával folyamatosan tartja a kapcsolatot. A felsővezetői egyeztetés évente egy speciálisan erre a célra összehívott fórumon zajlik, de a regisztereket gondozó hatóságok egymás közötti együttműködését segíti egy önálló munkacsoport is.

A kétoldalú együttműködési megállapodásoknak tehát kiemelt szerepe van, általánosságban megállapítható, hogy az alábbi területeket szabályozzák:

- jogi háttér,
- kapcsolattartók kijelölése,
- az átadandó adatok részletes leírása,
- átadás gyakorisága,
- minőségi kritériumok,
- adatvédelmi rendelkezések,
- változáskezelés,
- költségek rendezése.

A fenti területek közül talán a legproblémásabb minden állam számára a költségek rendezésének kérdése. A nemzeti statisztikai hivatalok folyamatos költségvetési megszorításoknak vannak kitéve, így nehézséget okozhat számukra, hogy nagyobb összegeket fizessenek ki az adatokért. Emellett kérdéses annak jogossága, hogy két államigazgatáson belüli szerv egy közfeladat ellátását kifizetés teljesítésétől tegye függővé. Ezt az ellentmondást sok

statisztikai hivatal úgy próbálta megoldani, hogy az adatért cserébe felajánlották munkatársaik szakértelmét, vagy más szervezetet vettek igénybe az adatkéréshez.

A fentiek alapján tehát megállapítható, hogy ugyan már létezik kialakult gyakorlat az uniós tagállamaiban az adminisztratív adatok felhasználására, azonban ez tagállamonként nagyon eltérő, sőt az is előfordul, hogy adott tagállamon belül, a különböző szervezetek vonatkozásában is más gyakorlat alakult ki. A fellépő problémák ugyanakkor általánosak:

- nem kellően erős a jogi felhatalmazás;
- nem biztosított az adatokhoz való térítésmentes hozzáférés;
- nem biztosított a statisztikai hivatalok bevonása az adminisztratív nyilvántartások tartalmának kialakításába, annak megváltoztatásába.

Tekintettel arra, hogy ezek a problémák akár az uniós adatszolgáltatási kötelezettségek teljesítését is veszélyeztethetik, részletesebb európai uniós szintű felhatalmazás és szabályozás szükséges.

A módosított 223/2009/EK európai parlamenti és tanácsi rendelet egy külön cikkben (17a.) kimondja, a válaszadók terheinek csökkentése érdekében nemzeti statisztikai hivatalok, valamint más nemzeti hatóságok és a Bizottság (Eurostat) jogosultak a közigazgatási nyilvántartásokhoz való azonnali és ingyenes hozzáférésre és azok azonnal és ingyenesen történő felhasználására, valamint e közigazgatási nyilvántartások adatainak a statisztikákba történő beépítésére olyan mértékben, amennyire ezek – az európai statisztikai programban meghatározott – európai statisztikák fejlesztéséhez, előállításához és közzétételéhez szükségesek.

Mindemellett a rendelet új szakasza alapján kötelező lesz a statisztikai hivatalokat megfelelő időben bevonni az új adminisztratív adatforrások kialakításába vagy a már létező nyilvántartások módosításába, megszüntetésébe vagy adatkörük csökkentésébe. Az adminisztratív adatok kezelői szintén kötelesek lesznek a statisztikai hivatalok számára megfelelő metaadatokat szolgáltatni, valamint elvégezni azokat a standardizációs feladatokat, amelyek alkalmassá teszik az adatállományokat a statisztikai célú felhasználásra. Mindezek megvalósítása érdekében a rendelet együttműködési mechanizmusok kidolgozására köteleznék a tagállamok intézményeit és az adatgazdákat.

Az európai statisztikákról szóló rendelet mellett az Európai Statisztika Gyakorlati Kódexe is számos új, az adminisztratív adatok felhasználására vonatkozó indikátorral bővült a 2011. szeptember 28-án kiadott módosításában.

A kódex 2. elvének (Felhatalmazás adatgyűjtésre) 2.2. ismérve mondja ki, hogy a nemzeti statisztikai hivatalokat jogszabályban kell feljogosítani az adminisztratív adatokhoz való hozzáférésre.

Mivel a jogszabályi felhatalmazás – ahogy azt fent már jeleztük – nem elégséges, ezért a kódex a 8. elven (Megfelelő statisztikai eljárások) belül a 8.8 ismérven keresztül kötelezi a statisztikai hivatalokat és a nyilvántartásokat kezelő hatóságokat, hogy az adminisztratív adatok statisztikai célú felhasználása iránti elkötelezettségüket külön megállapodásban rögzítsék.

A fentiekben túlmenően a 8. elven belül szereplő kilenc indikátorból még három vonatkozik az adminisztratív adatok felhasználására. Ezek rögzítik, hogy az adminisztratív nyilvántartások fogalmainak a lehető legjobban meg kell felelni a statisztikai céloknak, a statisztikai hivatalokat pedig be kell vonni a nyilvántartások kialakításába. Emellett a szervek

fentiek szerint formalizált együttműködésének ki kell terjednie az adminisztratív adatok minőségének biztosítására is.

A 9. elv (Adatszolgáltatói terhek csökkentése) megerősíti, hogy minden lehetséges esetben adminisztratív adatokat kell felhasználni a statisztikában annak érdekében, hogy az adatszolgáltatás megkettőzése elkerülhető legyen.

Mindezekre tekintettel megállapítható, hogy a nemzeti statisztikai hivatalok közigazgatáson belül betöltött szerepének erősítésével összhangban szükséges az adminisztratív adatokhoz való hozzáférési jogosultságok kiterjesztése, és egységes kezelése uniós szinten is. Erre alapozva könnyebben megtörténhet a nemzeti jogszabályok módosítása is, hiszen az európai statisztikákról szóló rendeletben foglaltak tagállami végrehajtása kötelező. Figyelembe kell venni ugyanakkor, hogy a jog és annak eszközei csupán a hátteret, alapot szolgáltatják, és valódi együttműködés, szemléletváltás nélkül nem valósítható meg a cél.

V. Adminisztratív adatok felhasználása a népszámlálásokban

1. A népszámlálás fogalma és megvalósítása

A népszámlálás egy ország lakosságának létszámára és jellemzőire vonatkozó adatgyűjtések fő pillére, mely nemzeti és helyi szinten rendszeres időközönként biztosítja a lakosság számának megállapításához a kiindulási alapot. A kis földrajzi területek vagy részpopulációk egyes szociális, demográfiai és gazdasági jellemzői tekintetében a népszámlálás lehet az egyetlen információforrás. Sok országban a népszámlálás szilárd keretet biztosít a statisztikai adatgyűjtések mintavételi kereteinek kidolgozásához is.²⁴ A népszámlálásokat a statisztikai hivatalok bonyolítják le.

A népszámlálás azt jelenti, hogy minden egyént és lakást egyedileg veszünk számba, és rögzítjük a jellemzőit. A statisztikai egységek csak ezzel az eljárással osztályozhatók teljeskörűen a különböző jellemzők szerint. A népszámlálásnak az alábbi öt alapvető jellemzője van:

1. Egyedi számbavétel: minden egyes személyt és lakást külön-külön össze kell írni.
2. Egyidejűség: a számbavételnek egy adott időpontra kell vonatkoznia, azaz egy pillanatfelvétel készül.
3. Egyetemesség: az adott területen, általában az adott ország területén teljeskörűen minden személyt és lakást össze kell írni.
4. Kisterületi adatok: a teljes körű összeírás alapján a népszámlálás kisebb területi egységekre is tudjon információval szolgálni.
5. Definiált periodicitás: a népszámlálásokat rendszeres időközönként, általában 10 évente kell megtartani, hogy ezzel longitudináisan összehasonlítható adatokat kapjunk.

A népszámlálásokat hagyományosan, történeti kialakulásuk óta területi összeírással végzik. A világ országainak döntő többségében ma is ilyen módon gyűjtik a census adatait. A népszámlálások végrehajtásának a modern világban, a technika, technológia fejlődésével és az

²⁴ Principles and Recommendations for Population and Housing Censuses. Revision 2. Department of Economic and Social Affairs Statistics Division. United Nations. New York, 2008. <http://unstats.un.org/unsd/censuskb20/Attachment455.aspx>

anyagi lehetőségek szűkülésével párhuzamosan már számos más módja is kialakult, és egyre bővül a különféle megvalósítások változatainak, típusainak száma.

Az egyedi számbavételre vonatkozó követelmény teljesülése elérhető területi információgyűjtéssel, azaz ún. összeírással, de olyan egy vagy több adminisztratív nyilvántartás használatával is, amelyben a szükséges információk fellelhetők, sőt ezeknek a módszereknek a kombinációja is lehetséges.

A három alpmódszer közé tartozik a hagyományos népszámlálás, amikor minden személyt és lakást területi összeírással veszünk számba; az adminisztratív adatforrásokra épülő népszámlálás, amikor nincs területi összeírás; illetve a kombinált népszámlálás, amikor adminisztratív adatforrásokat és teljes körű vagy mintavételes területi összeírást kombinálunk, és így nyerjük ki az adatokat. E fő módszereknek különféle altípusai is vannak, hiszen minden ország a saját legjobb tudása, eszköze és módja szerint bonyolítja népszámlálását. Fontos kiemelni azonban, hogy a módszerek szabad megválasztása mellett minden cenzusnak meg kell felelnie a népszámlálás alapvető jellemzőinek.

1.1. Adminisztratív adatforrásokon alapuló népszámlálás²⁵

Az adminisztratív adatforrásokon alapuló népszámlálás egyedi abban a vonatkozásban, hogy alkalmazásával a lakosságot nem keresik fel közvetlenül, hanem az egyes közcélú adminisztratív nyilvántartásokban, mint például a népesség-nyilvántartásban, az ingatlan-nyilvántartásban, a társadalombiztosítási nyilvántartásban, a különböző munkaügyi nyilvántartásokban (adóhivatali adatok stb.) található információkat kapcsolják össze és nyerik ki a személyekre és lakásokra vonatkozó adatokat.

Az a koncepció, hogy a statisztika népszámlálás típusú adatokat adminisztratív adatforrások felhasználásával állítson elő, nemzetközi szinten a 2000 körüli népszámlálásokkal összefüggésben merült fel, bár már 1970 óta vizsgálták ennek lehetőségét, és bizonyos tesztek el is végeztek. Finnországban 1981-ben, Dániában pedig 1991-ben már kizárólag adminisztratív adatokat felhasználva sikerült előállítani a népszámlálási adatokat. A koncepció mögött meghúzódó filozófia az, hogy kihasználják a rendelkezésre álló közigazgatási forrásokat, elsősorban azoknak a nyilvántartásoknak az integrálásával, amelyeknek fókuszában a háztartások, a lakások és a magánszemélyek állnak. Ezekhez az adatforrásokhoz egyedi szinten kapcsolódnak az egyének egyéb, például adózási, oktatási, foglalkoztatási információit tartalmazó nyilvántartásai. Bár az adatállományok összekapcsolása elméletileg lehetséges a magánszemélyek neve és egy másik alapadat (pl. születési ideje és helye vagy lakcíme) alapján, mégis hasznosabb, ha minden személynek, háztartásnak és lakásnak egyedi azonosító kódja van, amivel a különböző nyilvántartások hatékonyabb és megbízhatóbb összekapcsolása válik lehetővé.

Ennek a megközelítésnek a legfontosabb feltétele, hogy az országnak legyen egy kiváló minőségű és jó lefedettségű központi népesség-nyilvántartása, amihez kapcsolódik egy folyamatos adatfrissítést biztosító technológia is.

²⁵ A népszámlálások nemzetközi szakirodalmában regiszteresnek nevezett népszámlálást jelen kiadványban adminisztratív adatforrásokon vagy nyilvántartásokon alapuló népszámlálásnak nevezzük, mivel a statisztikai fogalmak hazai rendszerében a „regiszter” bizonyos feltételeknek megfelelő nyilvántartás, és a népszámlálásokhoz felhasználható nem statisztikai adatforrások nem tesznek eleget ezeknek az elvárásoknak (lásd II. fejezet).

Az adminisztratív adatforrás alapú népszámlálás elsődleges előnyei a hagyományos módszerhez képest – hasonlóan más, adminisztratív adatokon alapuló statisztikákhoz –, hogy a lebonyolítás költsége lényegesen alacsonyabb, és az adatok előállítása lehet gyakoribb is. E módszer nem terheli a lakosságot sem. Az adminisztratív nyilvántartások létrehozása és karbantartása azonban lényegesen költségesebb, mint amit egy népszámlálás önmagában indokolna, ezért ez jobbra csak akkor lehetséges, ha a statisztika mellett van olyan hasznos és hatékony közigazgatási cél, amihez a nyilvántartás hasznosítható.

A népszámlálások szempontjából az adminisztratív adatforrásokon alapuló statisztikák legnagyobb hátránya az a tény, hogy kizárólag azok az információk használhatók fel, amik a nyilvántartásokban rendelkezésre állnak. Az adminisztratív adatokból nem nyerhetők ki a háztartási és családi kapcsolatok, a családi állás, az élettársi kapcsolat, illetve a szenzitív kérdésekre (nemzetiséghez tartozás, vallás, egészségi állapot) várt válaszok sem. Ezen adatok egy része más információk alapján vélelmezhető, de így is maradnak olyan paraméterek, amik nem találhatók meg a nyilvántartásokban. További jelentős probléma ezekkel az adatokkal, hogy nem mindig a valóságot tükrözik (pl. a bejelentett és a tényleges lakóhely különbözik), és az eltérés mértéke adatforrásonként és országonként különbözik. Ezek a hátrányok korlátozhatják az időbeli és térbeli összehasonlíthatóságot.

1.2. Hagományos és adminisztratív adatforrásokon alapuló eljárás kombinálása

Azokban az országokban, ahol a különböző nyilvántartások nem tekinthetők egyformán fejlettnak vagy megbízhatónak ahhoz, hogy minden szükséges népszámlálási eredményt azokból nyerjenek ki, az adminisztratív adatokat a teljes körű összeírás alapjául lehet felhasználni. Egyes országokban (pl. Spanyolország) a 2000 körül végzett népszámlálások kérdőíveit részlegesen előre kitöltötték a rendelkezésre álló igazgatási adatok alapján. Beírták például a népesség-nyilvántartásból a lakás címét, a személyek nevét, valamint egyéb rögzített információkat. Ezeket az adatokat azután az interjú során megerősítették vagy korrigálták.²⁶

Ezen a módon az átlagos interjúidő lényegesen csökken, amely csökkenti az adatfelvétel költségeit. Emellett a válaszadói terheket is alacsony szinten tartjuk, ami javítja az adatok minőségét. Emellett összehasonlítva a népesség-nyilvántartásból és a terepmunka során gyűjtött információkat, értékelni lehet az igazgatási adatok pontosságát a lefedettség vonatkozásában.

Ha bizonyos változóknak a két adatforrásban fellelhető értékeit is összehasonlítjuk, értékelhetjük a nem statisztikai felhasználás céljával létrejött adatállomány minőségét, és ha az eredmény kedvező, lehetővé válhat, hogy bizonyos népszámlálási mutatókat gyakrabban publikáljunk kizárólag nyilvántartási adatok alapján. Ha rendszeresen végzünk ilyen típusú népszámlálást, azt úgy lehet tekinteni, mint egyfajta leltározást, amely lehetővé teszi a nyilvántartások alapján adódó népességszám korrekcióját. Erre azért lehet szükség, mert vannak olyan – például vándorlással összefüggő – események, amelyek nem lettek a nyilvántartásban regisztrálva.

A kombinált megközelítésnek is a teljes körű lefedettség az alapja, ezért itt is teljesül a hagyományos megközelítésnek az az előnye, hogy a teljes népességre nézve, egy, az adott

²⁶ Population and housing censuses. United Nations Statistics Division.

<http://unstats.un.org/unsd/demographic/sources/census/alternativeCensusDesigns.htm>

időszakra és viszonylag kis közigazgatási területekre vonatkozó pillanatfelvételt közvetít. Azok az országok, ahol ilyen típusú népszámlálást bonyolítottak le, azt tapasztalták, hogy a hagyományos gyakorlat hátrányai bizonyos mértékben csökkennek.

1.3. Adminisztratív adatforrásokon alapuló népszámlálás kombinálása mintavételes adatfelvételekkel

Egyes országokban (pl. Lengyelország, Németország) – ahol a nyilvántartások nem tartalmazzák az összes szükséges információt, mert például nincs háztartásra vonatkozó információ – az igazgatási adatokat mintavételes felmérésekkel egészítik ki. A mintavétellel csökkentik a területi összeírás költségeit. Ennek az eljárásnak a keretében először a nyilvántartások rekordjait elemi szinten összekapcsolják a mintavételes felmérések adataival. A következő lépésben az eredményektől függően az adatokat kitáblázzák vagy teljeskörűsítik. Ez a megközelítés hasonlít a népszámlálás rövid kérdőív/hosszú kérdőív paradigmájára, azaz, hogy amikor az alapvető kérdéseket minden egyének, illetve háztartásnak felteszik, míg az egyének és háztartások egy mintájával részletesebb kérdőívet töltenek ki, végül az utóbbi eredményeket teljeskörűsítik.

Ennek a megközelítésnek a fő tartalmi és módszertani előnyei és hátrányai hasonlóak a tisztán adminisztratív adatokon alapuló eljárásoknál felmerülőkhöz képest, amelyek esetében lényegében nincs visszautasítás, és a válaszmegtagadás csak a mintavételes adatgyűjtésekkel összefüggésben jelenthet problémát. A módszer hátránya a kizárólag adminisztratív adatforrásokon alapuló cenzushoz viszonyítva, hogy súlyozási probléma esetén a táblák előállítására a mikroadatokból több munkát igényel, illetve problematikus lehet bizonyos változók előállítására kisterületi adatokra, településekre, és ezáltal a népszámlálási idősorok folytatása is.

Ez a módszer közepes költségű a hagyományoshoz és a kizárólag adminisztratív adatforrásokon alapuló népszámláláshoz képest. Megvalósításának feltétele az adminisztratív adatállományok jó minősége mind az adattartalom, mind a technika tekintetében, valamint a nagyon pontos módszertani háttérmunka és tervezés a minta kialakításánál, illetve a végleges adatok előállításánál. További előnye, hogy a lakosság együttműködését jóval kisebb mértékben teszi szükségessé, így csökken az összeírás emberi erőforrás igénye.

A mintás felvétel pozitívuma a teljes körű területi összeíráshoz képest, hogy nem szükséges nagyszámú összeírót alkalmazni, így lehetővé válik, hogy a kevesebb összeíró pontosabb, megbízhatóbb munkát tudjon végezni, hiszen a legjobban teljesítő munkatársakat lehet kiválasztani. Ez javíthatja a minőséget.

Statisztikai területi összeírás során költségcsökkentő tényező a mobil eszközök használata, vagy az önköltés lehetősége. Ezek a módszerek a területi összeíráshoz kapcsolódnak, azaz még a hagyományos népszámlálás részeinek tekinthetők.

2. Nemzetközi tapasztalatok, 2011

2011-ben a 27 uniós országból 5 (Ausztria, Dánia, Finnország, Svédország, Szlovénia) kizárólag adminisztratív adatforrásokon alapuló népszámlálást tartott, 2 (Belgium és Hollandia) az adminisztratív adatforrások mellett meglévő mintás felvételeket is felhasznált, 6 ország (Észtország, Lengyelország, Lettország, Litvánia, Németország, Spanyolország) pedig adminisztratív adatforrásokat és területi adatgyűjtést, azaz kombinált népszámlálást hajtott

vége. Ez azt jelenti, hogy 2011-ben az EU tagországainak majdnem fele felhasználta az adminisztratív adatokban rejlő lehetőségeket, míg az előző ciklusban még ritka volt ez a megvalósítási mód (7 ország). Az átállás egy újabb módszerre általában fokozatosan történik. Eddig Ausztria az egyetlen olyan ország, amely radikális változtatással egy lépésben tért át a hagyományos összeírásról az adminisztratív adatforrások kizárólagos alkalmazására.

A különböző módszerek költségvetési szempontú összehasonlítására a 2011. évi népszámlálások adnak lehetőséget (1. tábla).

1. tábla

Egy főre jutó népszámlálási költségek néhány országban típusonként, 2011 körül

Népszámlálás típusa	Költség egy főre vetítve USA-dollárban					
	minimum		közép		maximum	
Hagyományos népszámlálás	1,03	Azerbajdzsán	7,60	Magyarország	18,21	Írország
	2,93	Románia	8,45	Görögország	19,09	Kanada
				13,85	Olaszország	40,17
Kombinált népszámlálás	0,19	Törökország	3,94	Litvánia	12,76	Németország
	2,56	Spanyolország	5,18	Lettország	13,98	Észtország
Adminisztratív adatforrásokon alapuló népszámlálás	0,03	Szlovénia	0,24	Finnország	0,71	Svédország
	0,07	Dánia	0,50	Norvégia	1,64	Ausztria

Forrás: Measuring population and housing, Practices of UNECE countries in the 2010 round of censuses, United Nations New York and Geneva, 2014.

Az adminisztratív adatforrásokon alapuló népszámlálások költsége a legalacsonyabb, közepes a kombinált módszeré és a hagyományos a legdrágább. Az eltérés egy-egy típuson belül számos tényezőtől függ, például az ország méretétől, gazdaságától, fejlettségétől, lakosságától stb. A költségek szóródása egy-egy típuson belül az alkalmazott technológia függvénye is. Ausztriában, ahol egy lépéssel tértek át a hagyományos népszámlálásról az új rendszerre, a legmagasabb az egy főre jutó költség azok között az országok között, ahol csak adminisztratív adatokat használnak. Ennek oka, hogy maga az átállás sok előzetes költséggel, a nyilvántartások létrehozásával, technikai fejlesztésével járt.

Népszámlálási típusonként nemcsak az összköltségek térnek el, hanem a költségek feladattípusok szerinti megoszlása is jelentősen különbözik.

*A népszámlálás költségeinek megoszlása típus és feladatok szerint
a 2011-es népszámlálások eredményei tükrében*

Feladat	A feladat részesezése		
	hagyományos	kombinált	adminisztratív adatforrásokon alapuló
	népszámlálás esetén		
Próbászámolás	3	1	0
Online kérdőív létrehozása	2	5	0
Nyomtatás	6	2	0
Térképes és egyéb geográfiai támogatás	6	3	0
Publicitás, kommunikáció	3	3	0
Területi összeírás	54	53	0
Adat-előállítás, ellenőrzés és kódolás	13	13	56
Projektmenedzsment és/vagy adminisztratív támogatás	4	13	11
Közzététel, publikáció és dokumentáció	4	4	16
Egyéb	7	4	18

Forrás: Measuring population and housing, Practices of UNECE countries in the 2010 round of censuses, United Nations New York and Geneva, 2014.

A legnagyobb eltérés a területi összeírás és az adat-előállítás költségében van. Míg a hagyományos népszámlálásnál a legjelentősebb költségvetési tételt a területi munkák teszik ki, addig a nyilvántartásokon alapuló népszámlálásnál a legnagyobb arányt az adat-előállítás, ellenőrzés és a kódolás. A kombinált módszernél a költségeloszlás hasonlít a hagyományoshoz, de a népszámlálás összköltsége alacsonyabb.

3. Hazai előzmények

A KSH 2007-ben kezdte meg az újfajta módszertan szerint az adminisztratív alapú népszámlálás megvalósíthatóságának feltérképezését, többek között a „Felkészülés a 2011-es népszámlálásra” című európai uniós projekt keretében. Ennek során első lépésként információt gyűjtött a hivatal arról, hogy milyen adminisztratív adatforrások vannak az országban, ezek milyen jogszabály alapján működnek, és milyen népszámlálási jellegű információkat tartalmaznak. A KSH munkatársai a hivatalos nyilvántartások feltérképezése során 9 adatgazda 19 adatgyűjtését találták valamilyen módon népszámlálási célból hasznosíthatónak. (A nyilvántartások száma és adatgazdája a 2007-es állapotokat tükrözi. A jelen fejezet 7. pontja tartalmazza a 2016-ban népszámlálási szempontból hasznosítható nyilvántartások listáját és rövid leírását.) Azon három adatforrásnál, mely a fő népszámlálási alapadatokat tartalmazza, közvetlen kapcsolatfelvétel is történt az adatforrások adatgazdáival (Földhivatal, Építésügyi hatóság, Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala (KEKKH)). A KSH az adatgazdákkal közösen vizsgálta az adatátadás-, átvétel jogszabályi hátterét, vagyis az adatok elérhetőségét. A projekt és a népszámlálási előkészítő munka fontos feladata volt a nemzetközi tapasztalatok összegyűjtése is. A vizsgálat a következő megállapításokra jutott.

A tanulmányozott 19 nyilvántartás közös jellemzője volt, hogy népszámlálási szempontból nem teljes körűek, nyilvántartási módjuk különböző, olykor egy rendszeren belül országosan sem egységesek. Az adminisztratív adatforrásokon alapuló népszámlálásoknak korábban említett gyengeségei ezekkel kapcsolatban is lényegében mind megjelentek.

A hazai jogszabályi környezet nagyrészt megfelelő egy adminisztratív adatokon alapuló népszámláláshoz, de szükség lenne még egyes ágazati törvények módosítására. A rendkívül sokféle módon nyilvántartott adatállományok más-más átvételi módjának kidolgozása, az adatok összekapcsolása nagy erőforrást igényel. A technikai nehézségek mellett az adatok nem megfelelő minősége szintén probléma, hiszen az átvett adatok nagy része népszámlálási szempontból nem teljes körű, a bejegyzések gyakran nem pontosak. Az állományok ellenőrzése és javítása is rendkívül összetett, nagy munkaerőigényű és költséges feladat. Az adminisztratív adatforrásokon alapuló népszámlálásra történő átállással járó módszertani és technikai fejlesztések az elején nagy anyagi ráfordítással járnak, s e ráfordítások nem kizárólag a statisztikai hivatal oldalán jelentkeznek, hanem az egyes adatgazdáknál is. Átállás esetén a felkészülési, előkészületi időszak is jóval hosszabb, mint egy hagyományos népszámlálás előkészítésénél. Azokban az országokban, ahol a hagyományos összeírásról teljes körű adminisztratív adatokon alapuló népszámlálásra térnek át, általában 10 évet vesz igénybe a teljes átállás, azaz rögtön egy népszámlálás lebonyolítása után belekezdnek az átállás megvalósításába.

Az adatállományok kezelésének technikai nehézségeinek feloldása előtt nagy jelentőségű feladat annak meghatározása, hogy a szokásos tematika hányad része nyerhető, illetve nem nyerhető ki a rendelkezésre álló nem statisztikai adatforrásokból. Speciális magyarországi jelenség a kettős bejelentettség, vagyis, hogy nemcsak lakóhelye, hanem tartózkodási helye is lehet egy Magyarországon élő személynek, és ez további bizonytalanságot okozhat a származtatott népszámlálási mutatók megbízhatóságát, minőségét illetően. E mellett másik hangsúlyos szempont, hogy a bejelentettség nem minden esetben fedi le az emberek tényleges lakó-, illetve tartózkodási helyét. Az áttérés a tényleges lakóhelyről a bejelentett lakóhelyre nagy módszertani kihívás és változás is egyben.

A 2011. évi népszámlálás előkészítésekor az igazgatási adatok vizsgálatának eredményeképpen az az összegző megállapítás született, hogy a fennálló problémák nem teszik lehetővé a kizárólag adminisztratív adatforrásokon alapuló népszámlálás megvalósítását.

Ennek ellenére a szóba jövő adatforrások között a KEKKH kezelésében lévő népesség-nyilvántartás adatait a 2011. évi népszámlálás többértően használta fel:

- Mint egy lehetséges hazai, adminisztratív adatokon alapuló népszámlálás alapnyilvántartását a KSH a többenél részletesebben vizsgálta. A népszámlálási próbafelvételek lebonyolításához
 - a 2005. április 1-jei állapot szerinti népesség-nyilvántartás részletes adatai a mikrocenzus adataival lettek összehasonlítva,
 - a 2008-as próbaszámlálás adatainak és a KEKKH azonos adatainak összehasonlítása is megtörtént.
- A népszámlálás előkészítése során a KSH címregiszterében szereplő címek ellenőrzéséhez, pontosításához szükség volt a KEKKH nyilvántartásban szereplő címadatakra.

- Hiányzó adatok pótlása.
 - A népszámlálások során első alkalommal történt meg, hogy a személyek meghíúsult összeírása esetén az adminisztratív adatforrást használták az alapvető demográfiai adatok pótlásához.
 - Amikor egy teljes cím összeírása hiúsult meg, az egység szintű imputáláshoz a jónak minősült adathalmazból választották ki a hiányzó címhez leginkább hasonlót, és azzal pótolták. A donor kiválasztása az adminisztratív adatok alapján történt.
- Mikro- és makroszintű validálás, vagyis az adatok érvényesítése egyedi és aggregált szinten.

3.1. A mikrocenzus és a próbanépszámlálás összehasonlítása a lakcímnilyvántartással

3.1.1. A 2005. évi mikrocenzus adatai

A népesség-nyilyvántartás adatállományából a személyi adatok (nem, születési idő (év, hónap), családi állapot, bejelentkezés módja) és a személyekhez kapcsolódó címadatok (település, közterület, házszám, épület, lépcsőház, emelet, ajtó) kerültek felhasználásra. A mikrocenzus címei közül csak azok lettek figyelembe véve, ahol a felvétel sikeresen lezajlott. Az összekapcsolás egysége a cím volt. Először településenként az azonos címekhez tartozókat gyűjtöttük össze, majd a személyek összekapcsolására került sor, első körben informatikai párosítással, majd manuálisan.

Az összehasonlítás azt az eredményt hozta, hogy a mikrocenzus címeinek 76%-a megtalálható volt a népesség-nyilyvántartásban is, de azoknak a címeknek az aránya, ahol valamennyi ott lakó személy is azonos volt, csak 44%-ot tett ki.

3. tábla

A 2005. évi mikrocenzus és a népesség-nyilyvántartás összehasonlításának eredménye

Egyezőségi szint	Egyező címek	
	száma, darab	aránya, %
	a mikrocenzusban	
Település	71 095	100
Közterület	70 701	99
Házszám	63 889	90
Teljes cím (lakás)	54 284	76
Legalább egy személy azonos a lakcímen	47946	67
Valamennyi személy egyező a lakcímen	31 539	44

Következtetések:

- A két állományban az eltérések kimutathatóak voltak, az azonban nem, hogy melyik a helyes vagy valós információ, például egy cím esetében.
- A népesség-nyilyvántartás adataiban az üres és a nem lakásként használt lakások nem szerepelnek.
- A címek manuális összekapcsolásához szükséges lenne a teljes lakókörnyezet címadataira, vagyis, hogy tudjuk melyik a jó, azaz pontos cím.

3.1.2. A 2008. évi próbanépszámlálás adatai

Az egyező címek aránya településenként eltérő volt, a legkisebb arányban a budapesti címek egyeztek meg szemben a kisebb települések nagyobb egyezőségével.

Következtetések:

- A KEKKH lakcímei a vizsgált településeken 62%-ban feleltek meg a tényleges helyzetnek.
- A nyilvántartási rendszer korszerűsítésének hatása még nem volt érzékelhető, továbbra is hiányoztak az üres lakások.
- A KEKKH címei népszámlálási célra csak teljes körű, településenkénti címpontosítás után használhatóak.
- A népesség számában az adatok településenként 1–12%-os eltérést mutattak, a népesség összetételében azonban nem voltak jelentősek az eltérések.
- Elemi szinten a nyilvántartás adatai a korábbi vizsgálatához hasonlóan csak részben fedték a próbanépszámlálás szerinti helyzetet a fiktív bejelentkezések és a bejelentkezés nélkül élők miatt.
- A nyilvántartásban szereplő demográfiai alapadatok ugyanakkor pontosabbak voltak, mint az összeírt adatok, hiszen dokumentumokon alapulnak.

4. Adminisztratív adatok felhasználásának feltételei egy hazai népszámlálás során

Áttekintve a rendelkezésre álló adminisztratív és statisztikai adatforrások tartalmát a KSH-nak a 2020. évi népszámlálás előkészítése során az adminisztratív adatforrások felhasználásának lehetőségét vizsgálva a minőségi kérdéseken túl tekintetbe kell vennie a szükséges jogi, technikai és erőforrás feltételek teljesülését is. Az, hogy a kínáló két lehetőség közül melyiket észszerű végrehajtani, a meglévő nem statisztikai adatforrások minősége mellett attól is függ, hogy a szükséges jogi, technikai és erőforrás feltételek hogyan teljesülnek.

4.1. Jogszabályi és technikai feltételek

A népszámlálás szempontjából az alapvető jogszabályi és technikai feltételek azonosak az adminisztratív adatok statisztikai célú felhasználásával összefüggésben már ismertettekkel (I. és IV. fejezet). A jogszabályi környezet a 2011 előttihez képest számos ponton kedvezőbbé vált, valamint folyamatban vannak olyan jogi változások, amelyek hosszabb távon több nyilvántartás felhasználhatóságát is lehetővé teszik.

Mint ahogy az egységes címhasználat a népszámlálás szempontjából kiemelten fontos kérdés, meg kell itt említeni a kormány által az állami és önkormányzati nyilvántartások együttműködésének általános szabályairól szóló 2013. évi CCXX. törvény (Interoptv.) felhatalmazásával hozott 345/2014. (XII. 23.) rendeletet a központi címregiszterről és a címkezelésről. A jogszabály alapján 2015 júliusában létrejött a központi, egységes és teljes címregiszter, mint közhiteles adatbázis, melynek célja a magyarországi címkezelés egységes jogszabályi környezetének és igazgatási folyamatainak kialakítása. A rendelet felállította a címek technikailag egységes szabványát, és a nyilvántartás felelőseként a KEKKH-t nevezi meg.

A jogszabályi környezet a KSH számára lehetővé teszi több adatforrás egyidejű felhasználása esetén a személyes adatok statisztikai célú összekapcsolását. Az összekapcsolás

ebben a körben vagy a személyi azonosítókból képzett technikai azonosítóval vagy lakcím alapján történhet.

4.2. Meglévő nyilvántartások felhasználhatósága

A különböző adminisztratív nyilvántartások számos népszámlálási információt tartalmaznak, de ha a közvetlen adatgyűjtés helyett ezek szolgálnának adatforrásként, az alábbi általános problémák megoldatlanul maradnának:

- a bejelentés nélkül élőkről nincs adat,
- a bejelentés nélkül dolgozókról nincs adat,
- az iskolai végzettségről nincs teljes körű adat,
- a családok, háztartások szerkezetének alapinformációi hiányoznak, nincs információ az élettársi kapcsolatokról,
- a szenzitív, azaz az anyanyelvre, nemzetiségre, vallásra és a fogyatékkal élőkre vonatkozó adatok csak részlegesen vagy egyáltalán nem állnak rendelkezésre,
- a lakásállomány egészéről sincs olyan elérhető nyilvántartás, ami megfelelő minőségű, átfogó és teljes körű adatokat biztosítana.

Az adminisztratív adatforrásokból származó személyes adatok összekapcsolása esetén kockázatot jelent, hogy az így nyert információk torzítanak a népesség számát és területi megoszlását, és a torzítás mértéke ismeretlen maradna. Ennek fő oka, hogy a lakcímbejelentések hiányosak, ugyanis nem tükrözik a lakosságnak az életvitelszerű tartózkodási hely szerinti valós megoszlását. Az eltérés mértéke településtípusok között változó, és nemcsak a lakosságzámban jelenhet meg, hanem érinthet egyéb területi szintű indikátorokat is, mint az iskolai végzettségre, családi viszonyokra, munkahelyre vonatkozó mutatók, de még a kisebbségi vagy a vallási adatok esetén is torzíthatnak a nyilvántartási adatok. Ez különösen egyes népességcsoportok esetén jelenthet problémát, mint például a határ mentén dolgozóknál, a külföldön bejelentett lakcímmel rendelkezőknél, vagy az olcsóbb parkolás vagy egyéb előnyök miatt egy lakóhelyen csak bejelentkezett, de ténylegesen máshol élő embereknel. A családok szerkezetét is befolyásolja, amennyiben a család tagjai a tényleges helyzettől eltérően különböző címeken vannak bejelentkezve. Az adminisztratív adatforrásokon alapuló népszámlálás megvalósíthatóságának jelentős akadálya, hogy a népszámlálási lakásadatokról ma Magyarországon nincs teljes körű nyilvántartás, így azokat statisztikai forrásokból kellene előállítani.

A korábban említett 345/2014-es kormányrendelet ugyan jelentős előrelépés a címek egységesítésének és egyszerűsítésének tekintetében, de népszámlálási szempontból, e rendelet hatályba lépése után is maradnak problematikus pontok. A szabvány felállításával nem történt meg a címek teljes körű területi bejárása, ezért továbbra is lehetnek olyan címek, amelyek nem szerepelnek, vagy rosszul szerepelnek a nyilvántartásokban. A lakott egyéb lakóegységek, kunyhók, lakókocsik, sufnik stb., ahol ténylegesen élnek emberek, nem szerepelnek a központi címregiszterben, pedig ezek is beleszámítanak a népszámláláskor összeírandó lakások, lakóegységek közé.

Az Interoptv. a 2015. január 1-jei hatályba lépéshez képest két év türelmi időt ad az adatgazdáknak arra, hogy nyilvántartásaikat alkalmassá tegyék együttműködésre, így az indulás után a következő népszámlálásig mintegy öt év áll rendelkezésre, ami vélhetően nem lenne

elegendő a jól előkészített, tesztelt és üzemszerűen, biztonságosan működő adatátvételek végrehajtására.

4.3. Humán erőforrás, végrehajtó szervezet

Az adminisztratív adatforrásokon alapuló népszámlálás megvalósításához, az adatátvételhez, adattisztításhoz megfelelő szaktudással bíró szakemberek szükségesek, akik az új típusú feladatokat el tudják látni:

- az adatátvételhez, az átvett adatok kezeléséhez értő emberi erőforrás, szakembergárda a statisztikai hivatalban,
- a népszámlálással foglalkozó szervezeti egység átstrukturálása az új típusú feladatokhoz,
- az informatikai támogatás növelése a statisztikai hivatalban,
- az adatgazdánál szakemberek és kapacitás biztosítása az adatátadáshoz és az együttműködéshez a KSH-val.

5. Adminisztratív adatforrásokon alapuló népszámlálás vagy statisztikai adatgyűjtéssel kombinált népszámlálás?

5.1. A módszertani választás szempontjai

A jelenleg rendelkezésre álló ismereteink alapján a népszámlálás újfajta módszertan szerinti megvalósíthatóságának ideje 2020-ig jogi, anyagi, szervezeti, adatminőségi, időbeli korlátok és problémák miatt még nem jön el. Reális megvalósíthatósága a kombinált népszámlálásnak van, azaz annak a módszernek, amikor az adminisztratív adatforrásokon alapuló alapadatokat statisztikai adatgyűjtéssel egészítik ki.

Amennyiben kizárólag adminisztratív adatforrásokon alapuló népszámlálás megvalósítására vállalkozna a hivatal, úgy jelentős adathiánnyal és minőségi problémákkal kellene szembenézni, ami sem az európai, sem pedig a hazai adatigényeket nem elégítené ki.

A népszámlálás módszerének kiválasztásakor nagyon fontos szempont, hogy az Európai Unió tagjaként Magyarországnak eleget kell tennie az európai népszámlálási rendeletnek, azaz az előírt adatokat időre és megfelelő minőségben elő kell állítani és az Eurostatnak átadni. A módszer kiválasztásában továbbra is szabad kezet kapnak a tagországok, de a minőség kritériuma olyan korlát és segítség is egyben, ami biztosítja az összehasonlítható adatokat produkáló népszámlálások lebonyolítását minden egyes tagországban.

A népszámlálási módszertan váltásánál többnyire a fokozatosság elvét követik az országok, azaz eleinte egy-két nem statisztikai adatforrást vonnak be az összeírás keretének kialakításához vagy a kérdőívek megszemélyesítéséhez, illetve az adat-előállításához. Már az adminisztratív adatforrások ilyen mértékű bevonása is előrelépésnek számít a korábbi hagyományos népszámlálásokhoz képest, és a külső adatok jövőbeni minél szélesebb körű felhasználására való felkészülést is segíti. (Számos uniós projekt jelenleg is arra hivatott, hogy az adminisztratív nyilvántartások felhasználhatóságát vizsgálja, illetve tesztelje. Magyarországon 2015 szeptemberében indult egy ilyen projekt.)

5.2. Statisztikai adatgyűjtéssel kombinált népszámlálás megvalósítása

5.2.1. Adatforrások

Egy kombinált népszámlálás megvalósításának kiindulási pontja, hogy tisztázzuk a tartalom és minőség alapján felhasználható adminisztratív adatforrásokban mely adatok állnak rendelkezésre, és mik azok az adatok, amelyeket csak statisztikai összeírással lehet biztosítani. Fontos, hogy az adminisztratív adatok biztos forrást nyújtsanak az alapadatokhoz, azaz a népesség- és a lakásszámokhoz.

A fő adminisztratív forrás a személyekre vonatkozóan a KEKKH kezelésében lévő személyiadat- és lakcímnnyilvántartás. Innen vehetők át a bejelentett lakcímmel rendelkezők főbb adatai:

- a) nem,
- b) születési idő,
- c) születési hely,
- d) születéskori lakóhely (történeti adatokban),
- e) lakóhely (lakcím-információk),
- f) tartózkodási hely (lakcím-információk),
- g) mióta lakik a lakcímen,
- h) állampolgárság,
- i) házasságkötésre vonatkozó adatok,
- j) termékenységi adatok (élveszületések száma, ideje és életben lévő gyermekek).

A lakások, épületek tekintetében az alapadatok átvétele adminisztratív adatforrásból egyelőre nem biztosított. A népszámlálási szempontból releváns lakásadatok egy részét a földhivatali nyilvántartások tartalmazzák, ám az adatok lekérdezése az adatgazdánál nagyon költségigényes, ezért a statisztikai felhasználás nem valósult meg. Ahhoz, hogy a földhivatali adatbázisok statisztikai célú hasznosíthatóságáról megfelelő információ álljon rendelkezésre, elengedhetetlen lenne az adatbázisok alapos elemzése.

Jelenleg a lakcímek statisztikai nyilvántartásának elsődleges forrása a 2011. évi népszámlálás címállománya, melyet a KSH a bontási és az építési statisztikából beérkező adatokkal aktualizál. Az újonnan épült lakások esetében az építési adatgyűjtés számos olyan adatot tartalmaz, mely a lakásállomány minőségi mutatójaként is hasznosítható lenne. Ez ugyanakkor két okból nem válthatja ki az összeírást. Egyrészt a régebben épült lakások minőségében beállt változásokat jelenleg semmilyen nyilvántartásból nem lehet nyomon követni; másrészt egyelőre nem biztos, hogy a használatba vételi engedélyezési eljárások valamennyi új építésű lakást regisztrálják. Az építésügyi hatósági eljárásokban zajló átalakítás, az elektronikus dokumentációs rendszer (ÉTDR) bevezetése lehetővé teszi, hogy az építési engedély kiadásától fogva nyomon lehessen követni egy lakásberuházás sorsát, így az engedélyezések elmaradt nyilvántartásából eredő torzítás várhatóan egyre kisebb lesz. Egyelőre azonban még jelentős eltérés tapasztalható, ha a kétféle adatforrást összevetjük. A 2001–2011 közötti időszakban a lakásépítési statisztika szerint 352 ezer lakás épült, míg a népszámlálás szerint 426 ezer. Az eltérésnek a fent említett regisztrációs problémán kívül oka lehet az építések, illetve a teljes lakásfelújítások időpontjának pontatlan felidézése. Azonban még egy kifogástalanul működő építésügyi nyilvántartás mellett is megoldatlan maradna a meglévő

lakások minőségváltozásának nyomon követése. A pótlólagos lakásberuházások nagyságrendje az új építésekéhez mérhető, megfigyelésük viszont az ismert nyilvántartásokra támaszkodva nem oldható meg.

Az alapadatok biztosításán kívül egyéb adminisztratív adatforrások is hasznosíthatók lehetnek. Ezek akár ki is válhatnak néhány változót, másrészt kiegészítő információval szolgálhatnak (pl. pótláshoz, validáláshoz). Ilyen adatforrások lehetnek a nyugdíjasok adatai, a regisztrált munkanélküliek adatai, a segélyezettek adatai, a bevándorlók adatai. A 4.2.-es részben felsorolt témakörökben azonban meglévő vagy a népszámlálással összekötött adatgyűjtéseket kell alkalmazni.

5.2.2. Adatátvétel, adatfogadás

A különféle nyilvántartásokból érkező adatok átvételére, fogadására és az adatok tisztítására a közelmúltban a statisztikai hivatalban jelentős fejlesztések történtek. A KARÁT és ADAMES nevű informatikai eszközök mind felhasználhatók népszámlálási célra. E rendszerek kifejezetten a különböző külső adatforrásokból érkező adatok kezelésére lettek kialakítva (lásd VIII. fejezet).

A KSH-ban az összekapcsolás a népszámláláskor cím alapján történt, de az EU statisztikai rendelete értelmében a népszámlálás módszertana szerint személyi alapon történő összekapcsolás is lehetséges. A személyi összekapcsolás alapja lehetne a név kiegészítve néhány egyéb információval, mint az anyja neve, születési hely és idő. Magyarországon a 2001. évi népszámlálással kezdődően a személy neve nem szerepel a kérdőíven a lakosság nagyobb együttműködési készsége érdekében, így a kiegészítő alapszolgáltatások olyan körére van szükség, amely egyértelműen beazonosítja a személyeket név nélkül is.

5.2.3. A mintavételes adatfelvétel

Az alapadatok adminisztratív adatforrásokból történő biztosítása mellett a további adatokat statisztikai adatgyűjtésből szükséges kinyerni. A rendelkezésre álló igazgatási adatforrások számbavétele és alapos módszertani és minőségi szempontú vizsgálata után lehet megállapítani, hogy mely népszámlálási információk nem állíthatók elő nyilvántartási adatokból. A KSH módszertani tervei szerint ezek egy megfelelően kialakított mintavételes felvétellel gyűjthetők be, aminek eredményei teljeskörűsíthetők. Az adatforrások integrálására és a mintából gyűjtött adatok extrapolálására kialakított módszertan alapján a minta nagyságának és jellegének meghatározása alapvető jelentőségű.

A mintavételes összeírás megvalósítható hagyományos módon vagy valamilyen informatikai eszközzel történő összeírással. A szokásos összeírási feladatoknál kérdőíveket és utasításokat kell tervezni, összeírókat kell szervezni, oktatni, le kell bonyolítani az összeírást, és fel kell dolgozni az adatokat. Az adatgyűjtési módszerek között választható lehetőségként szóba jön az internetes önkitöltés is, mint az adatszolgáltatók szempontjából is kényelmes megoldás, ami jelentősen csökkenti az adatgyűjtéssel összefüggő feladatok mennyiségét.

A személyiadat- és lakcímnnyilvántartáson alapuló kombinált népszámlálás esetében a személyekre vonatkozó adatok csak a bejelentett lakcímmel rendelkezőkre biztosítottak. Ez azt jelenti, hogy a szokásos lakóhelyre vonatkozó adat előállítására nem biztosítható, illetve az lesz a kiinduló koncepció, hogy mindenkinek a bejelentett lakcímét tekintjük a tényleges lakóhelyének. Ez komoly módszertani váltás, amely számos megfontolást igényel, és felmerül

annak a kockázata, hogy egyes változók torzulnak, megváltoznak a korábbi népszámlálási adatokhoz képest. Célszerű ezért a mintavételes felvételben is a különböző lakcímekre vonatkozóan részletes információkat gyűjteni, hogy javíthatóvá váljanak az esetleges bejelentettségéből fakadó eltérések. A lakcím-bejelentettséggel összefüggő problémák azonban nagy kihívást jelentenek ebben az esetben is.

Számolni kell azzal, hogy a mintából nem biztosíthatók település szintű vagy annál kisebb területi részletességű eredmények. A KSH a népszámlálási adatokból számos adatkérést teljesít a településnél kisebb szintekre, mint például a háztömbre vagy bizonyos szempontok szerint a településből lehatárolt területrésze. A mintavételes adatgyűjtéssel kombinált népszámlálásra való áttérés után ilyen adatok nem állíthatók elő. A települési adatok hiánya számos történeti idősort is megszakítana.

A módszertani váltás az EU számára történő adatszolgáltatási kötelezettség keretében átadandó adathiperkockákat²⁷ azonban nem érinti, mivel település szintre csak alapadatokat kérnek.

A jó minőségű adatok előállításának kritériuma a megfelelő minta kiválasztása megbízható becslési technikák alkalmazásával kombinálva. Ezeket a módszertani eszközöket úgy kell megválasztani, hogy az alapvető jellemzőkre a kisterületi adatok is biztosíthatók legyenek, ami a népszámlálással szemben az egyik legfontosabb elvárás. A mintát több körben célszerű tesztelni. Hasznos lehet az ilyen módszert alkalmazó országok mintavételi és becslési eljárásainak tanulmányozása is.

6. Kapcsolódás a 2016-os mikrocenzushoz

A 2016-os mikrocenzus is lehetőséget nyújt arra, hogy a szakemberek a következő népszámláláshoz további tapasztalatokat gyűjtsenek az alábbi területeken:

- a népesség-nyilvántartás adataival való munkák folytatása, bővítése,
- a KSH címregiszterének alkalmazása,
- mintavételi terv kidolgozása,
- az ADAMES adat-előkészítő rendszer alkalmazása,
- a mikrocenzus állományának összehasonlítása adminisztratív adatforrásokkal, tesztelesek,
- az adminisztratív állományokban található adatok tartalmának és a nyilvántartások rendszerének aktualizálása,
- az állami és önkormányzati nyilvántartások együttműködésének általános szabályairól szóló 2013. évi CCXX. törvénnyel összefüggésben már elindult kezdeményezések figyelemmel kísérése.

7. A 2021. évi népszámlálás során felhasználható legfontosabb adminisztratív adatforrások

Az alábbiakban felsoroljuk azokat az adminisztratív adatforrásokat (nyilvántartásokat), amik a hazai jogszabályok 2016. márciusi aktuális állapota szerint népszámlálási célra felhasználhatók. A „nyilvántartás adatai”-ban megadott ismervek felsorolása nem teljes körű, ezek csupán a

²⁷ A hiperkocka a kocka általánosítása háromnál több dimenzióra.

statisztikai felhasználás szempontjából lettek kiválogatva. A listák a nyilvántartások ügyvitellel összefüggő adatait nem tartalmazzák.

7.1. Személyiadat- és lakcímnyilvántartás

Adatgazda: Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala

Elrendelő jogszabály: 1992. évi LXVI. törvény a polgárok személyi adatainak és lakcímének nyilvántartásáról

Hatály: a Magyar Köztársaság területén élő magyar állampolgárok, a bevándorolt és a letelepedett jogállású, valamint a menekültként elismert személyek, valamint a szabad mozgás és tartózkodás jogával rendelkező személy.

A külföldön élő, illetőleg a Magyar Köztársaság területét a külföldi letelepedés szándékával elhagyó magyar állampolgárok akkor tartoznak a törvény hatálya alá, ha – a magyar állampolgárságuk igazolása után – kérik adataiknak a nyilvántartásba való felvételét és meglévő adataik további kezelését. Adataikat kérelmükre – az archív adatok kivételével – a nyilvántartásból törölni kell.

A nyilvántartás adatai: név; állampolgárság; menekült, bevándorolt, letelepedett jogállás, illetve a szabad mozgáshoz és tartózkodáshoz való joggal történő rendelkezés; nem; születési hely és idő; anyja neve; személyi azonosítója; elhalálozás helye és ideje; lakcím (lakóhely, tartózkodási hely); családi állapot; a házasságkötés helye.

A nyilvántartás történeti állománya tartalmazza a polgár minden eddigi, a törvény hatálya alá tartozó adatát és adatváltozását az érintett elhalálozásától, magyar állampolgárságának megszűnésétől, illetve a külföldi letelepedés szándékával külföldre távozásától, bevándorolt, menekült, letelepedett jogállásának megszűnésétől, illetve a szabad mozgás és tartózkodás jogával rendelkező személy tartózkodási jogának megszűnésétől számított 15 évig, kivéve, ha jogállásának megszűnését a magyar állampolgárság megszerzése eredményezte. Ha az érintett magyar állampolgársága a külföldi letelepedés szándékával történt külföldre távozás nyilvántartásba vételét követően szűnik meg, e tény a nyilvántartás történeti állományán át kell vezetni.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: A nyilvántartás szervei statisztikai célra – a népszámlálással kapcsolatos feladatok és a Központi Statisztikai Hivatal statisztikai célú adatkezelése kivételével – csak olyan módon szolgáltathatnak adatot, amellyel a polgár és az adat kapcsolata nem állítható helyre.

7.2. Földhivatali ingatlan-nyilvántartás

Adatgazda: Földmérési és Távérzékelési Intézet

Elrendelő jogszabály: 1997. évi CXLI. törvény az ingatlan-nyilvántartásról

Hatály: az ország valamennyi ingatlana

A nyilvántartás adatai: Az ingatlan adatai: a település neve, az ingatlan fekvése (belterület, külterület), a belterületen lévő ingatlannál a közterület neve, jellege, a ház száma, a helyrajzi szám és területnagyság; művelési ág és a művelés alól kivett terület elnevezése; az épület fő rendeltetés szerinti jellegét (lakóház, üdülő, gazdasági épület stb.), az ingatlan jogi jellegét (pl. társasház, szövetkezeti ház, műemlék, tanya stb.) stb., minőségi osztály, kataszteri tisztajövedelem; ingatlan-nyilvántartási szempontból szükséges egyéb adat.

A tulajdonosok adatai: természetes személyazonosító adatok, lakcím, személyi azonosító; illetve a szervezet megnevezése, székhelye, statisztikai azonosító.

Az ingatlan-nyilvántartásba bejegyezhető jogok: tulajdonjog, állami tulajdonban álló ingatlan esetében az állam tulajdonosi jogait gyakorló szervezet, és a vagyongazdálkodási jog; helyi önkormányzati tulajdon esetében a vagyongazdálkodási jog; a lakásszövetkezeti tagot megillető állandó használati jog, megállapodáson és bírósági határozaton alapuló földhasználati jog,
Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az ingatlanügyi hatóság kérelemre, ha jogszabály másként nem rendelkezik, az Elektronikus Kormányzati Gerinchálózaton (EKG) keresztül biztosít hozzáférést a Központi Statisztikai Hivatal számára, hogy az adatokat statisztikai célú felhasználásra a személyek azonosítására alkalmatlan módon lekérdezze.

7.3. OEP-adatbázis

Adatgazda: Országos Egészségbiztosítási Pénztár

Elrendelő jogszabály: (1) A társadalombiztosítás ellátásaira és a magánnyugdíjra jogosultakról, valamint e szolgáltatások fedezetéről szóló 1997. évi LXXX. törvény (Tbj.); (2) A kötelező egészségbiztosítás ellátásairól szóló 1997. évi LXXXIII. törvény; (3) Az egészségügyi és a hozzájuk kapcsolódó személyes adatok kezeléséről és védelméről szóló 1997. évi XLVII. törvény; (4) Az egészségügyi szolgáltatások Egészségbiztosítási Alapból történő finanszírozásának részletes szabályairól szóló 43/1999 Korm. rendelet; (5) Az egészségügyi szakellátás társadalombiztosítási finanszírozásának egyes kérdéseiről szóló 9/1993 (IV.2.) NM. rendelet és (6) Az egészségügyi ellátásban használt szakmai kódrendszerek és finanszírozási paraméterek karbantartásának jogi szabályozásáról szóló 6/1998 (III.11.) NM. rendelet.

A nyilvántartás hatálya, adatai és hozzáférés az adatokhoz résznyilvántartások szerint:

7.3.1. TAJ-nyilvántartás

Hatály: Az adatbázis az 1995-ben élő személyeket, valamint az azóta nyilvántartásban került személyek adatait tartalmazza.

A nyilvántartott adatok: személyi adatok (név, leánykori név, anyja neve, születési hely és idő); nem, családi állapot, állampolgárság, lakóhely (tartózkodási hely); TAJ-szám

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: A Tbj. 43.§ (2/a) bekezdés szerint az adatbázisból a tárolt személyes adatok „statisztikai célra felhasználhatók és statisztikai célú felhasználásra a Központi Statisztikai Hivatal részére személyazonosításra alkalmas módon átadhatók”.

7.3.2. Bejelentett személyek jogviszonyadatainak nyilvántartása

Hatály: Az adatbázis az 1998. január 31-én jogviszonyban álló természetes személyek adatait és 1998. január 31-től keletkezett jogviszonyok adatait tartalmazza.

A nyilvántartott adatok: név, TAJ, foglalkoztató neve, székhelye, nyilvántartási száma, jogviszony jogcíme, FEOR, jogviszony időtartama, heti munkaidő.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: A Tbj. 43.§ (2/a) bekezdés szerint az adatbázisból a tárolt személyes adatok „statisztikai célra felhasználhatók és statisztikai célú felhasználásra a Központi Statisztikai Hivatal részére személyazonosításra alkalmas módon átadhatók”.

7.3.3. Az egészségbiztosítási pénzbeli ellátásaival, a baleseti táppénzzel és az utazási költségtérítéssel kapcsolatos eljárás során kezelt adatok nyilvántartása

Hatály: Pénzbeli egészségbiztosítási ellátást (táppénz, terhességi-gyermekágyi segély, gyermekgondozási díj), baleseti táppénzt, valamint utazási költségtérítést igénylők.

A nyilvántartott adatok: személyi adatok (név, születési név, anyja neve, születési hely és idő), szükség esetén a gyermek személyi adatai, állampolgárság, lakóhely (vagy tartózkodási hely), TAJ, foglalkoztatóra vonatkozó adatok, az ellátás megállapításához szükséges jövedelmi adatok.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: A Tbj. 43.§ (2/a) bekezdés szerint az adatbázisból a tárolt személyes adatok „statisztikai célra felhasználhatók és statisztikai célú felhasználásra a Központi Statisztikai Hivatal részére személyazonosításra alkalmas módon átadhatók”.

7.4. ONYF-adatbázisok

Adatgazda: Országos Nyugdíjbiztosítási Főigazgatóság

Elrendelő jogszabály: (1) 1997. évi LXXX. törvény a társadalombiztosítás ellátásaira és a magánnyugdíjra jogosultakról, valamint e szolgáltatások fedezetéről (2) 1997. évi LXXXI. törvény a társadalombiztosítási nyugellátásról.

Hatály: (1) a Tbj. szabályai alapján biztosítottak minősülő személyek; és (2) a saját jogú és hozzátartozói nyugellátásban részesülő személyek.

A nyilvántartás adatai résznyilvántartások szerint:

7.4.1. KELEN (központi elektronikus nyilvántartási rendszer)

Társadalombiztosítási azonosítási jel, név, születési adatok; családi állapot, állampolgárság, lakóhely (tartózkodási hely); foglalkozás, munkahely, kereset, jövedelem, a levont és befizetett járulékok; szolgálati idő.

7.4.2. NYUGDMEG (nyugdíjigény elbírálását segítő rendszer)

Társadalombiztosítási azonosítási jel, név, születési adatok; családi állapot, állampolgárság, lakóhely (tartózkodási hely); foglalkozás, munkahely, kereset, a rokkantság fokára, az egészségkárosodás mértékére, a szakmai munkaképességre, a rehabilitálhatóságra, továbbá az egészségi állapotra vonatkozó adatok (ha a nyugellátás megállapítása az egészségi állapot figyelembevételével történik); az élettársra és az eltartott hozzátartozói minőségre vonatkozó adatok (ha a nyugellátás megállapítása az élettársi, hozzátartozói minőség figyelembevételével történik).

7.4.3. NYUFUR (az ellátások folyósításával kapcsolatos adatokat nyilvántartó rendszer)

Társadalombiztosítási azonosítási jel, név, születési adatok; családi állapot, állampolgárság, lakóhely (tartózkodási hely); foglalkozás, munkahely, kereset, nyugdíjak, járadékok és nyugdíjszerű rendszeres ellátások típusa, összege, szolgálati idő, nyugdíjazás éve.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adatok statisztikai célra felhasználhatók és statisztikai célú felhasználásra – személyazonosításra alkalmatlan módon – átadhatók.

7.5. Adóadatbázis

Adatgazda: Nemzeti Adó- és Vámhivatal

Elrendelő jogszabály: Az adózás rendjéről szóló 2003. évi XCII. törvény

Hatály: Magyarország területén lakóhellyel, szokásos tartózkodási hellyel rendelkező vagy egyébként itt tartózkodó és vagyonnal rendelkező vagy bevételt, jövedelmet (nyereséget) elérő magánszemély.

A nyilvántartás adatai: egyedi azonosító, név, anyja neve, születési idő, állampolgárság, magyarországi lakóhely települése, nyilvántartásba kerülés időpontja.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adózással összefüggő tény, adatot, iratot a hivatalos statisztikai szolgálathoz tartozó szervek statisztikai célra felhasználhatják, ha a titoktartási kötelezettség megtartását a feldolgozás során biztosítják és – amennyiben a statisztikáról szóló törvény másként nem rendelkezik – azt a későbbi egyedi azonosításra alkalmatlanná teszik.

7.6. Oktatási információs rendszer

Adatgazda: Emberi Erőforrások Minisztériuma – Oktatási Hivatal

Elrendelő jogszabály: (1) 2011. évi CXC. törvény a nemzeti köznevelésről és (2) 2005. évi CXXXIX. törvény a felsőoktatásról

A nyilvántartások:

7.6.1. Közoktatási információs rendszer

Hatály: (1) akit a közoktatásban pedagógus-munkakörben, illetve nevelő- és oktatómunkát közvetlenül segítő alkalmazotti munkakörben, pedagógiai előadó vagy pedagógiai szakértő munkakörben alkalmaznak. (2) tanulói jogviszonyt létesített személyek.

A nyilvántartott adatok: (1) az érintett neve, születési adatai, azonosító szám, a végzettségre és szakképzettségre vonatkozó adatok, a munkahely címe és típusa, valamint OM-azonosítóját. (2) tanuló neve, születési helye és ideje, azonosító száma, anyja neve, lakóhelye, tartózkodási helye, állampolgársága, OM-azonosítója, diákigazolványának száma, valamint az érintett nevelési-oktatási intézmény adatai.

7.6.2. Felsőoktatási információs rendszer

Hatály: (1) felsőoktatásban alkalmazottként foglalkoztatott, (2) hallgatói jogviszonyt létesített személyek.

A nyilvántartott adatok: (1) név, nem, születési név, születési hely és idő, anyja neve, állampolgárság, azonosító szám; lakóhely, értesítési cím; munkaviszonyra, közalkalmazotti jogviszonyra, megbízási jogviszonyra vonatkozó adatok; az adatokat igazoló okiratok azonosítására szolgáló adatok. (2) a hallgató (doktorjelölt) neve, születési neve, anyja neve, születési helye és ideje, állampolgársága, neme, lakóhelye, tartózkodási helye, értesítési címe; nem magyar állampolgár esetén a Magyarország területén való tartózkodás jogcíme és az erre jogosító okirat azonosító adatai; a jogviszonyra vonatkozó adatok; a jelenlegi képzésre vonatkozó adatok, a külföldi felsőoktatási résztanulmányok helye, ideje; a hallgatóra vonatkozó dokumentumok fajtája és azonosítója; a záróvizsga szakja, időpontja és eredménye, a kiadott okiratok tartalmára és azonosítására vonatkozó adatok; a hallgatói pályakövetéssel kapcsolatos adatok.

7.6.3. Diákigazolvány-rendszer

Hatály: tanulói vagy hallgatói jogviszonyt létesítettek, akik diákigazolványt igényeltek.

A nyilvántartott adatok: születési hely és idő, lakcím, tartózkodási hely, állampolgárság, azonosító száma, az iskola neve és címe.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adatok statisztikai célra felhasználhatók, és statisztikai felhasználás céljára személyazonosításra alkalmatlan módon átadhatók.

7.7. Idegenrendészeti nyilvántartás

Adatgazda: Bevándorlási és Állampolgársági Hivatal

Elrendelő jogszabály: (1) 2007. évi I. törvény a szabad mozgás és tartózkodás jogával rendelkező személyek beutazásáról és tartózkodásáról és (2) 2007. évi II. törvény a harmadik országbeli állampolgárok beutazásáról és tartózkodásáról.

Hatály: Három hónapnál hosszabb ideig Magyarországon tartózkodó, szabad mozgás és tartózkodás jogával rendelkező külföldi állampolgárságú személyek, valamint a beutazó harmadik országbeliek.

A nyilvántartás adatai a résznyilvántartások szerint:

7.7.1. Természetes személyazonosító adatok

Nevek; születési hely és idő; nem; anyja neve; és állampolgárság.

7.7.2. Vízumnyilvántartás

A harmadik országbeli állampolgár természetes személyazonosító adatai; a beutazás célja és a tartózkodás tervezett időtartama; a vízum kiadása vagy meghosszabbítása iránti kérelem elutasításának, valamint a vízum visszavonásának ténye és azok indoka; a be- és kiutazás ideje, a határátlépés helye; a szálláshely címe.

7.7.3. Hosszú távú tartózkodási engedélyek és bevándorlási engedélyek nyilvántartása

Az EU-csatlakozás előtt kiadott engedélyek adatait tartalmazza: természetes személyazonosító adatok; állampolgárság; külföldi lakóhely; az útlevél adatai; családi állapot; iskolázottság; foglalkozás; a bevándorlási kérelem adatai; az engedély kiadásának és visszavonására vonatkozó adatok; lakóhely címe Magyarországon.

7.7.4. Tartózkodási és letelepedési engedélyek nyilvántartása

Az EU-csatlakozást követően harmadik országból érkezők számára kiadott engedélyek adatai: természetes személyazonosító adatok; állampolgárság; külföldi lakóhely; az útlevél adatai; családi állapot; iskolázottság; foglalkozás; a bevándorlási kérelem adatai; az engedély kiadásának és visszavonására vonatkozó adatok; lakóhely címe Magyarországon; a tartózkodás célja; a bevándorló által használt és külföldön regisztrált gépjármű regisztrációs adatai. Az iskolázottság és a foglalkozás több nyilvántartásban is szerepel, ezek fontos adatok az integráció szempontjából. Fontos lenne ezeknek a mutatóknak a minél pontosabb és teljes körű számbavétele, mert a BÁH-adatok alapján készült statisztikákban meglehetősen hiányosan szerepelnek.

7.7.5. EGT-ből érkezők és családtagjaik nyilvántartása a további résznyilvántartásokra tagozódik

- i. A regisztrációs igazolással ellátott EGT-állampolgárok nyilvántartása.
- ii. A tartózkodási kártyával rendelkezők nyilvántartása.
- iii. Az állandó tartózkodási kártyával rendelkezők nyilvántartása.

Nyilvántartott adatok: személyes adatok, a tartózkodás jogszerűségét, a beutazást, tartózkodást és a lakóhelyet igazoló iratok azonosító adatai; lakóhely; az EGT-állampolgár családi állapota, ha ez másik személy jogszerű tartózkodását generálja.

- iv. A kiutasított vagy beutazási és tartózkodási tilalom hatálya alatt álló EGT-állampolgárok és családtagok nyilvántartása.

Nyilvántartott adatok: az EGT-állampolgár és a családtag személyazonosító adatai; a kiutasítás, illetve a beutazási és tartózkodási tilalom elrendelésének jogalapja, az intézkedés részletei; a kilépés helye, ideje, az EGT-állampolgár és a családtagja arcképmása, ujjnyomata.

- v. A személyes szabadságukban korlátozott EGT-állampolgárok és családtagjaik nyilvántartása.

Nyilvántartott adatok: az EGT-állampolgár és a családtagja személyazonosító adatai; a büntetőeljárás adatai; a rendkívüli esemény adatai, intézkedés részletei.

- vi. Az úti okmány, illetve a személyazonosító igazolvány elvesztését bejelentő EGT-állampolgárok és családtagjaik nyilvántartása.

Nyilvántartott adatok: az EGT-állampolgár és a családtagja személyazonosító adatai; az elvesztésként bejelentett úti okmány és bejelentés körülményeinek adatai; a tartózkodási jogot igazoló okmány adatai; az új úti okmány vagy személyazonosító igazolvány adatai.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adatok statisztikai célra felhasználhatók, és azokról személyazonosításra alkalmatlan módon a központi idegenrendészeti nyilvántartást kezelő szerv statisztikai célra adatokat szolgáltatathat. A résznyilvántartások adataiból a KSH részére statisztikai célú adatkezelés érdekében az adatok személyazonosításra alkalmas módon is átadhatók.

7.8. Menekültügyi nyilvántartás

Adatgazda: Bevándorlási és Állampolgársági Hivatal

Elrendelő jogszabály: 2007. évi LXXX. törvény a menedékjogról

Hatály: Azok a külföldiek, akik dublini eljárás hatálya alatt állnak, elismerés iránti kérelmet nyújtottak be, vagy menedékjogot élveznek.

A nyilvántartás adatai: természetes és egyéb személyazonosító adatok; a dublini eljárás alapján történt átvétel és átadás adatai; a menekültként, oltalmazottként, vagy a menedékesként való elismerés iránti kérelem adatai és az elbírálás; családi állapot, foglalkozás, iskolai végzettség; lakóhely, tartózkodási hely, szálláshely; a származási ország neve; a kérelem indoka; az együtt érkező családtagok adatai; jövedelmi, vagyoni helyzetre vonatkozó adatok.

Fontos lenne az adatbázisok egységesítése. Itt például szerepelnek a jövedelemre, vagyoni helyzetre vonatkozó adatok. Ezeket a különböző engedélyekkel rendelkező bevándorlóknál is be lehetne gyűjteni, ugyanaz az adatgazda.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adatok statisztikai célra felhasználhatók, és az azokat tartalmazó nyilvántartásokból személyazonosításra alkalmatlan módon statisztikai célra adatok szolgáltathatók.

A Központi Statisztikai Hivatal részére statisztikai célú adatkezelés érdekében személyazonosításra alkalmas módon átadhatók: név, születési hely és ideje, nem, a jelenlegi és korábbi állampolgárságára vagy hontalan státuszára vonatkozó adatok; a menekültként, oltalmazottként, vagy a menedékesként való elismerés iránti kérelem adatai; családi állapot, foglalkozás, iskolai végzettség; lakóhely, tartózkodási hely, szálláshely; a származási ország neve.

7.9. Külföldi állampolgárok adó adatbázisa

Adatgazda: Nemzeti Adó- és Vámhivatal

Elrendelő jogszabály: Az adózás rendjéről szóló 2003. évi XCII. törvény

Hatály: Magyarország területén lakóhellyel, szokásos tartózkodási hellyel rendelkező vagy egyébként itt tartózkodó és vagyonnal rendelkező vagy bevételt, jövedelmet (nyereséget) elérő magánszemély.

A nyilvántartás adatai: egyedi azonosító, név, anyja neve, születési idő, állampolgárság, magyarországi lakóhely települése, nyilvántartásba kerülés időpontja.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adózással összefüggő tény, adatot, iratot a hivatalos statisztikai szolgálathoz tartozó szervek statisztikai célra felhasználhatják, ha a titoktartási kötelezettség megtartását a feldolgozás során biztosítják és – amennyiben a statisztikáról szóló törvény másként nem rendelkezik – azt a későbbi egyedi azonosításra alkalmatlanná teszik.

7.10. E-közmű

Adatgazda: Lechner Lajos Tudásközpont Nonprofit Kft.

Elrendelő jogszabály: 324/2013. (VIII. 29.) Korm. rendelet az egységes elektronikus közműnyilvántartásról

Hatály: a víziközmű-szolgáltatást, az elválasztott rendszerű csapadékvíz-elvezetést, a szénhidrogén-ellátást, a távhőszolgáltatást, a villamosenergia-ellátást biztosító közművezetésekre, valamint az elektronikus hírközlést biztosító nyomvonalas építmények.

A nyilvántartás adatai: a közművezeték nyomvonala, elhelyezkedésének módja, a közműhálózati hierarchiában betöltött funkciója, szakági típusa, az általa szállított közeg, valamint a szállítás módja.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: az e-közmű által szolgáltatott adat, valamint a közművezeték tulajdonosáról, üzemeltetőjéről, szolgáltatói engedélyeséről és ezek elérhetőségéről szóló információ a felhasználók számára ügyfélkapus azonosítást követően nyilvános, és díjmentesen elérhető.

7.11. Elektronikus anyakönyvi nyilvántartás

Adatgazda: Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala

Elrendelő jogszabály: 2010. évi I. törvény az anyakönyvi eljárásról

Hatály: Magyarország területén bekövetkező valamennyi, illetve az anyakönyvezést kérvényező, külföldön élő magyar állampolgár anyakönyvi eseményeinek bekövetkezése, az azok alapján létrejövő családi kapcsolatok, valamint a házasság és a bejegyzett élettársi kapcsolat megszűnése.

A nyilvántartás adatai: az érintett személyazonosító adatai (az egyedi elektronikus anyakönyvi azonosítót, az érintett születési családi és utónevét, születési adatait, személyi azonosítóját, nemét, szülők születési nevét, állampolgársági adatok, házassági nevét, családi állapotát) és

- a születéssel kapcsolatban: származási hely, többes születés,
- a házassággal kapcsolatban: a házasságkötés helye és ideje, a házastárs adatai, a házasság megszűnésének, felbontásának vagy érvénytelenné nyilvánításának ténye, időpontja,
- a bejegyzett élettársi kapcsolattal összefüggésben: a létrehozás helye és ideje, a bejegyzett élettárs adatai, a bejegyzett élettársi kapcsolat megszűnésének, megszüntetésének, felbontásának vagy érvénytelenné nyilvánításának ténye, időpontja,
- a halálessel kapcsolatban: a haláleset helye és ideje, a holtak nyilvánítás ténye vagy a halál tényének bírósági megállapítása.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az elektronikus anyakönyvben kezelt adatok közül a Központi Statisztikai Hivatal a statisztikáról szóló törvényben és annak végrehajtásáról szóló rendeletben meghatározott, a népmozgalmi adatszolgáltatás körébe tartozó adatokat a központi nyilvántartó szervtől elektronikus formában, térítésmentesen jogosult átvenni.

7.12. Szociális nyilvántartás (CSTINFO)

Adatgazda: Magyar Államkincstár

Elrendelő jogszabály: 1998. évi LXXXIV. törvény a családok támogatásáról

Hatály: Magyarország területén élő magyar állampolgár, bevándorolt vagy letelepedett jogállású, továbbá a magyar hatóság által menekültként, illetve hontalanként elismert személy, a szabad mozgás és tartózkodás jogával rendelkező személy, összevont engedéllyel rendelkező harmadik országbeli állampolgár, ha a munkavállalást számára hat hónapot meghaladó időtartamra engedélyezték.

A nyilvántartás adatai: a családtámogatási ellátásban részesülő, valamint ha az ellátást más személyre tekintettel állapítják meg, úgy e személy, továbbá a fogyatékosági ellátásban részesülő személy természetes személyazonosító adatai, lakcím, tartózkodási cím, társadalombiztosítási azonosító jele, valamint az ellátás jogcíme és összegére vonatkozó adatok.

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az e törvény felhatalmazása alapján nyilvántartást vezető szervek a nyilvántartásban szereplő adatokat személyazonosító adatok nélkül statisztikai célra felhasználhatják, illetőleg azokból statisztikai célra adatot szolgáltathatnak.

7.13. Álláskeresők nyilvántartása

Adatgazda: Nemzetgazdasági Minisztérium

Elrendelő jogszabály: 1991. évi IV. törvény a foglalkoztatás elősegítéséről és a munkanélküliek ellátásáról

Hatály: Magyarország területén élő magyar állampolgár, menekültként, oltalmazottként vagy menedékesként elismert, továbbá a bevándorolt vagy letelepedett jogállású, valamint a szabad mozgás és tartózkodás jogával rendelkező személyek.

A nyilvántartás adatai: természetes személyazonosító adatok, állampolgárság, bevándorolt, letelepedett, menekült vagy oltalmazott jogállás, a menekültügyi hatóságnál a menekültként, oltalmazottként, vagy menedékesként történő elismerésre irányuló kérelem benyújtásának, vagy a harmadik országbeli állampolgár kijelölt helyen való tartózkodása elrendelésének ténye, lakcím (lakóhely, tartózkodási hely) és értesítési cím, az álláskeresői ellátás, a foglalkoztatást elősegítő támogatás, valamint a munkaerőpiaci szolgáltatás megállapításához és igénybeviteléhez és nyújtásához szükséges e törvényben és más törvényben meghatározott adatok,

Hozzáférés az adatokhoz az elrendelő jogszabály szerint: Az adatok statisztikai célra felhasználhatók és statisztikai célú felhasználásra – személyazonosításra alkalmatlan módon – átadhatók.

VI. A másodlagos adatok minőségügyi kérdései

Ahogy azt a III. fejezet részletesen is ismertette: másodlagos adatforrások gyakorlatilag a statisztika előállításának bármely munkafázisában felhasználhatók. Így az előbbieket minőségét is abból a szempontból ítélni lehet meg, hogy milyen mértékben járulnak hozzá a felhasználással előállított statisztikák minőségéhez. E fejezetben először bemutatjuk a KSH statisztikára vonatkozó minőségügyi keretrendszerét, majd ez után a másodlagos adatforrások minőségéhez közvetlenül kapcsolódó ESR-irányelveket, végül a KSH-ban felmerülő kérdéseket, és ezek illeszkedését a statisztika minőségügyi keretrendszerébe.

1. A KSH minőségügyi keretrendszere

A KSH minőségügyi keretrendszerét a hivatalban összegyűjtött tapasztalatokra, valamint a nemzetközi együttműködés keretében kialakított ajánlásokra és megismert gyakorlatokra építve dolgozta ki, illetve folyamatosan fejleszti. A munka elvi alapját a KSH minőségpolitikája és a hivatalos statisztikára elfogadott minőség fogalma jelenti. A minőségügyi keretrendszer struktúrája, elemei, eszközei, valamint működési rendszere ezekre, valamint a teljes körű minőségirányítás (Total Quality Management – TQM) alapelveire épül. A keretrendszer felhasználja a hivatalban kifejlesztett infrastruktúrát. A már elkészült vagy tervezett szakmai dokumentumok (minőségi irányelvek és más szabványok) illeszkednek a keretrendszerbe.

A nemzetközi együttműködés keretében a KSH több mint 15 éve bekapcsolódott az ESR minőségügyi keretrendszerének²⁸ kidolgozásába, ugyanígy részt vett az ENSZ Statisztikai Bizottságának hasonló projektjében²⁹ is.

²⁸ Az Eurostat honlapján elérhetők az Európai Statisztikai Rendszer (ESR) minőséggel kapcsolatos anyagai (<http://ec.europa.eu/eurostat/web/quality>), az ENSZ rendszerrel való megfeleltetés az ENSZ minőségi honlapján érhető el.

²⁹ A keretrendszer sémája, a hozzá tartozó útmutató, valamint szövszedet elérhető az ENSZ Statisztikai Bizottsága 2012 óta működtetett statisztika minőségbiztosítási honlapjáról. (<http://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx>)

1.1. A KSH minőségpolitikája, minőségfogalma

A statisztikai törvény kimondja, hogy a statisztika feladata, hogy a felhasználók számára hű képet adjon a valóságról, a természet, a gazdaság és a társadalom helyzetéről, változásairól.

Az ISO 9000/2005 definíció alapján a minőség annak mértéke, hogy a saját jellemzők egy csoportja mennyire teljesíti a követelményeket³⁰.

A statisztikára alkalmazva ez azt jelenti, hogy a hivatalos statisztikai adatok minősége mennyire felel meg a felhasználók igényeinek. A felhasználók különböző csoportjainak különbözők az igényei, ezért a statisztikák széles választékát kell előállítani, és a különböző statisztikák mellé meg kell adni mindazokat a minőségre vonatkozó információkat, amik alapján az adott felhasználási célra való alkalmasság megítélhető. Ezek az információk elsősorban az adat minőségére, az adat-előállítás folyamatára, valamint az előállító intézményre vonatkoznak. A KSH minőségpolitikájában³¹ kinyilvánítja a minőség iránti elkötelezettségét, és közzéteszi a minőség fogalmát, amit az alábbiakban ismertetünk.

Az adatminőség – összhangban az ESR minőségértelmezésével³², és az ISO minőségdefiníciójával – felhasználásra való alkalmasságot jelent, és az elvárt mértékű megfelelést az a következő, egymással kapcsolatban levő minőség-összetevőknek.

A statisztikai termék (a felhasználók számára átadott, publikált statisztikai adat) legyen:

RELEVÁNS – legyen alkalmas arra a célra, amire szánták kielégítve a felhasználók igényeit.

PONTOS – a statisztikai adat a lehető legnagyobb mértékben közelítse a jellemezni kívánt jelenséget leíró valós értékét.

IDŐSZERŰ – a leírt esemény, jelenség referenciaideje és az adat közlése között eltelt idő a lehető legrövidebb legyen.

IDŐBEN PONTOS – az adatközlés időpontja és a publikálásra eredetileg megcélzott időpont essen egybe (a tervezett időpontokat a KSH a tájékoztatási naptárban teszi közzé a honlapján).

HOZZÁFÉRHETŐ – a felhasználók lehető legszélesebb körének számára legyenek adottak azok a feltételek, amelyek mellett hozzájuthatnak a statisztikai termékekhez.

ÉRTHETŐ – a KSH-nak segítséget kell nyújtania a felhasználók számára a statisztikai adatok értelmezéséhez. Ezért a statisztikai adatok mellé rendelkezésre kell bocsátani a használatukat segítő információkat (metaadatokat) és a statisztikai adatok minőségére vonatkozó információkat is elérhetővé kell tenni.

ÖSSZEHASONLÍTHATÓ ÉS KOHERENS – a statisztikai adat legyen összhangban az egyéb, vele logikai, gazdasági és társadalmi szempontból kapcsolatba hozható adatokkal, és legyen alkalmas az időbeli, földrajzi területek közötti és tárgykörök szerinti összehasonlításra.

A felsorolt hét minőség-összetevő szerinti értékelés megvalósítására egységes minőségindikátorokat fejlesztettek ki a nemzetközi együttműködés keretében. A minőségindikátorok olyan kvantitatív vagy kvalitatív mérőszámok, melyek a termék, illetve az

³⁰ „The degree to which a set of inherent characteristics fulfils requirements.”

³¹ Elérhető a KSH honlapjáról: http://www.ksh.hu/docs/bemutakozas/hun/ksh_minosegpolitikaja_2014.pdf

³² Lásd európai statisztikai rendelet, valamint a Gyakorlati Kódex termékekre vonatkozó elvei.

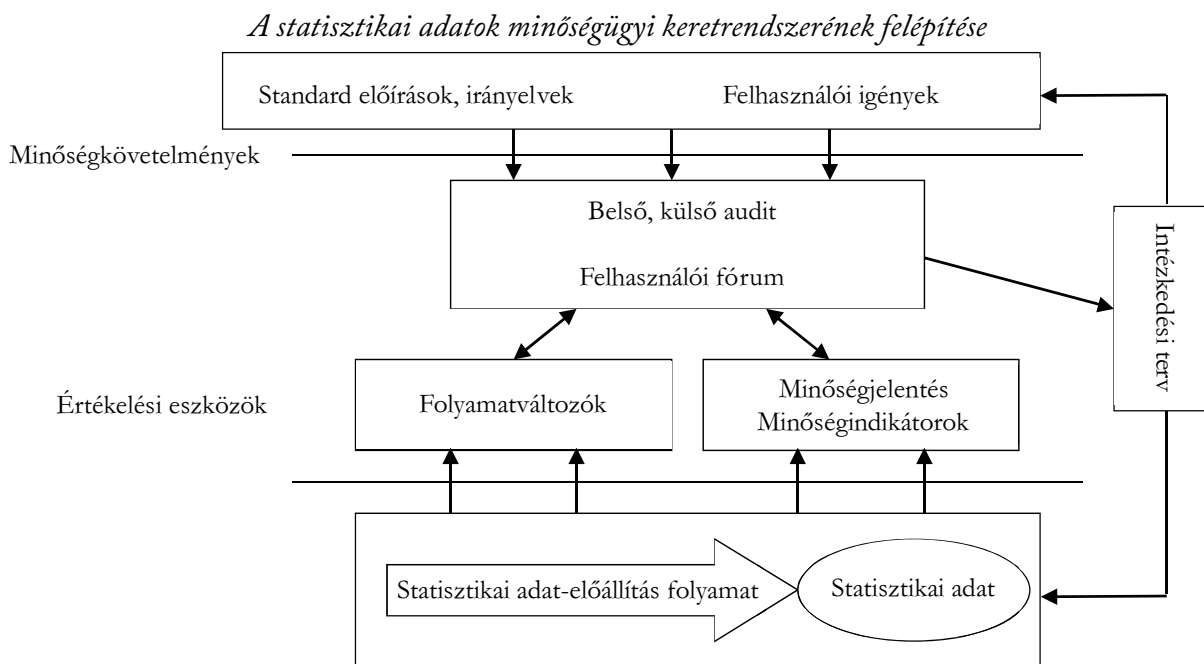
előállítási folyamat minőségét jelzik. A KSH jelenleg használt minőségindikátorai megtalálhatók a Minőségi irányelvek című dokumentum végén található mellékletekben.

A minőség jellemzője az is, hogy az adott minőségű statisztikai terméket milyen adatszolgáltatói és hivatali terhek, költségek mellett állítják elő.

1.2. A minőségügyi keretrendszer elemei

A minőségügyi keretrendszer olyan követelményeket fogalmaz meg, amelyeknek az adat-előállítási folyamat és a statisztikai adat meg kell, hogy feleljen. A követelményoldalhoz tartoznak a jogszabályok, szabványok, minőségi irányelvek³³, illetve a hazai és nemzetközi minőségi kritériumok. A keretrendszer másik fontos összetevője a mérés, melynek célja, hogy meghatározza és dokumentálja, hogy a folyamat és az adat mennyire felel meg a feltételeknek, kritériumoknak. A mérés eszközei a különböző indikátorok és a minőségjelentés. A rendszer harmadik pillére az *értékelés*. Ennek lényege, hogy a mért jellemzőket – önértékelés (pl. önértékelő kérdőív), illetve belső vagy külső audit segítségével – összevetjük az elvárásokkal. Az eredmények értelmezése, elemzése alapján az adatok folyamatos javítása érdekében fejlesztési, intézkedési tervek készíthetők. Ezt mutatja be az alábbi ábra.

7. ábra



Az adatok minőségét az adat-előállítási folyamat határozza meg. Sok esetben az adat minőségi jellemzőit nem is tudjuk közvetlenül mérni, csak az adat-előállítás folyamatából következtethetünk rájuk. Például az adat pontossága a statisztikailag becsült érték valós értékhez való közelséget jelenti, de a valós érték nem ismert. A nemválaszolási arány az egyik pontosságra vonatkozó indikátor, holott például egy 20%-os nemválaszolási arány mellett a becslés lehet torzítatlan is, és jelentős hibával terhelt is, attól függően, hogy a nemválaszolók véletlen mintának tekinthetők-e, vagy valamely csoportra koncentrálódnak. Az adat-előállítási folyamat

³³ A KSH minőségi irányelvei elérhetők a KSH honlapjáról: http://www.ksh.hu/docs/bemutakozas/hun/minosegi_iranyelvek_2014.pdf

közben a minőség monitorozása során beazonosított hibákat a munkafolyamat során csökkenthetjük vagy felszámolhatjuk, például – nemválaszolás esetén – pótlással/imputálással vagy súlyozással. Ezért a statisztikai munkafolyamatról is információkat kell szolgáltatnunk. A teljes garanciához a statisztikát előállító intézményre, szervezetre vonatkozóan is meg kell adnunk az alapvető információkat (az intézmény azonosító adatain túl olyan információk, mint a szakmai függetlenség, adatgyűjtésre való felhatalmazás).

A tágabb értelemben vett statisztikai adat-előállítási folyamat az adatszolgáltatónál kezdődik, és az adatok felhasználóinál ér véget. Az adatszolgáltató is része az adat-előállítási folyamatnak, a statisztika költségei/terhei részben ott merülnek fel. Másrészt a felhasználóhoz eljutott, értelmezett statisztikai adat jelenti a végterméket.

A hagyományos forgatókönyv szerint, ha a statisztikai adat iránt megfogalmazott igényt elfogadták, akkor a hivatal új adatfelvételt indít. Az eddigiekben a minőségügyi keretrendszer a KSH saját adatgyűjtéseire vonatkozóan mutattuk be. A nemválaszolás fokozódásával, illetve a statisztikai szervezetek erőforrásainak csökkenésével a figyelem előterébe kerültek az adatszolgáltatói terhek és a statisztika költségei. Az adattárolás és kezelés területén végbement technikai fejlődés is új lehetőségeket teremtett. Mindezek következtében a mások által, más célokból gyűjtött adatok átvétele, újrahasznosítása egyre nagyobb jelentőséget kap. Az adatok többcélú hasznosítása a statisztika jövőképeinek kulcselemévé vált. Az adatok többcélú hasznosításával előtérbe kerülnek a különböző típusú adatforrások statisztikai célú felhasználásával és a szakstatisztikák konzisztenciájával kapcsolatos kérdések. A KSH érdekelt ezeknek a nem statisztikai céllal, illetve nem a KSH statisztikai céljaival gyűjtött, vagyis másodlagos adatok statisztikai felhasználásában, melyek többsége jelenleg adminisztratív adat. A hivatal tervezi a big data típusú adatok jövőbeli kiaknázását is³⁴. Az adatáramlás, adatkapcsolás egyre nagyobb szerepet kap az adat-előállítási folyamatban. Megjegyzendő, hogy a másodlagos adatforrások minőségi kérdései nem különböznek alapvetően az elsődleges adatforrásokra vonatkozóktól, ezért érdemes ezeket közös rendszerben, integráltan kezelni. Az előzőekben ismertetett minőségügyi keretrendszer alkalmas erre.

A megközelítés általános alapja a teljes körű minőségirányítás (TQM), melynek alapelvei: vevőorientáltság, a vezetés elkötelezettsége, a munkatársak bevonása, folyamatközpontú irányítási rendszer, rendszerszemléletű irányítás, folyamatos fejlesztés, tényeken alapuló döntéshozatal és a beszállítókkal való hosszú távú kapcsolat.

A minőségügyi keretrendszer működése is ezeken az elveken alapul. A másodlagos adatforrásokat illetően ez azt jelenti, hogy ezek átvételekor a minőségükről megfelelő információval kell rendelkezünk (tényeken alapuló döntések), és az adatgazdákkal hosszú távú kapcsolatot (beszállítókkal való hosszú távú kapcsolat) kell kiépíteni.

2. A másodlagos adatforrásokra vonatkozó minőségi irányelvek az ESR-ben

Az ESR eddigi minőségügyi dokumentumai a másodlagos adatforrások közül csak az adminisztratív adatokat nevesítik, azaz csak statisztikai és adminisztratív forrásokat különböztetnek meg. Az Európai Statisztikai Rendszerhez tartozó minden statisztikára ugyanazok a minőségi elvárások, alapelvek érvényesek. Speciális kezelést igényel azonban az az

³⁴ KSH stratégia 2020 http://www.ksh.hu/docs/bemutakozas/hun/strategia_2020.pdf

adat, amit eredetileg nem statisztikai, hanem adminisztratív célra hoztak létre. Az Európai Statisztika Gyakorlati Kódexe tartalmazza az adminisztratív adatforrásokkal, azok kezelésével szemben támasztott követelményeket. A 2005-ben nyilvánosságra hozott kódexet felülvizsgálták, a jelenleg érvényes változatot 2011. szeptember 28-án publikálták. Ezzel egyidejűleg, alkalmazásának támogatása érdekében megalkottak egy a Gyakorlati Kódexet kiegészítő dokumentumot, az Európai Statisztikai Rendszer minőségbiztosítási keretrendszere (ESS QAF)³⁵ címmel. Az EU hivatalos statisztikájára érvényes minőségbiztosítási keretrendszer a Gyakorlati Kódex 4.–15. elvének végrehajtását elősegítő tevékenységeket, módszereket és eszközöket írja le, az elvek minden ismérvére vonatkozóan. A Gyakorlati Kódexben az adminisztratív adatforrásokra vonatkozó ismerveket a 2., 8., 9. és 10. elvek tartalmazzák.

A 2. elv az adatgyűjtésre való felhatalmazást fogalmazza meg. Ennek alapján a statisztikai szervezeteknek jogszabályi felhatalmazással kell rendelkezniük, hogy az adminisztratív nyilvántartások adataihoz statisztikai célra hozzáférjenek (2.2 ismerv).

A 8. elv foglalkozik a megfelelő statisztikai eljárásokkal. Az ide tartozó ismervek az adminisztratív adatgazdák és a statisztikai szervezetek megállapodáson alapuló együttműködését várják el, ami már az adminisztratív adatállományok kialakításakor is figyelembe veszi a statisztikai igényeket, és folyamatos, a statisztikai fogalmakat jól közelítő adminisztratív fogalmakat és biztonságos adatkapcsolatot eredményez (8.1, 8.7–8.9 ismérvek). Az ESS QAF a megvalósításhoz a statisztikai szervezetek – egymás közötti együttműködésen alapuló – feladataként írja le az adminisztratív adatok felhasználhatóságának feltérképezését, statisztikai feldolgozásának felelősségét, a jogszabályi környezet fejlesztésének figyelemmel kísérését, és az adminisztratív adatok gazdáival való együttműködést a megfelelő minőségű adatok biztonságos szolgáltatása érdekében. Termék/adatfelvétel szinten rendszeres egyeztetésen alapulva pedig az adminisztratív és statisztikai fogalmak, folyamatok dokumentálása, elemzése, fejlesztése, a különbségek kezelése járulhat hozzá az elvárások teljesüléséhez.

A 9. elv az adatszolgáltatói terhekkal foglalkozik. A terhek mérséklése érdekében a statisztikai szervezeteknek törekedniük kell az adminisztratív vagy más statisztikai adatforrások, illetve olyan eljárások használatára, melyek lehetővé teszik ezeknek az adatforrásoknak az összekapcsolását (9.4.–9.6. ismérvek). E célok megvalósítása érdekében az ESS QAF intézményi feladatként jelöli meg az európai együttműködési hálózatok eszközeinek fejlesztését, és az érintettek közötti, a koordinált modulok kialakítását szolgáló nemzeti szintű együttműködést. Termékszinten ezt segítik az adminisztratív források elérhetőségéről és jellemzőiről szóló információk, illetve az átvételt támogató alkalmazás. A nemzeti statisztikai rendszeren belül a cél az egymás közti adatforgalmat és az adatbázisok közös használatát támogató eszközök fejlesztése, valamint a különböző adatforrások összekapcsolását biztosító kulcsváltók összehangolt alkalmazása. Ezen az úton az egyik első lépés az adatállományok struktúrájának és átviteli formátumának, valamint a termelési és archív adatok tárhelyének dokumentációja.

³⁵ Quality Assurance Framework of the European Statistical System Version 1.2

<http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>

A 10. elv az erőforrások hatékony felhasználásáról szól. A költséghatékonyság érdekében lépéseket kell tenni az adminisztratív nyilvántartások használhatóságának javítására, hogy a közvetlen adatfelvételekre fordított erőforrásokat a lehető legalacsonyabb szinten tartsuk.

Érdeklődő olvasóknak ajánljuk az Európai Statisztikai Rendszer minőségbiztosítási keretrendszerének ide vonatkozó részletes útmutatásait, továbbá az ott hivatkozott tagállami jó gyakorlatok tanulmányozását. Röviden összefoglalva az adminisztratív adatforrásokkal kapcsolatos elvárások – szem előtt tartva az erőforrásokkal való takarékoság és az adatszolgáltatói terhek csökkentésének igényeit – a jogi környezetre, az adatgazdákkal való együttműködésre, az adatok használhatóságára (fogalmak és kapcsolási lehetőségek) és általában minőségére vonatkoznak.

Ugyanakkor a statisztikai rendszeren belül elvárás a többcélú adathasznosítás, amit az adatkapcsolást biztosító kulcsváltozókkal és technikai eszközökkel lehet megvalósítani.

Az adminisztratív adatok vagy statisztikai adatok másodlagos használatának módszertani kérdéseivel itt nem foglalkozunk.

A Gyakorlati Kódex 4. elvének megfelelően a statisztikai adatok minőségét rendszeresen mérni kell, és minőségjelentésben a felhasználók rendelkezésére kell bocsátani.

A minőségjelentések ESR kézikönyvének új kiadása³⁶ (2014) a Gyakorlati Kódex ajánlásaival összhangban részletesen tárgyalja a statisztikai folyamatokra és adatokra vonatkozó minőségjelentések tartalmát, összeállításának módszereit. A kézikönyv az általános, minden statisztikai folyamatra érvényes tartalom mellett a statisztikai folyamatok hat típusát különbözteti meg, amelyekhez esetenként speciális tartalmat javasol. Ezek közé tartozik az adminisztratív adatforrásokat felhasználó statisztikai folyamat, valamint a többféle adatforrást felhasználó statisztikai folyamat. Speciális tartalom mindössze két minőség-összetevő esetében ajánlott az ilyen típusú folyamatokhoz. A relevancia jellemzésénél ki kell térni az átvett másodlagos adatforrás fogalmainak ismertetésére, és minden esetben részletezni kell, hogy azok mennyiben térnek el a célzott statisztikai fogalmaktól, illetve milyen lépések (modellszámítások) történtek a közelítés érdekében. Ami a pontosságot illeti, a kézikönyv – adminisztratív nyilvántartások alkalmazása esetén – kiemelten ajánlja a lefedettség hiábára, a nemválaszolási hiábára, a fogalmak eltéréseire és a nyilvántartás változóinak hiábára vonatkozó jellemzést. Az eseményjelző rendszereknél a nem jelentett események rátája jellemzi a lefedettség hiányt. Speciális gond, hogy gyakran a nem jelentett eseményeket kizáró munkadefiníciót használnak³⁷, ami már relevanciát érintő probléma. A több adatforrást használó statisztikáknál a pontosságot jellemző speciális mutató a kapcsolható rekordok száma és aránya. Ugyancsak ebben az esetben feltételezhető, hogy a különböző adatforrások különböző időpontokban frissülnek, így a revíziók jellemzése is ajánlott. Az összes többi összetevő esetében az általános jellemzők alkalmazhatók.

Érdeklődő olvasóknak ajánljuk az ESS Handbook for Quality Reports 2014-ben megjelentetett kiadásának tanulmányozását, mert részletes magyarázatokat, jó példákat és további hasznos anyagokra való hivatkozásokat is tartalmaz. Más oldalról pedig az előző kiadáshoz képest újszerű, mert már figyelembe veszi a Sponsorship on Quality munkacsoport standard indikátorokra vonatkozó ajánlásait, és hivatkozik az időközben kidolgozásra került

³⁶ ESS Handbook for Quality Reports, 2014. <http://ec.europa.eu/eurostat/documents/64157/4373903/01-ESS-Handbook-for-Quality-Reports-2014.pdf/d6152567-a007-4949-a169-251e0ac7c655>

³⁷ Például, ha a nyilvántartás egységét úgy definiáljuk, hogy „a hatóság tudomására jutott” esemény.

metaadat-struktúrákra, az ESQRS-re és a SIMS-re³⁸ is, ami a minőségre vonatkozó információkat is tartalmazó standard metaadat-szerkezet. A 2016 elején elfogadott SIMS ver2³⁹ és ezzel összhangban az ESQRS ver2 a korábbi változathoz képest finomhangolásnak tekinthető, attól csak apró részletekben különbözik.

Távolabb tekintve hasznos az ENSZ vonatkozó rendszerének ismerete is, mivel hazánk az ENSZ egyik tagállama. A 2012-ben elfogadott, azóta elérhető komplex rendszer az „Irányelvek az ENSZ általános nemzeti minőségbiztosítási keretrendszeréhez” (NQAF)⁴⁰, melyhez fogalomtár és dokumentációs séma is tartozik. Szellemiségében hasonló az ESR rendszeréhez, mivel az adatszolgáltatói terhek növelésének elkerülése és adatminőség javítása érdekében hasznosítható adminisztratív adatforrások felhasználását javasolja. A legfontosabb stratégia az adminisztratív adatok használatának kockázatkezelésében a forrásadatok gazdáival való folyamatos kapcsolat fenntartása.

3. A másodlagos adatforrásokra vonatkozó alapelvek érvényesítése a KSH-ban

A másodlagos adatforrások minőségének rendszerszerű megközelítése azt igényli, hogy beazonosítsuk és nyilvántartásba vegyük ezeket. Ezért röviden ismertetjük a másodlagos adatforrások nyilvántartási rendszerét, amelynek kialakításakor tekintettel kellett lenni az ESR adminisztratív adatforrásokkal kapcsolatos, az előző részben ismertetett minőségi standardjaira. A KSH a gyakorlatban kihasználja a más adatforrásból átvett (eredetileg statisztikai célra vagy adminisztratív célra gyűjtött) adatok kezelésének hasonlóságait, és a másodlagos adatokat egységes rendszereként írjuk le.

3.1. A másodlagos adatforrások nyilvántartása

Az adatforrások nyilvántartása a KSH metainformációs rendszerének része.

A KSH által használt adatforrások a KSH honlapján az „Adatok” főmenü „Módszertani információ (metaadatok)” menüből közvetlenül elérhetők⁴¹:

- statisztikai adatgyűjtések,
- adminisztratív adatátvételek,
- nem adminisztratív adatátvételek,
- regiszterek, nyilvántartások.

Ezek a metainformációk a KSH-ban használt különböző típusú adatforrások metaadatait tartalmazó egységes adatbázisból származnak.

Az alábbi ábra mutatja a KSH honlapján is elérhető metarendszer szerkezetét.

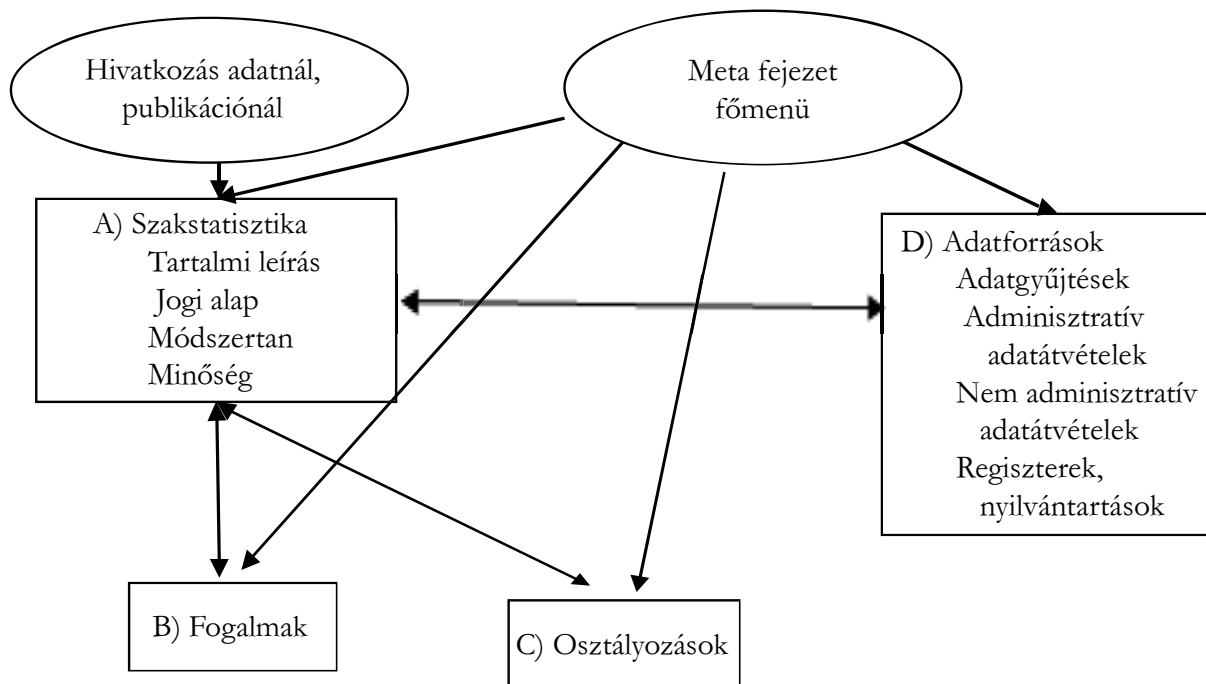
³⁸ ESQRS: European Statistical Quality Report Structure; SIMS: Single Integrated Metadata Structure.

³⁹ <http://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf/47c0b80d-0e19-4777-8f9e-28f89f82ce18>

⁴⁰ Guidelines for the Template for a Generic National Quality Assurance Framework.
<http://unstats.un.org/unsd/dnss/docs-nqaf/GUIDELINES%208%20Feb%202012.pdf>

⁴¹ Az adatforrások a felhasználó szakstatisztikák szerint csoportosítva is megjelennek.

Adatforrások sematikus ábrája a szakstatisztikák módszertani dokumentációs rendszerében (meta)



Forrás: Baracza Lajosné, KSH.

Az adminisztratív és nem adminisztratív adatátvételekből származó adatforrásokról a metainformációs rendszerben a honlapról elérhető adatok (tárgyévenként):

- adatforrás azonosítója és megnevezése,
- adatátadó szerv,
- adatátvétel típusa (adminisztratív adat – nem adminisztratív adat),
- gyakorisága,
- tartalma.

A szakstatisztikák dokumentációjának évi rendszerességgel történő aktualizálása során az adatforrásokra vonatkozó adatok is frissítésre kerülnek.

A KSH metainformációs rendszere sokkal összetettebb és az adatátvételekre vonatkozó információk köre is bővebb annál, mint ami a honlapon elérhető. A rendszer átfogó ismertetését lásd a 4. mellékletben.

6.3.2. A másodlagos adatforrások hazai minőségi kérdései

A másodlagos adatforrások minőségével kapcsolatos mérés, ellenőrzés a keretrendszer logikájának megfelelően több szakaszban valósul meg. Az első fázis a másodlagos adatok minőségmérése adatátvételkor, abban az állapotban, ahogy bekerül a KSH statisztikai rendszerébe (A). A második fázis a folyamatok minőségére vonatkozik, ennek része az adatátvétel folyamatának minősége és a további feldolgozási lépések minősége már a statisztikai folyamatra vonatkozó minőségi irányelveknek megfelelően (B). Végül, az utolsó fázis a másodlagos adatok felhasználásával készült statisztikai adatok minőségének mérése, jellemzése (C).

A. A másodlagos adatok minőségmérése átvételkor

Az előző fejezetekben láttuk, hogy másodlagos adatforrásokat a statisztikai munkafolyamat szinte valamennyi fázisában alkalmazhatunk. A másodlagos adatforrások minőségére vonatkozó megfontolások általános érvényűek, csak részben függenek a konkrét munkaszakasztól, felhasználási céltól. Az adat-előállítási folyamatban a másodlagos adatforrásból származó adatokat akkor tudjuk jól felhasználni, ha azok jó minőségűek, azaz az adatokra, folyamatra, szervezetre vonatkozó ismeretek rendelkezésre állnak, és azok jó minőséget tanúsítanak.

Két speciális probléma adódik: az adminisztratív adatokat eredetileg nem statisztikai, a mások által gyűjtött statisztikai adatokat nem a konkrét statisztikai célra gyűjtötték, továbbá az adatgyűjtés és az adatfeldolgozás egyes lépései más szervezetekben történtek. Felhasználáskor információkkal kell rendelkezünk az adatgazdáról, az adatgyűjtés és eddigi feldolgozás módszereiről és minőségéről, az adatok minőségi jellemzőiről. Az adatgazda ezeket az információkat jó esetben metaadatok formájában az adatokkal együtt átadja. Ezeket a felhasználó saját céljai szempontjából értékeli, valamint további vizsgálatok eredményeivel kiegészíti.

A másodlagos forrásból származó adatokra vonatkozóan is ugyanazok az elvek érvényesek, mint a közvetlen statisztikaiaknál, azaz a felhasználó statisztikai szervezetnek rendelkeznie kell minden olyan információval, ami alapján eldöntheti, hogy adott adatállomány alkalmas-e az adott felhasználási célra. Ez a minőségjelentés és minőségértékelés rendeltetése. Mivel az adatgazda saját feladataival összefüggésben állította elő az adatokat, az adatátvevő szervezet statisztikai céljait kellő mélységben nem ismeri, és nem mindig érdeke az adatok statisztikai célú újrahasonosítása, jellemzően a felhasználó statisztikai szervezetnek kell a másodlagos forrásból származó adatokra előállítani a minőségjelentést.

A fentiek alapján adódik, hogy a minőségjelentés (9. ábra) tartalma ki kell, hogy terjedjen az adatforrásra (adatgazda szervezet, adatgyűjtés célja stb.), az adatkezelés átadást megelőző lépéseire (ellenőrzések, javítások stb.), az átvett adatok minőségére (a szokásos minőségösszetevők mentén). Ezek az adott adatállomány általános jellemzői, amit elég, ha egy adatállományról egy felhasználó összeállít. Adott állomány több célra is használható. Különböző célokra különböző adatminőségi szempontok fontosak. Ezért az általános minőségjelentést felhasználáspecifikus részekkel célszerű kiegészíteni. Ezek egyrészt az adott célok szempontjából fontos minőségvizsgálatok eredményeit, illetve valamennyi jellemző adott szempontú értékelését jelenti. Például, ha egy másodlagos adatállomány csak egy szűk körre vonatkozik, igaz, azokra teljeskörűen, akkor egy statisztikai regiszter adatforrásaként nem nagyon hasznos, de ha ebben a csoportban nagy a statisztikai nemválaszolás, akkor imputáláshoz még igen hasznos lehet. Imputálásnál viszont a kapcsolhatóság, a kapcsolási és imputáláshoz felhasznált változók megfeleltethetősége, konzisztenciája elsődleges például a pontossággal szemben.

Másodlagos adatforrásból származó adatok minőségmérése, értékelése – szerkezet

	Adatforrás (A)	Folyamat (meta) (B)	Adatok (C)				
Átfogó, általános							
Speciális, adott felhasználási cél szempontjából			1. felh.	2. felh.	3. felh.

Az adatforrás minőségjelentésének, minőségértékelésének belső szerkezete különböző lehet, attól függően, mi látszik célszerűnek. Fenti struktúra egy lehetőség. Szakirodalomból ismeretes még több különböző modell. Daas et al.⁴² az adatforrás minőségének jellemzésére öt hiperdimenziót azonosít, ezek: adatgazda, relevancia, adatvédelem és biztonság, átadás, eljárások.

Egy másik megközelítés Peter van Nederpelt és Piet Daas műve⁴³, amely nem annyira a minőséget, hanem a minőséget befolyásoló tényezőket (49 tételt) veszi számba, ezeket jellemzi.

Ha mélyebben megnézzük e fent idézett két megközelítés indikátorait, azt találjuk, hogy az ezek alá besorolt indikátorok akár a fenti struktúrába is átrendezhetők lennének. Konkrét esetben célszerű az egyik modell mellett dönteni, de más modellt is átnézni, nehogy lényeges szempontok kimaradjanak az értékelésből.

Az értékelés egyrészt hatással van a további statisztikai feldolgozási folyamatra, másrészt visszacsatolható az adminisztratív adat tulajdonosának. Az azonban, hogy a visszacsatolás, folyamatos jobbítás megvalósuljon, speciális szempontokat is figyelembe kell venni. Az egyik ilyen, hogy a minőségi elvárások, esetleg eljárásbeli, jobbítást szolgáló intézkedések, jogosultságok az együttműködési megállapodásban vagy az abban foglalt felhatalmazással tehetők. A másik specialitás, hogy az adatvédelmi szempontokat tiszteletben kell tartani, azaz az egyedi adatok tekintetében nem kerülhet sor felfedésre statisztikai forrásból származó adatok alapján. Egy lehetőség, hogy csak átfogó, sokaságra vonatkozó ellenőrzési, javítási javaslatokat lehet megfogalmazni.

Minőségértékelés a 2010. évi kérdőív alapján

A KSH-ban 2010-ben egy munkacsoport kidolgozott egy minőségértékelő kérdőívet, elsősorban a Daas et al. cikkben ismertetett minta alapján. A kérdőív fejezetei:

- a kitöltő,
- adatforrás (szervezet, kapcsolat, átadás),
- metaadatok (azonosítás, adatkezelés az adatgazdánál),
- adatminőség (minőségi összetevők mentén, kiemelve a megfeleltethetőség és kapcsolhatóság tulajdonságait).

A kérdőív 17 oldal terjedelmű és 38 számozott kérdést tartalmaz űrlap formátumban.

A kérdőívet valamennyi ismert adatfelhasználó megkapta, 155 esetben töltötte ki a felelős az űrlapot.

⁴² Daas, Piet – Ossen, Saskia – Arends-Tóth, Judit: Framework of Quality Assurance for Administrative Data sources.

⁴³ Peter van Nederpelt és Piet Daas: 49 Factors that influence the Quality of Secondary Data Sources. Statistics Netherlands, The Hague 2012.

<https://unstats.un.org/unsd/dns/docsnqaf/49%20Factors%20that%20influence%20the%20Quality%20of%20Secondary%20Data%20Sources.pdf>

Az így összegyűjtött információkból adatátadó adminisztratív adatgazdánként értékelő összesítés készült javaslatokkal. Az általánosan kirajzolódó pozitív kép számos olyan részletet tartalmaz, ami javítási lehetőségeket fed fel. Ilyen például az együttműködési megállapodás hiánya, hozzá nem férhető volta, részlegessége (ugyanattól a szervezettől több adatátvétel esetén csak egyesekre vonatkozik), tartalma (minőségi elvárások).

A 2010-es megállapítások szerint alig több mint minden negyedik adminisztratív adatátvételben lehet egyedi rekordokat azonosítani, és 35%-uk csak publikálási céllal lett átvéve. Így nem lehet tudni, hogy az általános megelégedettség (89%) mögött a részletes ellenőrzés lehetőségének hiánya bújlik-e meg.

B. A másodlagos adatok átvételének és feldolgozásának minősége.

A másodlagos adatok speciális munkaszakasza az adatátvétel. A VIII. fejezet mutatja be, hogy a KSH-ban az utóbbi évek fejlesztései új lehetőségeket teremtettek a nem kérdőív formátumú adatállományok központi fogadására és továbbítására (KARÁT-rendszer). A KSH és az adatgazdák közötti kapcsolat informatikai rendszereibe már tervezéskor beépítésre kerültek az átvételt jellemző minőségindikátorok. Ezek az átvétel dokumentáltságát, összetettségét, végrehajtás pontosságát és időszerűségét jellemzik. Egy külön indikátorcsoport az adatgazdák adatszolgáltatói terheit jellemzi. A KARÁT-rendszer indikátorlistáját az 5. sz. melléklet tartalmazza.

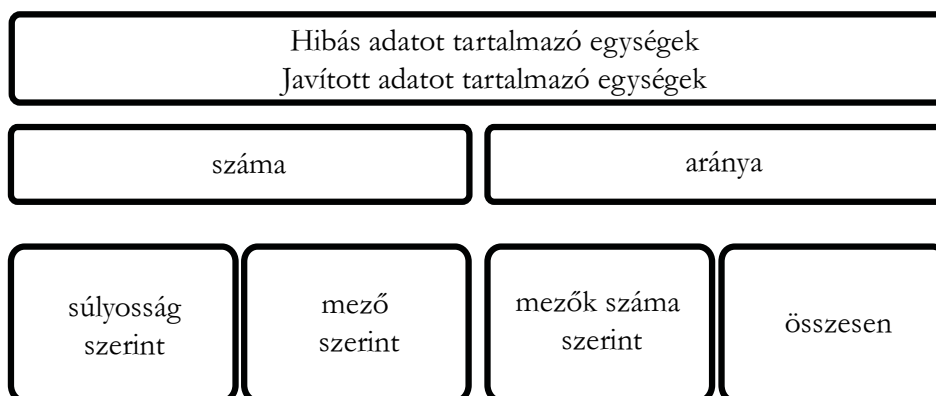
Az érkeztetés után a másodlagos adat egy ellenőrzési folyamaton megy át.

Az ADAMES-rendszer a KSH-ba nem kérdőív formátumban érkező másodlagos adatok feldolgozását megelőző előkészítést végzi, ideértve az elemzői, ellenőrzési és javítási feladatokat is.

Az ADAMES-rendszer keretében a 10. ábrán szereplő elemek kombinációjaként értelmezhető minőségindikátorok számíthatók.

10. ábra

ADAMES-rendszerben számított minőségindikátorok elemei



Bár ezek az indikátorok a bejövő adatok minőségét jellemzik, így az „A” kategóriába is sorolhatók lennének, tekintettel azonban arra, hogy a minőségellenőrzés már a konkrét felhasználási célhoz kapcsolódik, az adatfeldolgozás során számítandók. Ezek az indikátorok

alapul szolgálhatnak a 9. ábrán szereplő minőségjelentés felhasználásspecifikus részének kitöltéséhez.

A munkafolyamat a továbbiakban nem különül el a normál statisztikai munkafolyamattól, így az általános elvárásoknak kell megfelelnie.

A KSH 2014-ben elkészítette és nyilvánosságra hozta a minőségi irányelvek⁴⁴ aktualizált változatát. Ennek a dokumentumnak az a célja, hogy a statisztikai folyamatszakaszokkal (és részfolyamatokkal) szemben általános elvárásokat és ajánlásokat fogalmazzon meg a KSH dolgozói számára ahhoz, hogy a folyamat, végső soron pedig az output minél jobb minőségű legyen. A másodlagos adatokat illetően ez a dokumentum tartalmilag a statisztikai munkafolyamat egyes szakaszaiban a Gyakorlati Kódexben megfogalmazott elvárások helyi alkalmazási feltételeit tartalmazza.

C. A másodlagos adatok felhasználásával készült statisztikák minősége

Minden elkészített statisztika minőségét publikáláskor (adatkiadáskor) mérni, és minőségjelentésben dokumentálni kell, továbbá ezt az adatokkal együtt a felhasználó számára elérhetővé kell tenni. Mint azt már említettük, sokszor az előállított adat minőségét az adat-előállítási folyamat segítségével jellemezzük, közvetlen mérési lehetőség hiányában. Mind a KSH minőségjelentése, mind az európai standard minőségjelentés sémája tartalmaz olyan indikátorokat, amelyek specifikusak a másodlagos adatokat is felhasználó folyamatokra. Ez a terület jelenleg fejlesztés alatt áll, az ESR folyamatban lévő ADMIN projektjének⁴⁵ egyik nevesített célja a másodlagos adatok legnagyobb csoportját: az adminisztratív adatokat is használó statisztikák minőségjelentésének kidolgozása.

A KSH metainformációs rendszerében (4. melléklet) az egyes szakstatisztikák leírása tartalmazza a legfontosabb minőségre vonatkozó információkat. Egyelőre azonban nincs gyakorlat arra vonatkozóan, hogy a KSH honlapon is megjelenő szakstatisztikai dokumentáció részeként szerepeljen a felhasznált másodlagos adatok minőségével kapcsolatos információ. Az erre vonatkozó fejlesztés folyamatban van.

4. Várható jövőbeni feladatok

Az ESR bizottsága időközönként ún. szakmai felülvizsgálatot, peer review-t hajt végre annak vizsgálatára, hogy a nemzeti statisztikai hivatalok és az ESR egésze milyen mértékben felel meg az Európai Statisztika Gyakorlati Kódexében foglalt elveknek. A 2013–2015-ben lefolytatott ellenőrzések egyik kiemelt szempontja volt az adminisztratív adatok hasznosulása a hivatalos statisztikák előállításában. A magyar statisztikai rendszer felülvizsgálata nyomán megfogalmazott ajánlások meghatározóak a KSH rövid és középtávú fejlesztései tekintetében. Az adminisztratív adatokkal összefüggő ajánlás célul tűzi ki, hogy az illetékes magyar hatóságok a KSH-val együttműködve készítsék el a magyar közigazgatás nyilvántartásainak a regiszterét. Ennek alapján a KSH-nak értékelnie kell mindegyik adminisztratív adatállományt, mint a hivatalos statisztikák lehetséges adatforrásait, és megállapítani, hogy a továbbiakban hogyan hasznosíthatók.

⁴⁴ Minőségi irányelvek a Központi Statisztikai Hivatal statisztikai folyamataira.

http://www.ksh.hu/docs/bemutakozas/hun/ksh_minosegi_iranyelvek_2014.pdf

⁴⁵ <http://ec.europa.eu/eurostat/documents/42577/6906243/ADMIN+fact+sheet.pdf/cbb590b2-9d6f-439c-af2d-ca8b5e9cf1f7>

A fentiek teljesítése céljából ki kell dolgozni az adminisztratív adatforrások statisztikai célú hasznosításának felméréséhez szükséges módszertani háttérrel, majd az elkészült módszertan alapján el kell végezni a felmérést, és az eredményről összefoglaló jelentést kell készíteni.

További feladatot jelent az adminisztratív adatok minőségértékeléseinek rendszerszerűvé tétele. Ehhez ki kell dolgozni a szükséges informatikai és metainformációs háttérrel is. Gondoskodni kell a minőségjelentések belső nyilvánosságáról is.

VII. Együttműködés az adatgazdákkal

1. Az együttműködés „tétje” és mozgatórugói

A KSH folyamatosan szélesíti a magyar közigazgatás adminisztratív rendszereivel kapcsolatos ismeretanyagát, és aktívan alakít ki együttműködést azoknak az adatrendszereknek a gazdáival, ahol látja a statisztikai hasznosítás lehetőségét. Egyelőre azonban a nemzeti adatvagyon jelentős részét alkotó adminisztratív nyilvántartások statisztikai célú hasznosítása elmarad nemcsak a kívánatostól, hanem a lehetőségektől is, és ennek több oka is van. Fontos körülmény, hogy a közelmúltig az e cél megvalósulását elősegítő európai és hazai rendeletek és jogszabályok sem hozták meg e tekintetben a remélt eredményt. Mint arról a IV. fejezet részletesen beszámol, az európai statisztikákról szóló 223/2009/EK európai rendelet 2015. áprilisi módosítása szerint az európai statisztikát előállító intézmények jogosultak a közigazgatási nyilvántartásokhoz való azonnali és ingyenes hozzáférésre és azok felhasználására. Adatok átadásakor csatolni kell a metaadatokat is. A rendelet egyúttal 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.

Ez arra enged következtetni, hogy nem csupán az érdekeltséget segítő törvényi háttérrel kell biztosítani, hanem az együttműködés kultúráját is jóval magasabb szintre kell emelni. Az adminisztratív adatforrásokkal kapcsolatban kialakított rendszerszerű együttműködés az adatgazdákkal, illetve a KSH szerepének megerősítése ezekben a kapcsolatokban hozzájárulhat ahhoz, hogy az igazgatási forrásokból származó statisztikai adatok minősége megfelelő legyen, és javuljon a költséghatékonyság is. Az Eurostatnak a Gyakorlati Kódexben lefektetett elvek hazai megvalósulását vizsgáló 2015-ös jelentése⁴⁶ is rámutat arra, hogy a hatályos statisztikai törvény nem rendezi a KSH-nak az adminisztratív nyilvántartások megtervezésében, illetve – annak érdekében, hogy statisztikai célokra jobban felhasználhatók, megbízhatóbbak legyenek – az ezek kialakításában való részvételére vonatkozó jogosultságát.

Az újrahazsnosítás szempontjából kiemelten fontos a 2020 körüli népszámlálás, amelynek lebonyolításakor, más országok gyakorlatához hasonlóan, célszerű lenne a rendelkezésre álló nyilvántartásoknak a korábbihoz képest kiterjedtebb felhasználása (lásd V. fejezet).

⁴⁶ Jelentés a magyar statisztika európai uniós auditjáról.

A szakmai integráltság hiányosságait néhány, az alábbiakban részletezett „berögült” mechanizmus tartja fenn.

1.1. Az adminisztratív adatgyűjtések jogi szabályozása

Az adminisztratív adatállományok rendszerszerű statisztikai célú hasznosítása a legtöbb európai országban az utóbbi tizenöt év során merült fel, és az adatforrások körének ilyen módon történő bővítését a KSH is csak 2004-ben emelte be a stratégiai törekvései közé. A nem statisztikai céllal létrehozott nyilvántartások a különböző tárcák és főhatóságok (esetleg gazdasági szervezetek) kezelésében vannak, felhasználásukat, illetve adatgyűjtési rendszereiket különféle, egymással nem harmonizáló ágazati jogszabályok rendezik. Ezek a jogszabályok „történelmi” okok miatt többnyire nem vették/vehették figyelembe az adatok újrahasznosításának érdekeit, hiszen ilyen elvárás velük szemben korábban nem fogalmazódott meg.

Az adminisztratív adatállományok intézményes keretek közötti harmonizálásának igénye hazai kormányzati szinten az állami és önkormányzati nyilvántartások együttműködésének általános szabályairól szóló 2013. évi CCXX. törvényben (Interoptv.) fogalmazódott meg, és ennek értelmében az e-közigazgatásért felelős belügyminiszter lett a feladat gazdája.

A statisztikai törvény e téren a KSH részére eddig is biztosított ugyan bizonyos, a módszertani fejlesztéssel és koordinációval kapcsolatos jogosítványokat, de azok érvényesítése továbbra is a partnerek együttműködési hajlandóságától függ. Számottevő változást a módosított európai statisztikai rendelet és a magyar jogszabályok ahhoz való igazítása hozhat. Ennek keretében lehetőség nyílna arra, hogy – összhangban az Európai Statisztika Gyakorlati Kódexe 8.7–8.9 és 10.3 ismérveivel – a KSH erős jogosítványokat kapjon az adminisztratív adatforrásokhoz való statisztikai célú hozzáférésben, ideértve azt is, hogy a nyilvántartások kialakítása, fejlesztése, megszüntetése, illetve minőségbiztosítása csak a KSH egyetértésével történhet. Másrészt az adminisztratív források tulajdonosai számára is előírható lenne, hogy ezeket a tevékenységeket kizárólag a KSH egyetértésével végezhetik.

1.2. Az adminisztratív adatgyűjtésekkel kapcsolatos információhiány

Az adminisztratív regiszterek használhatóságával és megbízhatóságával kapcsolatos problémák csak kisebb részben fakadnak az adminisztráció működési nehézségeiből, nagyjából az a probléma, hogy az együttműködés kultúrája e téren egyelőre még fejlődőben van, és az adatgazdák nem minden esetben tudnak különbséget tenni adataik statisztikai és adminisztratív célú továbbhasznosítása között. Ez bizonytalansághoz vezet a kooperáció tartalmát illetően.

Az Interoptv. létrehozza a nyilvántartások regiszterét⁴⁷, ami a közfeladatot ellátó szervek által közfeladatuk ellátása során vagy azzal összefüggésben jogszabály által szabályozott nyilvántartásokra vonatkozóan tartalmazza a nyilvántartás és a nyilvántartó megnevezését, a rendelkező jogszabályra hivatkozást és a nyilvántartott adatok megnevezését. Egy ilyen regiszter alapul szolgál ahhoz, hogy a nyilvántartások rendszere koherensebbé váljon, továbbá jelentősen megkönnyítheti a statisztikai újrahasznosítást is. A hatályos jogszabályok szerint azonban a közadatvagyon teljes körű felmérése és a közadatkataszter elkészítése csak 2017 első negyed-évében várható.

⁴⁷ A 2015. évi CCXXII. törvény szerint „információforrások regisztere”.

Az audit jelentés is felhívja a figyelmet, hogy szükség van a magyar közigazgatás adminisztratív nyilvántartásainak leltárára, és annak közzétételére. E leltár alapján a KSH meg tudja vizsgálni, hogy mely nyilvántartások alkalmasak hivatalos statisztikai célú felhasználásra, és megállapításai közzétételével hozzájárulhat a nyilvántartások minőségének javításához.

A használhatóság szempontjából további komoly korlátot jelent egy egységes leíró rendszer hiánya, amely minden érdeklődőt tájékoztatna az adatállományok tartalmáról, és legfontosabb jellemzőiről. Az Interoptv. a regiszter mellett létrehozta az ún. adatmegnevezések jegyzékét⁴⁸ is, mely tartalmazza a nyilvántartások együttműködésre való alkalmasságának biztosítása szempontjából jelentőséggel bíró ismerveket, azok megnevezését és értelmezését. Tekintettel azonban arra, hogy a törvény szabályainak bevezetése fokozatosan történik, és a jegyzék a nyilvántartások közötti adatkapcsolati igényekkel párhuzamosan fog épülni, ezért használható közös metarendszer csak több év múlva áll majd rendelkezésre.

1.3. Szervezeti hiányosságok

Az adminisztratív adatok statisztikai hasznosításában érintett adatgazdák és a hivatalos statisztikát előállító szervezetek összessége között nem alakult ki a szükséges szakmai érdekazonosság. Nincs olyan mechanizmus vagy fórum, ahol a különböző érdekeket fontolóra vennék annak érdekében, hogy a Gyakorlati Kódex értékrendjének betartásával közösen állítsák elő a statisztikai termékeket. Az audit jelentés szerint az a tény, hogy egyre több adminisztratív forrás keletkezik, és ezzel összefüggésben a hivatalos statisztikák egyre kevésbé támaszkodnak elsődleges adatforrásként a mintavételen alapuló statisztikai felmérésekre, azt igényli, hogy ne csak egy központi hivatal létezzen, hanem jöjjön létre a nemzeti statisztikai hatóságok hatékony statisztikai hálózata. Mindezek a jelenleginél sokkal hangsúlyosabbá teszik az adminisztratív adatok statisztikai célú hasznosításában érintett szervezetek együttműködését és a KSH irányító, koordináló szerepét. A hivatalos statisztika hálózatán belül a statisztikai rendszer tagjaiból álló, az Országos Statisztikai Tanáccsal (OST) szorosan együttműködő koordináló szervezet foglalkozna egyebek mellett az adminisztratív adatforrások felhasználását érintő kérdésekkel is.

2. Az együttműködés során szerzett hazai tapasztalatok

Az adminisztratív nyilvántartásokat kezelő adatgazdák közötti szervezett együttműködés az OST keretei között 2007-ben indult, amikor létrehozták az OST Regiszter Munkabizottságát.

A munkabizottság megalakításának alapvető célja az volt, hogy a hivatalos statisztikai szolgálat tagjaival megismertesse az adminisztratív adatok statisztikai célú hasznosításának koncepcióját, hozzájáruljon a közös nyelv, közös gondolkodás kialakulásához. A következő három közép- és hosszú távú feladat megvalósítását tűzték ki a bizottsági tagok⁴⁹:

- az államigazgatási nyilvántartásokkal szemben támasztott statisztikai követelmények megfogalmazása,
- javaslat a statisztikai törvény tervezett módosításához,
- a következő népszámlálásra való felkészülés.

⁴⁸ A 2015. évi CCXXII. törvény szerint „adat- és iratmegnevezések jegyzéke”.

⁴⁹ Emlékeztető az Országos Statisztikai Tanács Regiszter Munkabizottságának 2007. június 28-i üléséről. KSH.

A legutóbbi pont kapcsán fogalmazódott meg az adminisztratív adatokon alapuló népszámlálásra való felkészülés gondolata, amelynek előfeltétele volt az adminisztratív nyilvántartások statisztikai célú hasznosíthatósága, illetve ennek törvényi megalapozása.

A munkabizottság helyzetfelméréssel kezdte a tevékenységét, amelynek a „...célja az volt, hogy képet kapjunk azokról a tervekről, lehetőségekről, amelyek az államigazgatásban mutatkoznak a statisztikai rendszer korszerűsítésére, az adminisztratív adatforrások hasznosítására vonatkozóan. Meghatározásra kerültek azok a minisztériumok, országos hatáskörű szervek, amelyek adminisztratív adatgazdaként, illetve adatigénylőként megjelennek. Továbbá ismertté váltak az adatállományok minőségi jellemzői, melyek megmutatják, hogy milyen fejlesztéseket, módosításokat kell elvégezni ahhoz, hogy a már hasznosított, sőt esetenként a bevonni tervezett adatok alkalmasak/alkalmasabbak legyenek a statisztikai felhasználásra.”⁵⁰

A későbbiekben a munkabizottságban inkább csak „apró lépések” megtételére volt lehetőség, tekintve, hogy a részvétel a munkabizottságban önkéntes volt, és a tagok jobbra szakmai érdeklődésük miatt vállalták a közreműködést. Mindazonáltal jelentősen hozzájárultak az államigazgatási nyilvántartások adatainak statisztikai célra történő felhasználásáról rendelkező 2212/2007. (XI.14.) kormányhatározat 2–3. pontjának végrehajtásához.

A kormányhatározat felhívta a Hivatalos Statisztikai Szolgálat (HSSZ) szerveit, hogy vizsgálják meg az államigazgatási nyilvántartások statisztikai célú hasznosításának bővítési lehetőségeit, különös tekintettel az azokból nyerhető statisztikai célú információk körének bővítésére, a statisztikák minőségének javítására és az adatszolgáltatók terheinek csökkentésére. A rendelkezésre álló információk alapján a HSSZ szervei a KSH irányításával áttekintették az adminisztratív nyilvántartásokat, és javaslatot dolgoztak ki a kormányhatározat végrehajtása keretében elvégzendő további feladatokra, a megvalósítás módjára, az együttműködők által vállalt feladatokra és a várható költségekre, amit az OST megtárgyalt és elfogadott.

A kormányhatározat hozzájárult az adatgazda és felhasználó szervezetek közötti párbeszéd elindulásához, illetve felerősödéséhez, de költségvetési fedezet hiányában a tervezett adatátvételek csak korlátozott számban valósulhattak meg.

2010-ben az OST tagjainak újonnan történt kinevezésével a korábban működő munkabizottságok megszűntek. 2011-ben az Egyszerű Állam című középtávú kormányzati program részeként megjelent a vállalkozói adminisztratív terhek csökkentésére irányuló rövid és középtávú kormányzati programról szóló 1133/2011. (V. 2.) kormányhatározat, amelynek nyomán a KSH kapcsolódó feladatainak elvégzésére OST-munkabizottságokat hozott létre, köztük „Az adminisztratív és hatósági nyilvántartások statisztikai célú felhasználását elősegítő munkabizottságot”. Tevékenységének eredményét ajánlásban fogalmazta meg⁵¹.

Fontossága miatt a dokumentumból érdemes kiemelni azt a szempontrendszert, amely alapján a szakmai programot megfogalmazták, ezek a következők:

- „jogsabályi környezet módosítása annak megfelelően, hogy jelenleg mi segíti, és mi akadályozza a nem statisztikai adatforrások statisztikai felhasználását,

⁵⁰ Adminisztratív adatforrások statisztikai hasznosításának bővítése a hivatalos statisztikai szolgálat keretein belül. KSH, munkaanyag, 2008.

⁵¹ Ajánlás az adminisztratív és hatósági nyilvántartások statisztikai célú felhasználásának elősegítéséhez. KSH, 2011.

http://www.ksh.hu/docs/bemutakozas/hun/ost/OST_2_MKB_ajanlas.pdf

- a közös nyelv, azaz közös fogalmak, konvertálható nomenklatúrák kialakítása, fejlesztése,
- a nyilvántartások minőségének mérése,
- a folyamatos konzultáció és kommunikáció megteremtésével az adminisztratív adatok gazdái és a statisztika között és
- egyéb technikai kérdések.”

Az ajánlás ismételten napirendre tűzte a statisztikai törvénnyel szembeni követelményeket, és kimondta, hogy a „nem statisztikai adatok statisztikai célú hasznosíthatóságának alapelveit a statisztikai törvénynek kell megfogalmaznia.”⁵² Az adatok megbízhatóságának érdekében pedig kiemelt kérdésként kezeli az adatminőség mérését.

A munkabizottság által kidolgozott ajánlást a KSH mellett nyolc másik szervezet is aláírásával megerősítette.

Itt érdemes megemlíteni, hogy a Fényes Elek Műhely (FEM)⁵³ kezdeményezésében tudományos jellegű műhelykonferencia is foglalkozott az adminisztratív adatállományok statisztikai célú hasznosításával⁵⁴, és az általuk megfogalmazott szakmai ajánlás sok tekintetben hasonló tartalmú, mint amit az OST említett munkabizottsága tett közzé.

3. A KSH és az adatgazdák együttműködési keretének meghatározása – együttműködési megállapodás

A KSH hosszú ideje együttműködési megállapodásokban rögzíti stratégiai partnereivel összefonódó tevékenységeinek kereteit és részleteit. Az esetek egy részében az együttműködés adatok cseréjére is kiterjed. A Gyakorlati Kódex 8.8. ismérve megfogalmazza az együttműködési megállapodások szükségességét az adatátvételekre: „A statisztikai szervezetek az adminisztratív adatok tulajdonosaival megállapodásokat kötnek, amelyek tartalmazzák a felek közös elköteleződését az adminisztratív adatok statisztikai célra való használatára.” Az együttműködési megállapodások egyúttal támogatják a 8.7. ismérvnek („Az adminisztratív adatok statisztikai céloknak való jobb megfelelése érdekében a statisztikai szervezetek részt vesznek az adminisztratív adatok tervezésében”) és a 10.3. ismérvnek („Előremutató lépések történnek az adminisztratív nyilvántartások használhatóságának javítására és a közvetlen adatfelvételeket szolgáló források korlátozására”) való megfelelést is. Az együttműködési megállapodások sablonjának tervezésénél figyelembe vettük azt is, hogy az európai statisztikáról szóló 223/2009/EK rendelet 2015-ös módosítása megerősíti a statisztikai szervezetek pozícióját az adminisztratív adatok statisztikai célú felhasználását illetően.

A KSH sablont dolgozott ki az adatgazda partnerszervezetekkel kötendő együttműködési megállapodásokra (3. melléklet) annak érdekében, hogy a KSH rendelkezésére álljon az adatállományok átvételéhez és feldolgozásához szükséges valamennyi adat, információ és

⁵² Idézett mű: 2. old.

⁵³ A Fényes Elek Műhelyt a Magyar Szociológiai Társaság Társadalomstatisztikai Szakosztálya és a Közgazdaságtudományi Egyesület hozta létre a statisztikai adatfelvételek és adathozzáférés, a statisztikai információk rendszer javításának érdekében. <http://www.mktudegy.hu/?q=fenyeselek>

⁵⁴ Harcsa István: Műhelykonferencia az adminisztratív adatállományok statisztikai célú hasznosításáról – különös tekintettel a népszámlálásra. Statisztikai Szemle. 2010. május http://www.ksh.hu/statszemle_archive/2010/2010_05/2010_05_549.pdf
A műhelykonferenciát a Fényes Elek Műhely szervezte, amelyen a téma kutatóin kívül részt vettek az adatgazdák képviselői is. A vita tanulságai alapján került megfogalmazásra az idézett szakmai ajánlás.

megállapodás függetlenül attól, hogy az adatátvétel statisztikai vagy egyéb adatokra vonatkozik, illetve, hogy erről rendelkezik-e, tájékoztat-e az OSAP vagy sem.

Az együttműködési megállapodás nemcsak az adatgazda által gyűjtött adatok átadására vonatkozik, hanem a vonatkozó metadatokra is, amelyek egyrészt az adatátvételt végző KARÁT és az adat-előkészítést támogató ADAMES IT-eszköz alkalmazásához szükségesek, másrészt tájékoztatást adnak az eredeti adatgyűjtés módszertanáról is. Ha az adatgazda rendelkezik információkkal az adatai minőségéről, a további felhasználáshoz azok is szükségesek. A statisztikai felhasználáshoz részletesen meg kell ismerni, hogy a nyilvántartásért felelős szervezet hogyan gyűjti, dolgozza fel, tartja karban az adatokat. Tudni kell azt is, hogy milyen módon történik az adatok tárolása, és milyen formátumban tudják átadni őket. Célszerű és hasznos annak biztosítása és tisztázása is, hogy a statisztikai hivatal részéről hogyan történjen a visszacsatolás az adatgazdához, ha az átvett adatokban valamilyen problémát észlel. Jó együttműködés kialakítása esetén a másodlagos adatforrás minősége is javulhat a statisztikai felhasználás eredményeképpen. A sablon szerinti megállapodások az átvett adatok gyűjtéséért felelős szervezetekkel erősítik a szakmai kapcsolatot annak érdekében, hogy a másodlagos adatok a KSH céljainak minél jobban megfeleljenek, illetve az adminisztratív adatok statisztikai célú felhasználása szakszerűen történjen.

Az együttműködési megállapodás az alábbi információkörökre terjed ki:

- a megállapodást kötő szervezetek adatai,
- a megállapodás időbeli hatálya, az együttműködés alapelvei, célja, tartalma, jogszabályi háttere, területei,
- kötelezettségvállalások az együttműködés területén,
- az adatállomány, az adatátvétel leírása és
- az adatgazdák kérdőíve a nyilvántartás módszertanáról és az adatok minőségéről.

Bár az adatgazdák kérdőívében feltett kérdésekre adandó válaszok legfőbb ismerői maguk az adatgazdák, mégis fel kell készülni arra az esetre is, ha ezek náluk nincsenek megfelelően dokumentálva. Ezért az együttműködési megállapodásban az adatgazdának csak arról kell nyilatkozni, hogy a kérdőívben felsoroltak közül melyek azok a metaadatok, amiket az adatoknak az együttműködési megállapodás hatálya szerinti (első) átadás alkalmával a KSH rendelkezésére bocsát. Ez biztosítja a fokozatosságot, és csak annyi terhet ró az adatgazdára, amennyit maga is vállalni tud.

Gondoskodni kell az együttműködési megállapodás rendszeres frissítéséről.

VIII. A másodlagos adatok korszerű informatikai kezelése a KSH-ban⁵⁵

A másodlagos adatok statisztikai célú hasznosításához gondoskodni kell arról, hogy ezek az adatminőség és biztonság szempontjait figyelembe véve, megfelelő körülmények között, megbízható csatornán keresztül jussanak el az adatgyűjtés, illetve karbantartás helyéről a statisztikai szervezethez. Az átvett adatoknak ezek után statisztikai adat-előkészítésen kell

⁵⁵ A fejezetben felhasználtuk a KARÁT- és az ADAMES-rendszerről a KSH Statisztikai és informatikai DOKTÁR-ában elhelyezett dokumentumokat is.

átadniük, aminek az a célja, hogy további felhasználásra alkalmassá váljanak, azaz a statisztikai adatrendszer részévé váljanak.

2012 novembere és 2014 szeptembere között lebonyolított EKOP1.A.2-es⁵⁶ konstrukciójú projekt keretében a KSH kifejlesztette a nem kérdőív formátumú adatállományok fogadására, az egyedi adatigényeket kielégítő adat-összeállítások küldésére alkalmas KARÁT-rendszert és a másodlagos adatok előkészítését támogató ADAMES-rendszert.

1. Az adatátvitel szempontjai és lehetséges eszközei

A külső adatforrások hasznosításának kezdeti időszakában az adatállományok mozgatására egyedi megoldások születnek, amiket az aktuális adottságok és a felhasználás módjai határozzák meg. Amikor viszont a statisztikai munkában már magas prioritást élvez és tervszerűvé válik az átvett adatok hasznosítása, a munkát hatékonyabbá teszi egy olyan központi eszköz létrehozása, aminek segítségével biztonságosan elvégezhető a külső adatállományok fogadása, tárolása és a felhasználás helyére történő továbbítása. A biztonság nem csak a fogadó szemszögéből értelmezendő, hanem egyúttal azt is jelenti, hogy az adatküldő szervezetek számára is garantálják az adatvédelmi szabályok betartását.

1.1. Általános szabványokat használó eszközök

Mielőtt rátérnénk a KSH-ban bevezetett, nem kérdőív formátumú adatok fogadására és küldésére alkalmas KARÁT-rendszer ismertetésére, a teljesség kedvéért szólni kell arról, hogy léteznek más megoldások is, amik a KARÁT-hoz képest magasabb szintű standardizáltsággal bírnak.

A statisztikai alkalmazás szempontjából az SDMX⁵⁷-rendszer tekinthető a legelterjedtebb általános szabványokat használó eszközhöz. Az SDMX 2001-ben jött létre a Nemzetközi Fizetések Bankja, az Európai Központi Bank, az Eurostat, az IMF, az OECD, az ENSZ Statisztikai Osztálya és a Világbank kezdeményezésére, és azokat a műszaki szabványokat és tartalom szempontú iránymutatásokat határozza meg, amelyek a modern információs technológia eszközeit használva megkönnyítik a statisztikai adatok és metaadatok cseréjét.

Az SDMX-rendszer lehetővé teszi, hogy minden intézmény megmaradhasson a saját definíciós rendszerének használatánál, miközben egy közös „szállítási” struktúrát alkalmaz az adatok és a metaadatok intézmények közötti cseréjéhez függetlenül attól, hogy kétoldalú vagy többoldalú adatcseréről van szó.

Az Európai Statisztikai Rendszerben folyó SDMX alapú adatküldési fejlesztésekkel összhangban jelenleg már a KSH több szakstatisztikája esetén is kötelező az SDMX-formátumban való adatküldés. Ezzel együtt a KSH az SDMX-szabványt nem használja a statisztikai adat-előállítási folyamat módszertani standardjaként, így a hivatalon belül deklaráltnak és a jövőre nézve is csak egy technikai, adattovábbítási standard marad. Az SDMX-standardot olyan statisztikai hivatalok alkalmazzák az adat-előállítási folyamatokra is,

⁵⁶ Elektronikus Közigazgatás Operatív Program. Ágazati pályázat a közigazgatási szolgáltatások belső folyamatainak megújítására. Célja: több közigazgatási szakrendszer informatikai fejlesztése az elektronikus ügyintézés elterjesztése, illetve a közigazgatási munka hatékonyabbá tétele érdekében.

⁵⁷ Statistical Data and Metadata Exchange.

amelyeknél a folyamatra alkalmazott rendszerek kevésbé integráltak⁵⁸, mint a KSH-ban, és az SDMX segítséget jelent a módszertani standardizálásban. A KSH szempontjából sem a külső körülmények – sok, eltérő informatikai feltételekkel rendelkező partnerszervezet –, sem pedig a belső viszonyok – a rendszerek nagyfokú integráltsága és fejlettsége – nem indokolják az SDMX használatát.

Ez a rendszer azonban több szervezet közötti kölcsönös adatcserét nem tesz lehetővé, mert az valamennyi résztvevő számára alkalmas szabványok kidolgozását tette volna szükségessé. A fejlesztés minden résztvevőtől lényegesen nagyobb erőforrást és sokkal több időt igényelt volna, mivel saját rendszereikben is alkalmazniuk kellett volna ezt a szabványt az adatok SDMX-nek megfelelő továbbításához. Mindezek miatt érthető, hogy egyelőre az általános szabványokat használó eszközök világszerte csak kivételesen fordulnak elő. Egy ilyen megoldás az összes partner hasonló fokú érdekeltségét feltételezi.

1.2. A KSH egyedi megoldása

A KSH adatátvételeihez fejlesztett IT-eszköz a másodlagos adatok adatgazdáival való egyedi kétoldalú kapcsolatait harmonizálja figyelembe véve nemcsak a saját, hanem az adatküldő partnerszervezetek informatikai lehetőségeit is. A KSH-ban kifejlesztett eszköz úgy lett kialakítva, hogy az adatfogadáson felül az adatfelhasználók egyedi kérésére készített adatösszeállítások célba juttatására is alkalmas legyen. Tekintve azonban, hogy az adatkiadás nem kapcsolódik közvetlenül a másodlagos adatok kezeléséhez, a funkcióit sem részletezzük, legfeljebb csak a teljesség kedvéért említjük meg őket.

Az alábbiakban azokat az elvárásokat írjuk le, amik a KARÁT-rendszerrel szemben megfogalmazódtak, hogy a KSH és az adatgazda partnerek közötti bilaterális kapcsolatok központosított kezelésére alkalmas legyen.

A rendszernek kettős funkciót kell teljesíteni.

1. Rugalmas megoldást ad az adatállományok fogadásához és továbbításához.
 - A szolgáltatásokat előzetesen feltöltött metaadatok vezérlik, és egységes eljárásokat alkalmaznak.
 - Az adatküldő szervezet számára lehetőséget nyújt arra, hogy az adatállományokat és a hozzájuk tartozó metaadatokat egy felületen töltsse fel.
 - Elvégzi az állományok formai ellenőrzését.
 - Az átvett állományokat lehetőség szerint adatbázisba tölti a további feldolgozás számára, az adatbázisba nem tölthetőknél pedig biztosítja az állományok letölthetőségét a jogosult statisztikusok számára.
 - Az adatkérő partnereknek átadandó adatállományokat a rendszer egy kezelő felületen megjeleníti, ahonnan a külső felhasználó eléri és letöltheti a szükséges metaadatokkal együtt.
 - A rendszer az adatok feltöltésével, ellenőrzésével, letöltésével kapcsolatos valamennyi lépést naplózza.
2. Lehetőséget biztosít az adatforgalom irányítására és a folyamat dokumentálására is.

⁵⁸ Például azért, mert nincs saját metaadatszabványuk vagy hatékonysági okokból „megéri” az SDMX-szabványt belső metaadat-standardként implementálni.

- A rendszer lehetőséget biztosít arra, hogy a teljes folyamatot nyomon lehessen követni a nyilvántartásba vételtől addig, hogy az állomány megérkezéséről a feladó nyugtát kap. Az adatállományok mozgását a rendszer naplózza.
- A naptár szolgáltatása nyilvántartja a teljesítendő adatátadási feladatokat, és a határidő előtt e-mailben figyelmeztet az adatátadási kötelezettségre, és a határidő után az elmaradt teljesítésre.
- Előre rögzített minőségi indikátorok állnak elő a folyamat során.

Az eszköznek az alábbi további tulajdonságokkal kell rendelkeznie:

- a hálózat, amelyen az adatforgalom, illetve az üzemrend folyik, nagy méretű fájlok zökkenőmentes küldését, fogadását is biztosítja;
- a hálózat eleget tesz az adatállományok biztonsági követelményeinek;
- a kezelőfelület egyszerű, felhasználóbarát és biztonságos;
- a statisztikai szervezet munkatársainak és a külső felhasználóknak a rendszerhez való hozzáférését, bejelentkezésének módját jogosultsági szabályok szerint biztosítja;
- a külső felhasználók a felületen csak a rájuk tartozó részeket látják, és csak a számukra engedélyezett műveleteket végezhetik;
- a rendszer biztosítja az adatállományok megfelelő ideig történő tárolását, gondoskodik az archiválás lehetőségéről;
- az üzemrend idő- és költséghatékony módon ütemezi a feladatokat, például nagy állományok ütemezett (akár éjszakai) küldése;
- az eszköz használata a partnerszervezetek informatikai és adminisztratív rendszerében a lehető legkevesebb változtatást igényli;
- az elektronikus adatfogadás szervesen kapcsolódik az adat-előkészítő rendszerhez.

Ha a rendszer szolgáltatásai egy web alapú felhasználói portálon keresztül érhetőek el, lehetővé válik, hogy az eszköz általános, olcsó, és egyszerűen kezelhető legyen.

1.3. A megvalósult KARÁT-rendszer

A KARÁT elsődleges feladata a KSH-ba érkező adatátvételek és küldések ellenőrzése, hibás formátumok esetén az átvétel megtagadása, újraküldetése. A felhasználói jogosultság kezelése biztosítja az egyedi adatok védelmét. Az alapfunkciók mellett naptárszolgáltatás, e-mailes figyelmeztetés tájékoztatja az érdekelteket a határidőkről és egyéb feladataikról. Itt történik a beérkező fájlok alapvető ellenőrzése is, hogy csak a formailag alkalmas adatállományok kerülhessenek be a statisztikai adat-előállítási folyamat további fázisaiba.

Emellett a KSH-ból kifelé irányuló adatforgalom számára is biztonságos adattovábbítást, letöltési lehetőséget biztosít.

A KARÁT-rendszer a következő három alkalmazásból áll.

1. A belső KARÁT-rendszer, ami csak a KSH belső informatikai rendszerében érhető el, és ahol az alábbi funkciók működnek:
 - az adatátvételek karbantartása (tervezése, szervezése, érkeztetése, másolása),
 - az adatátvételek lekérdezése (állományok, metaadatok letöltése, feladatlisták lekérdezése),
 - az adatkiadások lekérdezése,
 - az adatforgalom monitorozása,

- tevékenységek, események adminisztrálása.
2. Külső KARÁT a KSH honlapon – adatátadás, amelyen az adatgazdák/adatbenyújtók tölthetik fel az adatállományokat a KSH statisztikai rendszere részére.
 3. Külső KARÁT a KSH honlapon – adatkiadás, amelyen az adatkérők tölthetik le az adatkiadási nyilvántartás (ADKI) rendszer által átadott adatállományokat (ennek ismertetése nem témája jelen leírásnak).

1.3.1. Belső KARÁT-rendszer

A KARÁT-rendszerbe azok a KSH által végrehajtott adatátvételek kerülhetnek be, amelyek a KSH adatforrás nyilvántartásában (metarendszerben) szerepelnek.

A KARÁT működése az adatátvételi terv (KARÁT-metaadat) leírásához kötött. A terv előkészítése az adatátadó partnerekkel való egyeztetést igényel, amelyben meghatározzák, az átadás kapcsolattartóit, határideit, az átadandó adatállományok tartalmát, formáját, szerkezetét, mellékleteit, az átadás csatornáját. Az előkészítés során el kell dönteni, hogy az adatokat automatikusan adatbázisba akarja-e tölteni a statisztikus a további feldolgozás érdekében, vagy nem standard eszközökkel dolgozza fel, ezért csak az állományok átvételére, letöltésére van szükség.

A KARÁT-metaadatok és -terv leírása a KARÁT „tervezés” felületén történik (vonatkozási idő, határidők, üzenetek, küldendő állományok, a küldendő állományok feldolgozásának módja, szerkezete). Több adatgazdás adatátvételek esetén mód van az adatgazda csoportok számára eltérő tervváltozat megadására is.

A „szervezés” funkció hozza létre az adatátvételi terv alapján az adott adatátvételi időszak adott adatbenyújtóhoz tartozó feladatlista tételét, amelyre az adatátvételi feladatok (adatfeltöltés, üzenetküldés, érkeztetés stb.) épülnek, és amelyen ezek megtörténte regisztrálásra kerül. Az itt megadott információk teszik lehetővé az adatátvételek minőségének ellenőrzését is. Emellett a szervezés funkció módot ad az évközbéli kapcsolattartó, határidő módosítások átvezetésére, a hibásan beküldött adatok újraküldésének engedélyezésére is.

Az adatállományok érkeztetése automatikus, ha azokat a KARÁT-rendszeren keresztül küldték. Lehetőség van azonban a más módon beérkező adatállományok érkeztetésére és az állomány belső KARÁT-rendszerben való feltöltésére is.

A beérkezett állományok feldolgozására két mód van.

1. Adatbázisba tölthetnek az átvett állomány szerkezetének megfelelően. A rendszer az adatbázisba töltött másodlagos adatokat elérhetővé teszi a statisztikai adat-előállítási folyamat következő lépése, az adat-előkészítés számára. Ennek eszköze az ún. ADAMES-rendszer.
2. Változatlan módon letölthetők a KARÁT-rendszerből, és egyedi módon felhasználhatók.

A rendszerben az adatátvételek tervezési és végrehajtási folyamata, állapota nyomon követhető, ellenőrizhető az adatátadó partner adatfeltöltő tevékenysége.

1.3.2. Külső KARÁT-rendszer

Az adatállományok beküldése és a KSH-től kért adatok szolgáltatása, átvétele biztonságos csatornákon történhet. A feltöltési, illetve letöltési lehetőségek egyike a KSH honlapján található KARÁT-felület, a másik a kormányzati hálózaton keresztül a hivatali kapu használata.

Mód van továbbá webszolgáltatáson vagy IBM WebSphere MQ-n keresztüli fel-, illetve letöltésre is.

A KSH honlap adatátadó és adatkiadó KARÁT-felületeire a felhasználók KARÁT regisztráció alapján vagy a KARÁT-hoz szükséges információk egyszeri megadása után ügyfélkapun keresztül léphetnek be. Amennyiben az adatátadó szervezetnek kérdőíves adatszolgáltatási kötelezettsége is van, az adatküldés teljesíthető a KSH elektronikus adatgyűjtő rendszerén, az ELEKTRÁ-n keresztüli belépés után is.

Az OSAP-ban előírt vagy megállapodás alapján teljesítendő adatátvételi feladatokról az adatgazdákat, illetve a helyettük az adatátadást teljesítő adatbenyújtókat a KSH értesíti a feladataikról az adatküldési évet megelőzően. Amennyiben az átadandó állomány szerkezete rendelkezésre áll, igény esetén a KSH mintaállományt is küld, amelybe az adatok betölthetők.

A KARÁT-felületen történő adatbenyűjtáshoz a benyújtók számára rendelkezésre áll a „feltöltés felhasználói felület”, amelyen a külső felhasználók feltölthetik az állományokat, valamint szükség szerint rögzíthetik az adatállományt kísérő adatokat (szerkezetleírást, az adatállományban alkalmazott értékkészleteket). A feltöltött állományokat a modul verziókezelten adatbázisban tárolja, és elindítja az adatátvétel állományainak feldolgozását.

A többi csatornán beérkező adatok esetében a feldolgozás automatikusan történik. A KARÁT-rendszerben eseménynapló működik, amelybe az előzőekben felsorolt adatátvételekkel kapcsolatos tevékenységek bejegyzésre kerülnek. Az adatátvétellel kapcsolatos tevékenységekről (pl. az állomány sikeres vagy hibás beküldéséről) a rendszer visszaigazolást küld a kapcsolattartónak.

A rendszerbe xml/xsd, csv, txt, xls, xlsx, doc, docx, pdf, zip és egyéb típusú fájlok tölthetők fel. Amennyiben azonban a KSH statisztikusa igényli az adatok adatbázisba töltését, az állomány típusa csak xml/xsd, csv vagy txt lehet.

Az egyéb formátumban érkező adatállományokat a rendszer fogadja, de további feldolgozása a KARÁT-rendszerben nem biztosított.

A KARÁT adatátadó rendszer felületei, az alábbi szolgáltatásokat biztosítják az adatbenyújtók számára:

- Adatátadás – feladatlista, a bejelentkezett felhasználó által felügyelt vagy számára kijelölt feladatok állapota és a kiválasztott feladat történetének megjelenítése.
- Határidőnapló – bejelentkezett felhasználó által felügyelt, vagy számára kijelölt feladatok határidei havonta és naptár nézetben.
- A feltöltés felülete, amelyen a benyújtók feltöltik az adatállományokat és a csatolandó szerkezetleírásokat, értékkészlet-leírásokat, egyéb dokumentációt.
- Az adatátvételre vonatkozó metaadatok (adatállomány-szerkezet vagy értékkészlet) leírását segítő felület (ha nem a KSH írja elő a szerkezetet).

2. Az adat-előkészítés szempontjai és eszközei

2.1. Az adat-előkészítés alapelvei

Mielőtt a másodlagos adatokat a statisztika fel tudná használni, statisztikai adat-előkészítésen kell átesniük, amely során megtörténnek azok az átalakítások és javítások, amik lehetővé teszik, hogy az adatok a kitűzött rendeltetésnek megfeleljenek. Ennek során a cél egy statisztikai

felhasználásra alkalmas adatállomány előállítása egy vagy több adatforrás felhasználásával, amelyben az adatok a statisztikai sokaság azonosítható egységeire vonatkoznak.

Az átvett adatok lehetnek elemiek, vagy aggregáltak, és mindkét fajta állomány esetén szükség lehet az adatok szerkesztésére. A magas feldolgozottságú adatok azonban nem igényelnek elektronikus adat-előkészítést, mert az adatkezelő rendelkezésére álló ellenőrzési lehetőségek nagyon korlátozottak, és az adatokat közvetlenül – jórészt publikálásra – használják fel.

Egy statisztikai szervezet általában fel van készülve arra, hogy a saját adatgyűjtéseiből származó eredmények adat-előkészítését korszerű eszközzel végezze, ami biztosítja az egységes adatjavítási szempontok érvényesítését, a kontrollálhatóságot, a dokumentáltságot, illetve az előkészített adatállomány továbbítását és archiválását. Az átvett adatok kezelése azonban a közvetlen adatgyűjtésektől eltérő rendszert tesz szükségessé alapvetően azért, mert nem az újrahasonosító szervezet által kitűzött statisztikai céloknak megfelelően lettek összegyűjtve, és ez a következő lehetséges következményekkel jár:

- a célsokaság egysége az átvett adatforrásban gyakran nem azonos a statisztikai szervezet által definiált szakstatisztikában használttal;
- a használt definíciók és nomenklatúrák az adatgazdánál és a statisztikai szervezetnél eltérőek lehetnek, emiatt konvertálásra lehet szükség;
- az átvett adatgyűjtés és a statisztika prioritásai különbözők lehetnek, és ezért a hasznosításra kerülő adatok minősége, ellenőrzöttsége nem megfelelő, illetve nem egyenletes;
- az átvett adatok vonatkozási ideje nem egyezik meg a statisztikában használttal;
- az átvett bonyolult adatrendszerekből nehéz kinyerni a statisztika igényeinek megfelelő információkat;
- az ágazati politikai szempontok változásával a nyilvántartás/adatgyűjtés tartalma, szerkezete módosulhat;
- biztosítani kell, hogy az átvett adatok szükség szerint összekapcsolhatók legyenek más adminisztratív vagy statisztikai adatforrásból származó adatokkal;
- az átvételek technikai megvalósítása heterogén, az adatformátumok eltérőek.

A másodlagos adatokhoz kialakított rendszer segítségével az átvett adatok adat-előkészítése a statisztikai eszközökkel gyűjtött adatokéhoz hasonló módon vezérelhető, illetve végezhető, így annak ellenére, hogy a fogadás és a betöltés olyan funkció, amely egy átvett adatállomány egészére vonatkozik, az ellenőrzés és javítás lépései rekordonként is végrehajthatók. Hasonlóság érvényesül abban az értelemben is, hogy bizonyos funkciók (pl. a hiányzó kérdőívek/rekordok pótlása, aggregált adatok részletes elemzése) most sem ezzel az eszközzel történnek.

- Egy egységes adat-előkészítő rendszerrel szemben elvárt általános alapvető jellemzők, hogy
- funkciói az egyedi feladatokhoz rugalmasan alakíthatók legyenek,
 - adjon lehetőséget az adatok összetett szempontok szerinti ellenőrzésére,
 - dokumentálja az alkalmazott funkciókat,
 - az elvégzett műveletek legyenek rekonstruálhatók,
 - az adatállományok megbízható tárolása és archiválása is legyen biztosított,
 - az ellenőrzések és adatjavítások más által is előhívhatók legyenek, és a jogosultak által kontrollálható módon működjenek.

Az adat-előkészítő egy olyan általános, szabványos eljárásokat alkalmazó keretrendszer, melybe az egyes szakstatisztikák alkalmazásai beilleszthetők. Segítségével a felhasználók önállóan dolgozhatnak, és a kapcsolódó metarendszer megfelelő előkészítése után munkájuk ellátása során általában nincs szükség informatikus segítségére.

A másodlagos adatok előkészítését végző informatikai eszköz segítségével akkor lehet a különböző állományokkal összefüggő eltérő szerkesztési feladatokat elvégezni, ha rendelkezik az alábbi sajátosságokkal:

- az adatátvételek jellemzőit, az állományokkal végzendő feladatokat a metarendszerben leírja, dokumentálja;
- lépésekre, szakaszokra bontással, moduláris megvalósítással a folyamatot átláthatóvá, hatékonyan fejleszthetővé és biztonságosan üzemeltethetővé teszi;
- szabványos, egységes megoldásokat alkalmaz;
- jogosultság kezelése biztosítja a szenzitív adatok védelmét;
- támogatja az egyes átvételekhez kapcsolódó feladatokat, funkciókat (pl. ellenőrzés, adatjavítás) opcionális megvalósíthatóságát;
- folyamatkövető, vagyis a lépések dokumentáltak, és a folyamat állapota lekérdezhető;
- a lehető legkisebb mértékben igényel programozói kapacitást, főként a leírt metaadatok alapján üzemeltethető, illetve olyan eszközrendszerre épít, amellyel a felhasználók saját maguk valósíthatják meg az igényeiknek megfelelő funkciókat.

A másodlagos adatok statisztikai célú használata során nagyon gyakran nem elegendő egyetlen adatforrás használata a szükséges változók előállításához, hanem más adatot, adatforrást is be kell vonni. Ezért az átvett adatok szerkesztésére használt eszköznek alkalmasnak kell lennie arra, hogy kettő vagy akár több „külső” adatállományt is párhuzamosan lehessen használni. Fontos, hogy ha statisztikai számításunkhoz egy speciális információ hiányzik, ami a rendelkezésre álló adatokból közvetlenül nem áll elő, ún. származtatott változót hozhassunk létre. Ebben az esetben a rendelkezésre álló információk együttes figyelembevételével kell becslést készíteni a hiányzó változóra, ami, ha nem is teljesen pontos, de legalább jó közelítést adhat.

Az eszköz használata esetén minden adatátvételnél általánosan biztosítani lehet, hogy adatbevétel, javítás csak az arra kijelölt dolgozók végezzenek, hogy minden ellenőrzés az állomány összes egységére kötelezően megtörténjen, hogy indoklás nélkül ne lehessen hibát elfogadni, valamint, hogy semmiképpen ne maradjon az állományban súlyos kategóriába sorolt hiba.

Egy ilyen IT-rendszer összességében hozzájárul ahhoz is, hogy az adatok időben rendelkezésre álljanak elősegítvén az időszerű és minél gyorsabb publikálást, miközben megfelelő idő marad a validálásra és elemzések készítésére is.

2.2. Az adat-előkészítő rendszer szükséges funkciói

Az adat-előkészítő rendszernek alapvetően három funkciót kell elvégezni, az adatok ellenőrzését, javítást; a minőségi indikátorok előállítását, továbbá az ellenőrzési, javítási folyamat állapotának és az adatoknak a lekérdezhetőségét.

2.2.1. Adatok ellenőrzése, javítása

A statisztikai adatállomány ideális esetben „tiszt” és következetes abban az értelemben, hogy nincsenek ellentmondások az egyéni adatokon belül. Az adatellenőrzés során a valószínűtlen információkat kell megkeresni.

Az ellenőrzések, javítások kétféleképpen történhetnek.

1. Képernyőn, a szakstatisztikus folyamatos jelenléte, közbeavatkozási lehetősége mellett.
2. A háttérben, miközben a szakstatisztikus egyéb feladataival tud foglalkozni. Igény szerint ellenőrizhető, hogy a folyamat éppen hol tart, és miután befejeződött az ellenőrzési eljárás, hibalisták állíthatók elő, amik a hiba helyét, jellegét és súlyosságát tartalmazzák.

A rendszernek az állományban meg kell jelölni a hibás és a pótolta, javított adatokat, jelezve a hiba súlyosságát, és arra is lehetőséget kell biztosítani, hogy egy-egy adathoz megjegyzést fűzhessünk.

2.2.2. Minőségi indikátorok előállítása

A másodlagos adatok minősége, vagyis a rendszerhibák és az egyedi hibák, illetve a hiányosságok jellege és előfordulási gyakorisága még akkor is befolyásolhatja a felhasználással előállított statisztikai mutatók minőségét, ha a szerkesztés során logikai szempontból az adatok konzisztenssé váltak. Ezért annak érdekében, hogy jó döntést hozzassunk a másodlagos adatforrások statisztikai célú használhatóságáról, továbbá, hogy biztosítsuk a statisztikai tevékenység átláthatóságát, dokumentálhatóságát, az adat-előkészítés során a rendszer előre meghatározott minőségi indikátorokat állít elő. Ezek a mutatók az adatok pontosságáról, a hibatípusokról és azok gyakoriságáról, valamint összekapcsolandó állományok esetén az illeszkedés mértékéről adnak információt.

A minőségi indikátorok előállítása nemcsak a statisztikai újrahasonosítás szempontjából nyújthatnak hasznos információkat, hanem hasznosak lehetnek az adatgazdák számára is. Nagyon fontos ugyanis, hogy azokban az adatforrásokban, ahonnan statisztikai célra átveszünk adatokat, lehetőleg minden fajta szisztematikus hiba ki legyen javítva. Ha szoros kapcsolatot tartunk fenn az adatállományt karbantartó felelős szervezettel, aminek keretében tájékoztatjuk a talált hibák típusáról, jellegéről, ezeket az információkat hasznosítani tudják az adatállomány tisztítása során. Nemzetközi tapasztalatok alapján számos olyan helyzetről lehetne beszámolni, amelyben statisztikusok ismertek fel hibatípusokat, amiket utólag javítottak, és ebből haszná lett mind az adatgazdának, mind pedig a statisztikai szervezetnek.

2.2.3. Lekérdezés

Ez a funkció információt nyújt az adatellenőrzési és javítási folyamat lépéseiről, valamint az adattartalomról. Ha a lépéseket a rendszer adminisztrálja, információt tudunk kinyerni a folyamat készülségéről és állapotáról, továbbá arról is, hogy egy-egy lépés sikeres volt-e vagy sikertelen, illetve az adat-előkészítés során fellépő hiba esetén annak jellegéről.

A lekérdezési funkció lehetőséget ad hibalisták megtekintésére is. Az adattartalom lekérdezéséhez olyan felületet célszerű biztosítani, amivel a felhasználók rugalmasan a saját igényeik szerinti lekérdezéseket valósíthatnak meg, aminek formátumai későbbi, ismételt felhasználásra rögzíthetők.

2.3. A megvalósult ADAMES-rendszer

Az ADAMES-rendszer a KSH-ba nem kérdőív formátumban érkező adatok adat-előkészítésére alkalmas, amibe beletartoznak az elemzői, ellenőrzési és javítási funkciók is. Ez az informatikai eszköz a KARÁT-on keresztül beérkezett adatállományok kezelésére készült, elősegítve az adat-előkészítési folyamatok KSH-n belüli egységes kezelését és a kapcsolódó rendszerekkel való egységes kommunikációt. A KSH adatbázisába betöltendő adatok az ADAMES-ben kapják meg a metakódjaikat, ezért ezekben az esetekben alkalmazása kikerülhetetlen, és feltételezi a szükséges metainformációk előkészítését is. A már előkészített adatok az adatfeldolgozó rendszerbe (EAR), vagy egyéb célrendszerekbe (ADÉL, regiszterek, tájékoztatási rendszerek és termelési adatbázis) kerülnek.

Az ADAMES-rendszernek az alábbi főbb funkciói vannak.

- Feladattervek létrehozása. Az adat-előkészítés folyamatát előre meg kell tervezni. Az egyes feladatterveket a felhasználó a kialakult legjobb gyakorlata alapján rugalmasan, egyedileg paraméterezzhetően tudja összeállítani. A rendszer olyan grafikus felületet biztosít a felhasználó részére, ahol a megszokott Windows-technológiák segítségével gyorsan és egyszerűen lehet terveket összeállítani.
- Szerepkörök kezelése, státuszkezelés. A feladattervekhez csak jogosultságokon keresztül lehet hozzáférni, amelyeket egy adminisztrátori szerepkörrel rendelkező felhasználó tud egyedileg beállítani. Az egyes feladattervek státuszát az adat-előkészítői folyamathoz igazodva tudja az arra jogosult felhasználó a rendszerben állítani, ezáltal a feladatterv életciklusában bekövetkező státuszváltozások határozzák meg a terven végrehajtható funkciók körét. Ezen funkciókat csak a meghatározott szerepkörökkel lehet elérni.
- Elemzés. A rendszer olyan elemző, grafikonmegjelenítő felületet biztosít, ahol a szakstatistikus felhasználó az adatokból elemzést tud készíteni, azok helyességét, formaiságát „ránézéssel” ellenőrizni, hogy ezeket az információkat a feladatterv készítése során felhasználhassa.
- Naplózás. A rendszer minden tevékenysége (akár manuális, akár automatikus) naplózásra kerül, amely bejegyzések könnyen visszakereshetőek, ellenőrizhetőek.
- Riportolhatóság. Az elkészült feladattervek több formátumban (Excel, pdf) is riportolhatóak, nem csak a feladattervek szerkezete, tartalma, hanem a feladatterven belüli adatok is.

A feladatterv szakaszolható, a szakaszokat műveleteknek nevezzük, amik műveleti lépésekből építhetők fel. A műveletek között a rendszer az adatokat ideiglenes táblákban tárolja, így hiba esetén nem szükséges az egész folyamatot az elejéről indítani, elegendő a hibás szakaszt megismételni.

Az adat-előkészítés folyamata a következő fő műveleti lépésekből állítható össze.

- Adatáttemelés. A felhasználó ebben a szakaszban határozza meg, hogy a KARÁT-rendszer által biztosított forrásadatok mely részét fogja felhasználni az adat-előkészítés során. Szűkítheti a táblázatok oszlopait, és szűrheti azok sorait.
- Ellenőrzés, elemzés. Ebben a szakaszban az átemelt állományokon különböző ellenőrzések végrehajtására kerülhet sor. Ilyen ellenőrzések lehetnek például egy oszlop szabálynak való megfelelése, valamint különböző oszlopok közötti

összefüggések megléte. Az ellenőrzéseken túl lehetőség nyílik az adatok keresztábrás elemzésére, valamint grafikus megjelenítésére is.

- **Javítás.** Ebben a szakaszban a felhasználó által megadott feltételeknek megfelelő adatok javítására kerül sor egy táblázat vagy konstans érték alapján. Az előkészítés során kétféle javítás végezhető el az adatokon: automatikus és manuális. Automatikus javítás során a tipikus hibákra létrehozott javító algoritmusok futnak le. Manuális javítás során a felhasználó, egy külön erre létrehozott felületen megnézheti, valamint módosíthatja az ideiglenes tábla tartalmát.
- **Átalakítás.** Ebben a szakaszban az adatok struktúrájának átalakítására kerül sor, hogy az megfeleljen a célrendszernek.
- **Minőségi mutató számítás.** Az ADAMES-rendszerben minőségi mutatók számítására is van lehetőség a következő kategóriákban: elfogadott hibamutatók – hibamutatók – javítási mutatók.
- **Adattovábbítás.** A feladatterv utolsó lépéseként az adat-előkészítő rendszer output sémájában külön táblákba kerülnek a kész, átadásra váró adattáblák. A célrendszer számára történő átadás csak akkor történhet meg, ha csak olyan súlyos hibák vannak, amiket a témafelelős jóváhagyott, és csak olyan elfogadható hibák vannak, amiket indoklással ellátott. A célrendszer számára történő átadás csak akkor történhet meg, ha minden mutató le van írva a metában és az ADAMES-meta párosítás már megtörtént. A célrendszerek az output sémátábláiból tudják átvenni az adatokat saját felhasználásra.

E szakaszok csupán az adat-előkészítés jellemző elméleti felbontási egységei, de a gyakorlatban egy feladat ennél több vagy kevesebb műveletből állhat, és egy műveleten belül a lépések számát, fajtáját és sorrendjét csak az határozza meg, hogy az adott szakasznak milyen szerepet szánunk.

A rendszerbe külső tábla is beolvasható lehetővé téve az ellenőrzött állomány adatainak összehasonlítását máshonnan származó értékekkel. Az ADAMES-en belül egyszerűbb adat-összekapcsolások is megvalósíthatók.

Az összeállított feladatterv a szerkesztés ideje alatt tesztüzemmódban futtatható. A futások ilyenkor a felhasználó saját gépén történnek. A véglegesített feladattervek esetén éles futtatást lehet végezni, és ebben az esetben a feladatterv a rendszer szerverén fut. A futtatás eredménye minden esetben megjelenik. A véglegesített feladatterv egy következő időszaki adat-előkészítés esetén újrafuttatható, de szükség esetén a véglegesítés felfüggesztésével tovább is szerkeszthető.

3. A KARÁT- és ADAMES-rendszer illeszkedése a KSH informatikai környezetébe

A KSH informatikai rendszere integrált, ami a központi szoftverek kétféle szempontú összehangoltságát is jelenti. A rendszer elemei egyrészt interfészekon keresztül kapcsolódnak egymáshoz, másrészt a működésükhöz az általános paramétereket a metarendszerekből veszik át. Az alábbi ábra a KARÁT és az ADAMES illeszkedését mutatja a KSH informatikai környezetébe (11. ábra).

A KARÁT és az ADAMES működését tekintve az adatkiadási nyilvántartás és a meta szerepéről a korábbiakban már volt szó.

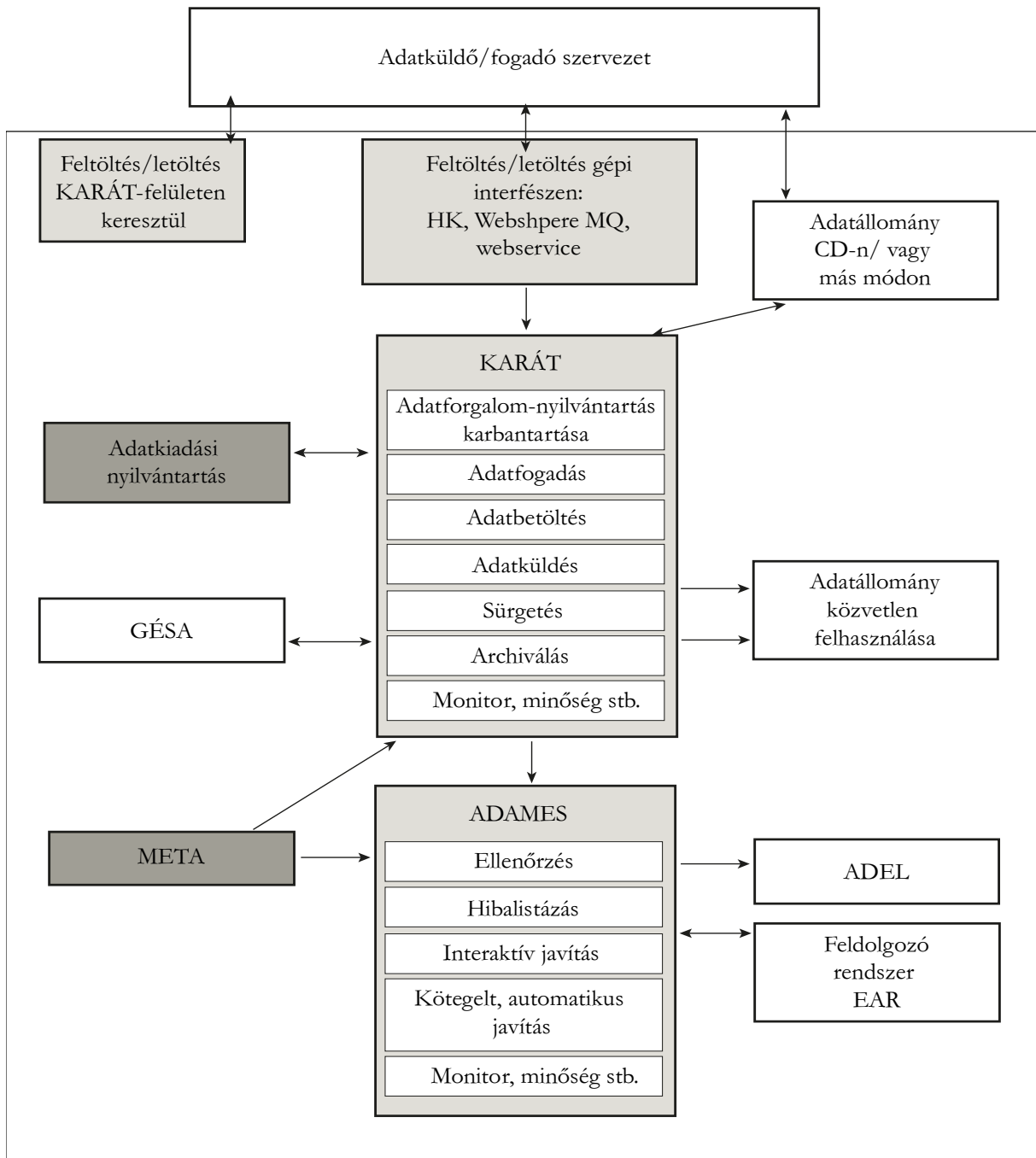
A GÉSA-rendszert a KSH adatgyűjtés-szervezési funkcióinak támogatására hozták létre. Ez szolgáltatja a KARÁT részére is az adatgazdák, illetve az adatbenyújtók aktuális listáját és

az adatátvételeknek az adatgyűjtésekkel közös metaadatait. A GÉSA-rendszer az intézményi gazdaság- és társadalomstatisztikai felvételek sokaságának, felvételi keretének és mintájának (adatszolgáltatói körének) leírására szolgál. Ha az adatátvétel egy GÉSA-ban leírt felvétel meghatározott részsokaságához szolgáltat adatot, az átvett adatok, statisztikai egységek teljessége ellenőrizhető a KARÁT-rendszer segítségével.

Az ADEL rendszer a KSH saját adatgyűjtéseiből származó adatok szerkesztésére készült. Ha az átvett adatokat közvetlen statisztikai adatgyűjtéssel kapcsoljuk össze, az ADAMES által szerkesztett másodlagos adatokat az ADEL veszi át, hogy az elsődleges adatokkal együtt folytassa az adat-előkészítést.

Az EAR az adatfeldolgozás egységes eszköze, ami jellemzően az adat-előkészítést követően veszi át az adatokat, de az is előfordulhat, hogy az adatok előkészítését olykor meg kell szakítani bizonyos feldolgozási funkciókkal (pl. imputálás), és ilyenkor az ADAMES és az EAR felváltva kezeli az adatokat.

A KARÁT- és az ADAMES-rendszer átfogó környezeti diagramja



IX. Másodlagos adatforrásból származó adatok felfedés elleni védelme

1. A felfedés elleni védelemről általában

A Központi Statisztikai Hivatal széles körű tájékoztatási tevékenységet végez, melynek keretében nagy mennyiségű feldolgozott aggregált információt állít elő a gazdaság- és társadalomstatisztika számos területéről. Az aggregált információkat rendszerint táblázatos formában hozza nyilvánosságra; ezeket az adatokat tekintjük táblázatos adatoknak. A statisztikai adat-előállítási folyamat eredményeként a KSH-nál olyan adatállományok is keletkeznek, amelyekben egy statisztikai egység egy rekordot képez, és a táblázatos adatok előállításának elsődleges forrásait jelentik. Ezeket az adatállományokat mikroadatoknak nevezzük. Az adatvédelmi tevékenység az aggregált adatok, illetve a mikroadatok védelmét egyaránt szolgálja.

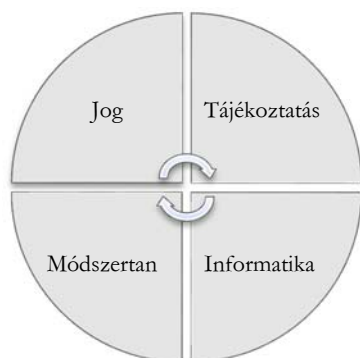
A mikroadatok és táblázatos adatok statisztikai adat-előállítási folyamatok eredményeként állnak elő. Az adatgyűjtéstől az adatok publikálásáig és archiválásáig a statisztikai célra kezelt adatok védelmét biztosítani kell a jogosulatlan hozzáféréstől. A vonatkozó jogszabályok kiterjednek az adatkezelés, adattovábbítás, adatfeldolgozás során betartandó adatbiztonsági követelményekre. Mivel az adatvédelmet általában a jogi környezetből eredeztetik, természetszerűleg adódik, hogy jogi rendelkezések írják le a kereteit. Az adatvédelem általánosan használt fogalom, noha egyértelműen meghatározni nehéz. Habár a jogszabályi környezet sem ad definíciót rá, kiindulásként azt mondhatjuk, hogy minden az adatvédelem tárgykörébe tartozik, amit a releváns jogszabályok a tárgykörbe sorolnak (kiemelten a statisztikai törvény, valamint az információs önrendelkezési jogról és információszabadságról szóló 2011. évi CXII. törvény). Példaként említhető a KSH-ban használt módszertani adatvédelem két alapeleme, amelyek a statisztikai törvényből és a hozzátartozó végrehajtási rendelet rendelkezéseiből erednek. Ezek a következők:

Az 1993. évi XLVI. törvény a statisztikáról 17. § (2) kimondja, hogy „nem lehet nyilvánosságra hozni természetes és jogi személy, valamint a jogi személyiséggel nem rendelkező adatszolgáltatóval kapcsolatba hozható adatot, azaz egyedi adatot”.

A 170/1993.(XII.3.) Korm.rendelet a statisztikáról szóló 1993. évi XLVI törvény végrehajtásáról 19. §-a kimondja, hogy „összesítve sem lehet nyilvánosságra hozni olyan adatot, amelynél az adatszolgáltatók száma háromnál kevesebb”.

A gyakorlatban az adatok védelme alatt a jogi vonatkozásokon felül több mindent értünk. Ide tartozónak érezhetjük a fizikai adatbiztonságot (pl. adatokat tároló meghajtók belső hálózatoktól izolált környezetben való tárolása, infrastrukturális védelem az adatvesztés megakadályozásáért, erősen szabályozott jogosultság az adatokhoz való hozzáféréshez stb.), az adatok célhoz kötött kezelését vagy annak gyakorlatát, hogy az egyes statisztikai egységekre (pl. természetes személyekre, háztartásokra, vállalatokra stb.) vonatkozó információk megismerését matematikai-statisztikai módszerek alkalmazásával is biztosítjuk.

A statisztikai adat védelemének négy területe



A statisztikai adat-védelemnek négy fő területe van.

A **jog** elsősorban a jogszabályok, illetve alkalmazott jogi garanciák (szerződések, kötelezettségvállalások stb.), a **módszertan** a felfedés elleni védelem, az **informatika** az informatikai biztonsági kérdések, a **tájékoztatás** pedig főként az adathozzáférési csatornák működtetése területén illetékes.

A hivatalos statisztika által gyűjtött információk statisztikai célból kerülnek felhasználásra, azaz gyűjtésük célja, hogy valamely jelenséget összevont, aggregált információk alapján jellemezzenek. A statisztikai célhoz kötődve kezeljük a tudományos célú hozzáférés biztosítását is, ami a hivatalos statisztikai célra kezelt adatok felhasználásának másik fő területét alkotja.

A hivatalos statisztikák előállítására céljára felhasznált közvetlen statisztikai adatgyűjtések, illetve másodlagos adatforrások adatvédelmi vonatkozású problémákat is felvethetnek. A következőkben ezeket tekintjük át átfogó jelleggel.

2. Elsődleges és másodlagos adatforrásból származó adatok sajátosságai a felfedés elleni védelem szempontjából

A felfedés elleni védelem egyik kulcsfogalma a felfedési kockázat, ami az azonosítási kockázat, valamint új információ nyilvánosságra kerülésének együttes veszélyét jelenti. A felfedési kockázat megítélése során az egyik legfontosabb szempont, hogy a kérdéses állományban szereplő egyedekről az adatkérő milyen információval rendelkezik, illetve rendelkezhet (azaz milyen, az érintett egyedekre vonatkozó ismerethez fér hozzá). Azzal az információval, ami a „támadó” birtokában van, megtörténhet egy egyed azonosítása, az azonosításon keresztül pedig további olyan információt is megismerhet egy egyedről, amelyről eddig nem volt információja.

A statisztikai célú adatgyűjtés egy részénél témát, vizsgált sokaságot, vonatkozási időt stb. tekintve több forrásból is elérhető információk. Amennyiben a gyűjtött statisztikai adat témájában van hozzáférhető publikus információ, úgy azok szerkezete többnyire eltér a statisztikai célra gyűjtött adatok szerkezetétől. Például a gazdaságstatisztikai adatgyűjtések adattartalma, részletezettsége a statisztika céljainak megfelelően lett kialakítva, míg publikus forrásból bárki számára hozzáférhető a vállalatok bizonyos egyedi információi, jellemzően azok, amelyek az éves beszámolóban szerepelnek, az ott előírt adattartalomban és bontásban, ami nem egyezik meg a statisztikai adatgyűjtés szerkezetével.

Mivel a statisztikai célra kezelt egyedi adatok nem használhatók fel adminisztratív célokra, ezért biztosított, hogy a statisztikai célra gyűjtött adatok egyedi azonosításra alkalmas módon nem férhetőek hozzá.

A tudományos célú felhasználás támogatása keretében kutatók számára a statisztikai célra gyűjtött adatállományokhoz a KSH több csatornán keresztül biztosít tudományos célú

hozzáférést. Minden esetben garantálva azt, hogy az egyedekre vonatkozó információk megismerhetőségének lehetősége minimális.

Egy statisztikai szervezet másodlagos adatforrásai és a saját statisztikai célra gyűjtött adatai adatkezelése közötti különbség egyik legfontosabb oka az, hogy a másodlagos adatforrásból származó adatokat jellemzően nem vagy nem kizárólag statisztikai célra, illetve nem az adott statisztikai szervezet céljaira gyűjtik, ezért más előírások, szabályok határozzák meg védelmük kereteit, mint a közvetlen statisztikai célra gyűjtött adatok esetén. A másodlagos adatforrások gazdáinál eltérő lehet az adathozzáférés gyakorlata, vagy egyáltalán nem biztosítanak hozzáférést az adatokhoz külső személyek, szervezetek részére, vagy egyedi megállapodások alapján csak bizonyos szerveknek adják át az adatokat vagy azok egy körét stb. Ezért a saját statisztikai célra gyűjtött adatokkal szemben a másodlagos adatforrásból származó adatoknál általánosságban nem élhetünk azzal a feltételezéssel, hogy az állományokban lévő információk külső személyek, szervek számára nem ismertek vagy nem megismerhetők, ezáltal egy táblázatos adatkiadás esetén is megnőhet az azonosítás és a felfedés veszélye, ha nincs megfelelő felfedés elleni védelem.

A hozzáférhetőség és a felfedés kockázatának egyik sajátossága, hogy a statisztikai célra kezelt adatokból a Hivatalos Statisztikai Szolgáltatáson kívülre egyedi azonosításra alkalmas állomány nem kerül ki. Fordított helyzetben, a másodlagos adatforrások egyedi azonosítókkal történő átadása a hivatalos statisztika számára működőképes konstrukció, hiszen a statisztikai adatkezelés szabályai garantálják, hogy a statisztikai adat-előállítási folyamatba becsatornázott adatok más célra, egyedekhez köthető információtartalommal ne legyenek felhasználhatók.

Mindezekből következik, hogy a statisztikai kezelésben lévő adatokat – függetlenül azok forrásától, illetve az adatforrás elsődleges vagy másodlagos jellegétől – ugyanolyan jogi, módszertani, informatikai eszközökkel védjük. A különbséget az jelenti, hogy a felfedés kockázatának megítélésekor a másodlagos forrásból származó adatok felfedési kockázata jellemzően magasabb, azaz indokolt lehet az érintett adatok erősebb módszertani védelme.

3. Adathozzáférési csatornák a KSH-ban

A KSH adatai jelenleg több csatornán juthatnak el a felhasználóhoz⁵⁹. A különböző csatornákon különböző statisztikai felfedés elleni védelmi megoldásokat kell, illetve lehet alkalmazni. Erre azért van szükség, mert az egyes csatornák a hozzáférhető adatok részletezettségében, a hozzáféréssel rendelkezők körében, valamint a hozzáférés feltételeiben különböznek egymástól.

A csatornák egészét tekintve elmondható, hogy a felfedés elleni védelem szemszögében a rendező elv az ún. egyenkockázat elve, azaz bármilyen csatornán is történjen a hozzáférés, a felfedési kockázatnak azonosan alacsony szinten kell maradnia. Ezt hivatott biztosítani a jogi, illetve felfedés elleni védelmi eszközök kombinált, ámde eltérő mértékű alkalmazása az egyes csatornákon történő adathozzáférés során. A megfelelő módszer megválasztásakor arra kell törekedni, hogy a felfedési kockázat mértéke és az adatminőségben bekövetkező veszteség egyensúlyban maradjon.

⁵⁹ Részletesen lásd: <http://www.ksh.hu/adatigenyles>

Adathozzáférési csatornák a KSH-ban⁶⁰

Egyedi kérésre összeállítandó táblázatos adatok	Nyilvános mikroadatfájlok	Kutatósobai hozzáférés	Távoli hozzáférés	Távoli végrehajtás	Anonimizált mikroadat-kiadás
– Jogi védelem: Gyenge	– Jogi védelem: Gyenge	– Jogi védelem: Erős	– Jogi védelem: Erős	– Jogi védelem: Erős	– Jogi védelem: Erős
– Módszertani védelem: Erős	– Módszertani védelem: Erős	– Módszertani védelem: Gyenge	– Módszertani védelem: Gyenge	– Módszertani védelem: Gyenge	– Módszertani védelem: Közepes

A különböző adathozzáférési csatornák más-más erősségű jogi, illetve felfedés elleni védelmet igényelnek, és különböző felhasználói igényeket szolgálnak ki.

3.1. Táblázatos adatok

Az egyedi kérésre összeállított táblázatos adatok esetén az adatok védelmét teljes egészében a felfedés elleni védelemnek kell ellátnia, nincs lehetőség az alapvető adatvédelmi szabályozásokat meghaladó jogi védelemre, bár egyes jogi eszközök (pl. adatkérővel történő szerződés-kötés) itt is alkalmazhatók. Módszertani szempontból ebbe a típusba sorolandó a közvetlen nyilvánosságra hozatal is, ami tipikusan az információk honlapon vagy évkönyvekben való megjelentetését jelenti. Ez logikailag nem is lehetne másképp, hiszen a honlapon, kiadványokon keresztül az információ megismerésének lehetősége bárki számára nyitva áll, a KSH-nak az ilyen csatornán elérhető adatok felhasználása felett a nyilvánosságra hozatal után nincs érdemi kontrollja. Éppen ezért a KSH-nak a nyilvánosságra hozatalt megelőzően biztosítania kell, hogy a nyilvánosságra kerülő adatokban egyedekre vonatkozó információ ne szerepeljen. A felhasználó ugyanakkor térítés ellenében kérhet olyan egyedi kérésre összeállított táblázatos adatot, amely adattábla nem található meg közvetlenül az évkönyvekben vagy például STADAT-táblákban. Ebben az esetben a KSH előállítja az adatkérőnek az igényelt táblázatos adatot, amelyhez az adatkérőnek el kell fogadnia a táblázatos adatokra vonatkozó felhasználási feltételeket.

3.2. Nyilvános mikroadatfájlok

A KSH bárki számára hozzáférést biztosít nyilvános mikroadatfájlokhoz is. Ezeket az állományokat is erős felfedés elleni védelemmel szükséges ellátni, mivel ebben az esetben sincs lehetőség a jogi garanciákat bevonni.

3.3. Tudományos célú adathozzáférési csatornák**3.3.1. Anonimizált mikroadatfájlok**

A KSH tudományos célból anonimizált mikroadatot adhat ki kutatók részére. Ebben a csatornában már lehetőség nyílik a felfedés elleni védelem mellett jogi védelem alkalmazására is. Ez a kettő együtt adja az adatok megfelelő mértékű védelmét. Anonimizált mikroadat-kiadás során azért nem alkalmazhatunk pusztán felfedés elleni védelmet, mert az rontaná az adatok minőségét, adott kutatási célra való felhasználhatóságát. Anonimizált mikroadat anonimizálási módszertanának kidolgozása során csak olyan felfedési forгатókönyveket veszünk figyelembe,

⁶⁰ Az egyes adathozzáférési csatornák bemutatását lásd később, illetve részletesebben itt: <http://www.ksh.hu/adatigenyles>

melyek az adott anonimizált mikroadatkerésre érvényesek lehetnek. Ehhez párosul az a jogi védelem, hogy harmadik fél számára nem adható tovább az igényelt mikroadat, csak az adott kutatási célhoz használható fel és az adatkérő csak adott időtartamra jogosult azt használni.

Ebben a csatornában a KSH-nak közvetlen kontrollja van a hozzáférésre kiadott állományok felett. A KSH minden anonimizált mikroadatkerés esetén kutatói akkreditációt végez, azaz vizsgálati szempontoknak veti alá a kutató személyét és a kutatás intézményi hátterét, továbbá megbizonyosodik arról, hogy a hozzáférés tudományos célra történik. Anonimizált mikroadat-állományhoz csak kutató férhet hozzá. A KSH ellátja a kiadandó állomány előzetes felfedés elleni védelmét, hogy statisztikai egységekre vonatkozó információ ne legyen megismerhető, valamint jogi eszközök segítségével biztosítja, hogy az információt az adott kutatókon kívül mások számára ne tegyék megismerhetővé.

3.3.2. Mikroadatokhoz való hozzáférés biztonságos környezetben

Biztonságos környezetben a hozzáférhető adatok részletezettebbek, mint a nyilvános állományokban. Mikroadatokhoz biztonságos környezetben való hozzáférésnek a KSH-ban háromféle módja van: kutatószoba, távoli hozzáférés és távoli végrehajtás.

Kutatószoba és távoli hozzáférés

A kutatószobában és a távoli hozzáférés során a kutató a mikroadatokhoz csak a KSH biztonságos környezetében férhet hozzá. Tényleges mikroadat kiadásra nem kerül sor, az adatkérő a mikroadatot „nem viheti el”, csak a kontrollált környezetben belül rendelkezhet vele. Ebben az esetben a tudományos célú anonimizált mikroadat-kiadáshoz hasonló helyzet áll elő, ám még erősebb jogi védelemre van szükség, mert a kutató ebben a csatornában tud a legrészletesebb adatokkal dolgozni. Ennek megfelelően kevésbé szigorú felfedés elleni védelmet alkalmazunk, lényegében közvetlen azonosítóktól megfosztott állományhoz enged hozzáférést a KSH, bár még ebben az esetben is alkalmazhatók további felfedés elleni módszerek. A KSH számára ezeknek a csatornáknak az a nagy előnye, hogy az adatok nem hagyják el a KSH informatikai környezetét, és az is biztosított, hogy a kutatási eredmények nem tartalmaznak adatvédelmileg aggályos információt, hiszen az eredmények kiadását megelőzően minden esetben sor kerül a kutatási eredmények utólagos adatvédelmi szempontú ellenőrzésére (ún. output checking-re).

Nemzetközi tendenciák szerint jelentősen felértékelődően van a távoli hozzáférés. A KSH-nak jelenleg egy hozzáférési pontja van, a szegedi kirendeltségen. Távoli hozzáférés esetén is ugyanazok az adathozzáférési szabályok érvényesek, mint a kutatószobai környezet esetén, annyi különbséggel, hogy az adott hozzáférési pont a KSH szegedi kirendeltségén helyezkedik el, nem Budapesten, megkönnyítve ezzel a fizikai távolságból eredő adathozzáférést (nem kell Budapestre utazni a kutatószobai környezet eléréseért).

Távoli végrehajtás

Távoli végrehajtás során a kutató programkódot juttat el a KSH-hoz, melyet a KSH munkatársa, a KSH belső, védett hálózatán futtat le, jellemzően mikroadatokhoz kapcsolódva. Távoli végrehajtás esetén a kutató az érintett mikroadat-állományokhoz nem fér hozzá, csak a kész kutatási eredményeket kapja kézhez utólagos adatvédelmi szempontú ellenőrzést követően. Csupán annyi feladata van a kutatónak, hogy az adminisztrációs feladatokat elvégezze

(igénybejelentő adatlap, titoktartási nyilatkozat, szerződés stb.) és megírja a programkódját, minden más a KSH feladata.

4. Másodlagos forrásból származó adatok felfedés elleni védelmének speciális kérdései

A fentiek tükrében a másodlagos forrásból származó adatok felfedés elleni védelmével kapcsolatban az alábbi gyakorlat az irányadó a KSH-ban.

4.1. Tudományos célú hozzáférés

A másodlagos forrásból származó adatok statisztikai felhasználásának leggyakrabban az a célja, hogy a statisztikai adatgyűjtéseket kiváltva vagy mintavételi keretként, pótláshoz vagy validálásra felhasználva statisztikák előállításához alkalmazhatók legyenek, azaz beépüljenek a statisztikai adat-előállítási folyamatba. A KSH a folyamat során bevont minden adathoz biztosítja, hogy az információkat kizárólag statisztikai célra használja, azaz nem fordulhat elő, hogy a statisztikai célra kezelt adat olyan személyekhez, intézményekhez kerül, akik, amelyek nem statisztikai célra használnák fel azokat. Ebből eredően az adatok csak a statisztikák tényleges előállításához kerülnek felhasználásra, illetve a statisztikai célra gyűjtött adatokhoz hasonlóan tudományos célból hozzáférhetővé tehető kutatók számára. Ha az átvett adatokat tovább feldolgozzák, azaz az átvett állományok az átvételkori állapotukhoz képest módosításra kerülnek, úgy ezeket a feldolgozott adatokat lehet adatkérés keretében kiadni, kutatószobában hozzáférhetővé tenni stb.

Az átvett adminisztratív adatokat eredeti, átvételkori állapotuknak megfelelő teljes tartalmukban (teljes változókészletben) alapszabályként csak akkor lehet harmadik félnek kiadni, hozzáférhetővé tenni, ha ezt az adminisztratív adatgazdával megkötött megállapodás az adathozzáférés módjairól és feltételeiről egyértelműen tartalmazza. Ilyen rendelkezés hiányában az adatátvétel keretében felhasznált adminisztratív adatforrás nem tehető hozzáférhetővé, illetve hozzáférésre vonatkozó igény érkezése esetén a hozzáférési céljára külön hozzájárulást kell kérni az adminisztratív adatgazda szervezettől. Ennek alapja, hogy az adatátvétel elsődleges célja statisztikai adatok előállítása és nem pedig egyéb formában történő közzététel (pl. adatkiadás, kutatószobai hozzáférés).

Amennyiben a kutató csak másodlagos adatforrásból származó adatot kíván megismerni, úgy két lehetősége van. Vagy az eredeti adatgazdájához fordul, és az adatgazda által meghatározott feltételek mellett ott kap hozzáférést. Ez esetben a kutató az adatgazda szervezetre érvényes adatkezelési, adatvédelmi szabályok mellett férhet hozzá az adatokhoz. A másik lehetőség – ha az adott másodlagos adatforrást a KSH felhasználja –, hogy a kutató a KSH-tól kér hozzáférést. Amennyiben az adatkérés alapján az adatkérőnek mind a sokaságot, mind pedig a változókat tekintve az adminisztratív adatgazdától átvett állománynak kifejezetten az átvételkori teljes állapota szükséges, úgy az adatgazdával kötött együttműködési megállapodásban foglaltak szerint adható hozzáférés, vagy ilyen rendelkezések hiányában az adatgazdától az adott célra írásos hozzájárulást szükséges kérni.

Az utóbbi esetben, ha az adminisztratív adatgazda megadja az írásos hozzájárulását, a kutató a statisztikai célra gyűjtött adatok kezelésével megegyező gyakorlat szerint, azaz felfedés elleni védelem mellett kaphat csak hozzáférést az adatokhoz, feltéve, hogy a hozzáférési igény

megfelel minden olyan feltételnek, amit a statisztikai célra kezelt adatokkal szemben a KSH általánosságban támaszt (pl. tudományos cél megléte stb.).

4.2. Alkalmazott felfedés elleni módszerek

Nem indokolt az alkalmazott felfedés elleni módszerek megválasztásánál különbséget tenni aszerint, hogy az adatok saját statisztikai adatgyűjtésből vagy másodlagos adatforrásból származnak, hiszen mindkét típusú adat statisztikák előállítását szolgálja. A statisztikai célra kezelt adatoknál minden esetben szükséges a felfedés elleni védelem alkalmazása, amennyiben valamely csatornán adat kerül nyilvánosságra. Mivel a gyakorlatban az adatállományok védettsége különböző felfedés elleni módszerek segítségével is biztosítható, azaz nincs „egyetlen jó megoldás” a felfedési kockázat kellően alacsony szintjének eléréséhez, a felfedés elleni védelmet az adott adatkéréstől függően kell ellátni.

Különböző kutatók különböző célból ugyanazon állományhoz való hozzáférés során is eltérő védelmű állományokhoz férhetnek hozzá. A védelem ellátását ugyanis nagyban befolyásolja, hogy a kutatók az állományok mely mutatóihoz szeretnének hozzáférni, és mely információk a legfontosabbak számukra. Bár a KSH nagyobb felvételeihez (pl. munkaerő-felmérés, közösségi innovációs felvétel stb.) az Eurostat is kidolgozott ajánlásokat a felfedés elleni védelem módszertanához, a gyakorlat azt mutatja, hogy az egyedi igény szerint összeállított állományokra mutatkozik a legnagyobb kutatói igény. Ennek következménye, hogy anonimizált mikroadat-kiadás esetén az alkalmazandó módszerek megválasztása a kért egyedi állomány függvényében alakítható. Biztonságos környezetben való hozzáférés esetén az alkalmazott módszerek megválasztása kevés terhet ró a KSH-ra, hiszen azon csatornában a hangsúly a kutatási eredmények utólagos adatvédelmi szempontú ellenőrzésén van.

4.3. Adathozzáférési csatornák

A másodlagos – különösen az adminisztratív – forrásból származó adatok felfedés elleni védelme esetében hangsúlyosan igaz, hogy a leghatékonyabb eljárás, ha a hozzáférés alapszabály szerint valamilyen biztonságos környezetben történik. Amennyiben ugyanis kombinált állományt ad ki a KSH, azaz olyat, amelyik másodlagos adatforrás(ok)ból és közvetlen statisztikai adatgyűjtés(ek)ből származó információk összekapcsolásával állítható elő, a másodlagos adatforrás megismerésével és a kombinált állománnyal való együtt kezelésével a statisztikai célú adatgyűjtésekből származó eredmények adatszolgáltatókhoz való kötésének valószínűsége megnő.

Elképzelhető, hogy az adatgazda szervezetnél az adott másodlagos adatforrások megismerésének lehetősége jelenleg erősen korlátozott vagy lehetetlen, azaz a másodlagos adatforrás gazdája nem ad hozzáférést harmadik személy, szervezet részére. Ez a gyakorlat azonban bármikor változhat, és a megváltozó hozzáférési feltételek pedig utólag is növelhetik a KSH által már kiadott kombinált állományok felfedési kockázatát, ami az adatgazda számára is nagy kockázatot jelentene, és a hivatalos statisztikával, valamint a KSH-val szembeni bizalmat is csökkentené.

5. Másodlagos forrásból származó adatok felfedés elleni védelmének elvei

A fentiek alapján a másodlagos forrásból származó adatok felfedés elleni védelme során az alábbi alapelveket érvényesítjük.

1. elv: Hozzáférés tudományos célból

A statisztikai célra gyűjtött adatokhoz hasonlóan a statisztikai célra felhasznált másodlagos adatforrásokhoz önmagukban vagy más statisztikai célra kezelt állománnyal összekapcsolva hozzáférést biztosítunk kutatók számára tudományos célból.

2. elv: Hozzáférés biztonságos környezetben

A másodlagos forrásból hozzáférhető adatok hozzáférési gyakorlatának sokrétűsége miatt a statisztikai célra felhasznált másodlagos adatforrásokhoz önmagukban vagy más statisztikai célra kezelt állománnyal összekapcsolva hozzáférést csak biztonságos környezetben biztosítunk. A fenti két elvre érvényes, hogy a másodlagos adatforrásból származó adatokat csak abban az esetben lehet adatkérés keretében kiadni, hozzáférhetővé tenni, ha az átvételkori állapotukhoz képest a statisztikai adat-előállítási folyamat során módosításra kerültek. Az érintett állományokat eredeti, átvételkori állapotukban teljes tartalmukban nem lehet kiadni, hozzáférhetővé tenni, csak abban az esetben, ha az adminisztratív adatgazdával megkötött megállapodás ezt egyértelműen tartalmazza, vagy egyedileg írásban engedélyezi.

3. elv: Megfelelő felfedés elleni védelmi módszerek

A részben vagy egészben másodlagos adatforrásokból származó adatok felhasználásával készült táblázatos adatokat publikálást megelőzően – a statisztikai adatgyűjtések eredményei alapján előállt táblázatos adatoknál alkalmazott gyakorlattal megegyezően – adatvédelmi szempontú ellenőrzésnek vetjük alá.

A statisztikai célra felhasznált másodlagos adatforrásokon vagy más statisztikai célra kezelt állománnyal összekapcsolt adatokon biztonságos környezetben létrejövő kutatási eredményeket adatvédelmi szempontból ellenőrizzük. Az ellenőrzés során azonos szabályokat alkalmazunk minden biztonságos környezetben elérhető állományból készített kutatási eredményre, függetlenül attól, hogy az adatok statisztikai adatgyűjtésekből vagy másodlagos adatforrásokból származnak.

Fogalomtár

Adatgazda: az adatokat gyűjtő, karbantartó intézmény.

Adminisztratív adat: a közigazgatási/közfeladatot ellátó szervezet hazai jogszabályban vagy EU-s követelményben meghatározott igazgatási feladatának végrehajtása során keletkező adat, gyűjtését jogszabály rendeli el. Ide tartoznak a nyilvános, a közhitelű és egyéb nyilvántartások, valamint a hatósági ellenőrzési, gazdasági vagy egyéb tevékenységgel járó adatgyűjtések.

Adminisztratív adatforrás: adminisztratív adatokból felépülő meghatározott adathalmaz, (regiszter, nyilvántartás vagy egyéb adatállomány), metaadathalmaz, adatbázis vagy metaadattár, ahonnan az adatok, illetve metaadatok rendelkezésre állnak.

Adminisztratív adatforrásból származó adat: olyan adat, ami adminisztratív adatforrásból származik függetlenül annak feldolgozottságától, illetve a feldolgozás módjától.

Adminisztratív cél: adminisztratív cél, amikor a közigazgatási szervezet jogszabályban meghatározott igazgatási feladatának végrehajtása érdekében egyedi adatot kezel.

Big data: nagy mennyiségű, nagy sebességű és változatos adatok.

Editálás (szerkesztés): az adatgyűjtés során felvett adatok egyes hibáinak felderítésére és kiigazítására szolgáló eljárás.

Az **imputálás** egy olyan eljárás, melynek során értéket rendelünk valamely specifikus adathelyhez, amelynél a válasz hiányzik vagy nem használható.

Másodlagos adat: az adminisztratív adat, valamint a magánszektorból és a kutatóintézetektől származó adat. (*Use of secondary and mixed sources for official statistics: in-depth review by the Bureau of the Conference of European Statisticians, CES, 59. plenáris értekezlet*). A KSH szemszögéből nézve tágabb értelemben ide tartoznak mindazok az adatgyűjtések, amiket nem a KSH gyűjtött. Idetartozik a Hivatalos Statisztikai Szolgálat más szerveinek az OSAP keretében végrehajtott statisztikai adatgyűjtései, a nem hivatalos statisztikai adatgyűjtések és a vállalati adatgyűjtések is. A másodlagos adatok közé tartoznak a big data adatai is. Jelen kiadvány ez utóbbi értelemben használja a fogalmat, és ebben a vonatkozásban a másodlagos adat szinonimájaként használjuk az átvett adat fogalmát. Ha egy szakstatisztika szemszögéből definiáljuk a másodlagos adatot, ide kell sorolni a fentiekén kívül a más szakstatisztikák által gyűjtött adatokat is.

A **minőség** az ISO 9000/2005 definíció alapján: annak mértéke, hogy mennyire teljesíti a saját jellemzők egy csoportja a követelményeket. („The degree to which a set of inherent characteristics fulfils requirements.”) A statisztikára alkalmazva ez azt jelenti, hogy a hivatalos statisztikai adatok minősége mennyire felel meg a felhasználók igényeinek.

Minőségjelentés a KSH-ban: olyan strukturált formában készített jelentés, amely a statisztikai adat vagy adatfelvétel minőségét írja le a minőség-összetevők mentén, szövegesen és számszerűen kifejezhető minőségindikátorok felhasználásával.

Minőségmutató (minőségindikátor) olyan kvantitatív vagy kvalitatív információ, amely a termék és a folyamat minőségét jelzi.

OSAP: Országos Statisztikai Adatfelvételi Program. A Hivatalos Statisztikai Szolgálat évente összeállított statisztikai célú adatgyűjtési programja.

Statisztikai cél: statisztikai célú az az adatkezelés, ami a sokaság jellemzése céljából történik.

A **statisztikai nyilvántartás** valamely sokaság egyedeit, azok azonosításához, eléréséhez, jellemzéséhez szükséges tulajdonságokat adott időpontra leíró jegyzék. A nyilvántartás vonatkozhat több egymással kapcsolatban álló egyed típusra is (mint például intézmények és telephelyeik).

A **statisztikai regiszter** valamely sokaság egyedeinek azonosítására, elérhetőségére és az adatfelvételi folyamatot támogató jellemzőkre vonatkozó rendszeresen karbantartott nyilvántartás. A regiszterből elérhetők a sokaság aktuális és történeti állapotai, a változások oka, hatálya, forrása. A regiszter a sokaság egyedeire (egységeire) vonatkozó adatokat strukturáltan, adatbázisban (KSH, meta) tárolja.

Tudományos cél: tudományos célú a statisztikai adatok kezelése, mikor a világról, annak társadalmi, gazdasági és természeti jelenségeiről, e jelenségek összefüggéseiről szerzett ismereteink gyarapodnak. (az 1995. évi CXIX. törvény a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről (az ún. direkt marketing törvény) fogalommagyarázata alapján).

Validálás alatt az egymással összehasonlítható adatok (statisztikai adatfelvételek, adminisztratív adatforrások) mikro- vagy makroszinten történő ellenőrzését, szükség szerinti javítását értjük. Az összehasonlítás mikrovalidálás esetén elemi szinten, makrovalidálás esetén aggregált (pl. területi aggregátum) szinten történik.

Felhasznált irodalom

- 1133/2011. (V. 2.) kormányhatározat a vállalkozói adminisztratív terhek csökkentésére irányuló rövid- és középtávú kormányzati programról. www.magyarokzlony.hu/pdf/9026
- Adminisztratív adatforrások statisztikai hasznosításának bővítése a hivatalos statisztikai szolgálat keretein belül. KSH. Munkaanyag. 2008.
- Ajánlás az adminisztratív és hatósági nyilvántartások statisztikai célú felhasználásának elősegítéséhez. KSH. 2011. http://www.ksh.hu/docs/bemutakozas/hun/ost/OST_2_MKB_ajanlas.pdf
- Átvett adatok adat-előkészítése egységes rendszerben – Statisztikai elvárások. A KSH projektterve az „Egységes adatgyűjtés-szervezési, adat-előkészítési és adat-feldolgozási módszerek fejlesztése” című főirányon belül. KSH munkaanyag, 2010. július.
- Brackstone, Gordon J: “Statistical Issues of Administrative Data: Issues and Challenges”, in "Statistical Uses of Administrative Data – An International Symposium", organised by Statistics Canada, 23–25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988).
- Business Registers – Recommendations Manual, Eurostat Methodologies and working papers, European Union, 2010. <http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf>
- Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community. Official Journal of the European Union. L 076, 30/03/1993 P. 0001 – 0011. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993R0696:EN:HTML>
- Daas, Piet – Ossen, Saskia – Arends – Tóth, Judit: Framework of Quality Assurance for Administrative Data. Paper for the 57th session of the International Statistical Institute, Durban, South Africa, 2009. https://www.researchgate.net/publication/268187814_Framework_of_Quality_Assurance_for_Administrative_Data_Sources
- The Eurostat Concepts and Definitions Database (CODED). http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL_GLOSSARY&StrNom=CODED2&StrLanguageCode=EN
- Erdész Tiborné: A személyiadat- és lakcímnnyilvántartás, okmánykiadás összesített adatai, Népszámlálási füzetek (szerk.: Vetési Iván), Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala (KEKKH), 2010. http://www.nyilvantarto.hu/letoltes/kiadvanyok_2010_jan_szoveg.doc
- Az Európai Statisztika Gyakorlati Kódexe. Az Európai Statisztikai Rendszer Bizottsága, 2011. http://www.ksh.hu/docs/bemutakozas/hun/az_europai_statisztika_gyakorlati_kodexe.pdf
- Harcza István: Műhelykonferencia az adminisztratív adatállományok statisztikai célú hasznosításáról – különös tekintettel a népszámlálásra. Statisztikai Szemle, 2010. május. http://www.ksh.hu/statszemle_archive/2010/2010_05/2010_05_549.pdf
- Jelentés a magyar statisztika európai uniós auditjáról, 2015. július. http://www.ksh.hu/docs/bemutakozas/hun/audit_jelentes_2015.pdf
- Központi Adatfogadó és továbbító Rendszer Átvett állományokhoz (KARÁT) – Statisztikai elvárások. A KSH projektterve „Az adminisztratív nyilvántartások, adatforrások statisztikai célú felhasználásának bővítése” című főirányon belül. KSH munkaanyag, 2010. június.

A KSH minőségpolitikája, 2014.

http://www.ksh.hu/docs/bemutakozas/hun/ksh_minosegpolitikaja_2014.pdf

Laney, Douglas: 3D Data Management: Controlling Data Volume, Velocity and Variety. Gartner. Retrieved 6 February 2001.

<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

Minőségi irányelvek a Központi Statisztikai Hivatal statisztikai folyamataira.

http://www.ksh.hu/docs/bemutakozas/hun/minosegi_iranyelvek_2014.pdf

National Quality Assurance Frameworks

<http://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx>

van Nederpelt, Peter – Daas, Piet: 49 Factors that influence the Quality of Secondary Data Sources.

Statistics Netherlands. The Hague, 2012. <http://unstats.un.org/unsd/dnss/docs-nqaf/49%20Factors%20that%20influence%20the%20Quality%20of%20Secondary%20Data%20Sources.pdf>

Népmozgalmi kézikönyv, KSH, 2015.

http://www.ksh.hu/2016_torveny_altal_elrendelt_adatgyujtesek

Population and housing censuses. United Nations Statistics Division.

<http://unstats.un.org/unsd/demographic/sources/census/alternativeCensusDesigns.htm>

Principles and Recommendations for Population and Housing Censuses. Revision 2. Department of Economic and Social Affairs Statistics Division. United Nations. New York, 2008.

http://unstats.un.org/unsd/publication/seriesM/seriesm_67Rev2e.pdf

Register-based Statistics in the Nordic Countries - Review of Best Practices with Focus on Population and Social Statistics, United Nations, New York and Geneva, 2007.

http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

Resolution adopted by the Economic and Social Council on 24 July 2013 2013/21. Fundamental Principles of Official Statistics, United Nations. <http://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>

SDMX content-oriented guidelines. annex 4: metadata common vocabulary.

http://unstats.un.org/unsd/dnss/docs-nqaf/04_sdmx_cog_annex_4_mcv_2009.pdf

Sindoni, Giuseppe – Allen, John – Junker, Claudia: Overview of the XBRL Pilot Project, Eurostat, 2008.

<https://www.yumpu.com/en/document/view/17356474/overview-of-the-xbrl-pilot-project-circa-europa>

Statisztikai és Informatikai Doktár, KSH belső rendszere.

Szóke Katalin: Big data a statisztikában (munkaanyag). KSH, 2015.

Use of Administrative Sources for Business Statistics Purposes – Handbook on Good Practices, Eurostat, 2000.

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/CA-24-99-897/EN/CA-24-99-897-EN.PDF

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

Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices. UNITED NATIONS. New York and Geneva, 2011.

<http://unstats.un.org/unsd/EconStatKB/Attachment442.aspx>

Wallgren, Anders and Britt (Statistics Sweden): Register-based Statistics. Administrative Data for Statistical Purposes. Wiley Series in Survey Methodology, John WILEY & Sons, Ltd, England, 2007.

Wallgren, Anders and Britt: To understand the Possibilities of Administrative Data you must change your Statistical Paradigm! Joint Statistical Meetings (JSM) - Section on Survey Research Methods, Miami Beach, Florida, 2011.

http://www.amstat.org/sections/srms/proceedings/y2011/Files/300347_64422.pdf

What does “Big Data” mean for official statistics? UNECE, 2013.

<http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170614>

XBRL tájékoztató információk. Magyar Tőkepiaci Szakemberek Egyesülete.

http://www.capmarketprof.hu/text/doc/text21013101001/XBRL_tajekoztato_informaciok.doc

Mellékletek

1. melléklet: A másodlagos adatforrások hasznosítási területei a KSH-ban⁶¹

A KSH által átvett mintegy 200 adminisztratív adatforrást tekintve a legtöbb, szám szerint 40 adatállományt, adat-összeállítást a Nemzeti Adó- és Vámhivatal (NAV) ad át a KSH számára. A második helyen az Országos Egészségbiztosítási Pénztár (OEP) és a Nemzeti Bank (MNB) áll 18–18 adatátvétellel. A Magyar Államkincstár (MÁK) 17, a Nemzetgazdasági Minisztérium (NGM) 13 adatkört biztosít a hivatal számára. 19 szervezet egyenként 2–9, és 29 intézmény 1–1 tétellel járul hozzá a KSH adatforrásaihoz.

Az OSAP keretén belül átadott adatkörök számát figyelembe véve a KSH két „legnagyobb” adatátadó partnere az Emberi Erőforrások Minisztériuma (28) és a Nemzeti Fejlesztési Minisztérium (25). A Belügyminisztériumtól 10 adatállományt kap a hivatal, és a további hat szervezettől egyenként legfeljebb 4-et. Az átvett adatállományokat adatgazda szerint részletezve az 2. melléklet tartalmazza.

Jelen áttekintés a KSH egésze szempontjából veszi számba az adatforrásokat. Egy-egy statisztikai terület azonban nemcsak közvetlen adatgyűjtéseket, illetve adatátvételeket használ, hanem más szakstatisztikáktól is átvesz olyan adatot, ami az ottani feldolgozás eredménye. Az alábbiakban szakstatisztikáinként csak azokat a másodlagos adatforrásokat említjük meg, amit az a terület közvetlenül alkalmaz, de nem szólnak azokról, amik egy másik szakstatisztika közvetítésével válnak adatforrássá. Az alábbi struktúra a „regiszterek, statisztikai nyilvántartások karbantartása” kivételével a szakstatisztikák nomenklatúráját követi, de csak azokat az elemeket ismerteti, amelyek adatforrásai között külső szervezettől átvett másodlagos adatok is szerepelnek.

Az alábbi ismertetés célja, hogy bemutassuk, az adatátvételek hogyan hasznosulnak a KSH statisztikai adat-előállításának folyamatában. Tekintettel azonban arra, hogy az egyes szakstatisztikák által használt adatforrások változhatnak, továbbá az illetékes adatgazdák is módosulhatnak, szükséges leszögezni, hogy a másodlagos adatforrások alábbi számbavétele a 2016. júniusi állapotot tükrözi. A szakstatisztikákhoz tartozó leírások terjedelme semmiféle kapcsolatban nincs azoknak a gazdaság és társadalom leírásában betöltött fontosságával, kizárólag a felhasznált adatforrások számát tükrözi.

1. Regiszterek, statisztikai nyilvántartások karbantartása

A statisztikai regiszterek és nyilvántartások⁶² általában egy vagy több adminisztratív adatforrásra épülnek. A statisztikai felhasználás igényeit kielégítő, további szükséges információk vagy olyan statisztikai adatgyűjtésből származnak, ami kifejezetten a karbantartást szolgálja, vagy a regiszter/nyilvántartás szempontjából használható adatokat tartalmazó egyéb statisztikai felmérésből.

A KSH-ban kialakított és karbantartott regisztereket és nyilvántartásokat témájuk, felhasználási területük, illetve karbantartásuk módja szerint a II. fejezet részletezi.

⁶¹ 2016. júniusi állapot.

⁶² Statisztikai egységek olyan nyilvántartása, amely nem éri el a regiszter kritériumait. Lásd II. fejezet!

2. Népeség, népmozgalom

2.1. Népszámlálás

A KSH a 2011. évi népszámlálás során sokrétűen használta fel a KEKKH kezelésében lévő népeség-nyilvántartás adatait. (Lásd V. fejezet)

2.2. Népmozgalom, népesség-továbbvezetés

A népmozgalmi statisztika⁶³ alapját sok ország gyakorlatával ellentétben Magyarországon a KSH által lebonyolított teljes körű statisztikai adatgyűjtések alkotják, amelyek azonban nem függetlenek a hivatalos nyilvántartásoktól. A kérdőívek kitöltéséért az illetékes hatóság képviselője a felelős, és a szolgáltatott adatok között szereplő azonosító az anyakönyvi folyószám. A kérdőíven felsorolt alapinformációk megfelelnek a hivatalos regisztráció tartalmának kiegészítve olyan adatokkal, amik kizárólag statisztikai célokat szolgálnak. A halottvizsgálati bizonyítvány esetében oly mértékben összekapcsolódik az adminisztratív és a statisztikai funkció, hogy egy hat példányos kérdőíven történik az adatszolgáltatás, illetve a hatósági igazolás az érintett személyek, illetve intézmények felé. A nyomtatvány átírható vegypapíron készül úgy, hogy az egyes példányok tartalma ne sértse az adatvédelmi szabályokat.

Míg a többi természetes népmozgalmi esemény mindegyikét anyakönyvezik, a magzati halálozásokat nem, ezért indokolt, hogy a statisztikai úton gyűjtött adatok érvényesítéséhez felhasználják az OEP a terhességmegszakításokra és a magzati halálozásokra vonatkozó adatait is.

A vándorlásstatisztika bázisát adminisztratív adatforrások alkotják. A belföldi vándorlás indikátorai a lakó, illetve tartózkodási hely bejelentettsége alapján készülnek, amihez a KEKKH szolgál lakcímváltozásra vonatkozó adatokkal.

A nemzetközi migráció statisztikája is adminisztratív adatokra épül. A KEKKH személyiadat- és lakcímnyilvántartása a Magyarországon lakcímrel rendelkező külföldiek, és a külföldre távozó, illetve visszatérő magyarok bejelentett adatait biztosítja, a Bevándorlási és Állampolgársági Hivatal pedig valamennyi Magyarországra bevándorló, innen kivándorló, illetve itt tartózkodó külföldi adatait. Ezeket egészítik ki az OEP által kezelt, tajszámmal rendelkező külföldi állampolgárok adatai.

A népesség számának két népszámlálás közötti statisztikai becslése a népmozgalmi események egyenlege alapján készül, így ugyanazok a külső adatforrások kerülnek felhasználásra, amiket az előzőekben felsoroltunk.

⁶³ A népmozgalmi statisztika körében megfigyelésre kerül minden Magyarországon bekövetkezett születés, haláleset, házasságkötés, bejegyzett élettársi kapcsolat létesítése, élettársi nyilatkozat nyilvántartásba vétele, házasság felbontása, valamint érvénytelené nyilvánítása, bejegyzett élettársi kapcsolat felbontása, megszüntetése, a nyilvántartott élettársi kapcsolat megszűnése, valamint lakcímváltozás. (1993. évi XLVI. törvény a statisztikáról. 10. § (1) bek.) A szakstatisztikai rendszerben ide soroljuk még a magzati veszteségeket és a nemzetközi vándorlást.

3. Társadalom

3.1. Munkaerőpiac

3.1.1. Alkalmazásban állók, keresetek

Az alkalmazásban állók és keresetek statisztikája gazdag adatforrásrendszerre épül, és az adatgyűjtések mellett több másodlagos adatforrást is felhasznál. Ezek közül a legjelentősebb a MÁK illetmény-számfejtési rendszerének adatai, amelyből a KSH a helyi és központi költségvetési szervek munkaügyi adatait (létszám, keresetek, munkaóra, munkaerőköltség) veszi át. Emellett, a MÁK statisztikai célokból begyűjti és átadja a KSH-nak a költségvetési szervek üres álláshely adatait is. Ebben a szakstatisztikában felhasználásra kerülnek a NAV nyilvántartásaiból átvett adatok is a munkavállalók és egyéni vállalkozók járulékaikról, valamint a családi adókedvezményről. A KSH másodlagos adatforrásként hasznosítja a NGM statisztikai adatgyűjtésének az egyéni bérekről és keresetekre vonatkozó adatait is.

3.2. Önkormányzati ingatlanvagyon

Az önkormányzatok az általuk nyilvántartott ingatlankataszter alapján szolgáltatják az adatokat a tulajdonukban lévő ingatlanvagyonról. A KSH a MÁK Somogy Megyei Igazgatóságától kapja az összesített adatokat.

3.3. Lakáscélú hitelek

A szakstatisztika felhasználja a NAV által az ingatlan-adásvételi illetékügyek kezelése során gyűjtött adatokat.

3.4. Lakáspiaci árak

A lakáspiaci árak alakulásának megfigyelése az illetékkiszabási eljárásban rögzített adatokon alapul. Az anonimizált adatokat a NAV havonta bocsátja a KSH rendelkezésére. Az átadott adatok valamennyi, a magánszemélyek által kötött és a teljes lakás átruházásával járó adásvételre vonatkozóan tartalmazzák a lakások eladási ára mellett azok legfontosabb adatait is.

3.5. Lakásépítési költségindex

Az egyes anyagköltségek az ÉGSZI Senior Kft. építőipari anyagárak adatbázisából származnak.

3.6. Egészségügy, baleset

Az egészségügyi statisztika lényegében a nyilvántartások fejlődésével együtt alakult, ezért az adminisztratív és statisztikai célra használt adatforrások a kezdetek óta jelentős átfedést mutatnak. Korábban a regisztereket vezető egészségügyi intézmények állították elő a statisztikai indikátorokat is, és hosszú időn keresztül az egészségügyért felelős minisztérium fogta össze az ágazat statisztikai tevékenységeit. Az informatikai lehetőségek fejlődésével ma már minden közfinanszírozott szolgáltatást elektronikusan tartanak nyilván, és ezek az információk a statisztika potenciális adatforrásai.

3.6.1. Egészségügyi alapellátás

Az Egészségügyi szolgáltatók és működési engedélyük nyilvántartása, melyet az Országos Tisztifőorvosi Hivatal (OTH) küld meg a KSH részére, a házi orvosokra vonatkozó statisztikai adatgyűjtés felvételi keretként szolgál. Az OEP-nek a házi orvos és házi gyermekorvosi praxisok betegforgalmi és morbiditási adatait a szakstatisztika a közvetlen adatgyűjtés kontrolljaként, a hiányzó adatok pótlására használja. Az NGM foglalkozás-egészségügyi alapszolgálatok munkájára vonatkozó statisztikai adatgyűjtése is újrahasznosításra kerül a szakstatisztika keretében.

3.6.2. Egészségügyi szakellátás, kórház

A járó- és a fekvőbeteg szakellátás statisztikájának adatforrásai az OEP adminisztratív adatállományai, melyekből a KSH évente vesz át adatokat. A felhasználók ennek alapján kapnak képet a kórházak kapacitásáról, betegforgalmáról és a perinatális intenzív centrumokról.

A gondozóintézeti ellátás statisztikája több, éves gyakoriságú adminisztratív adatátvételeken alapul. Az Országos Korányi Tbc és Pulmonológiai Intézetől a tüdőbeteg-gondozókról, az OEP-től a bőr és nemibeteg, valamint a pszichiátriai, az OTH-tól pedig az addiktológiai gondozókról érkeznek adatok. Az Országos Epidemiológiai Központtól (OEK) érkeznek az adatok a bőr- és nemibeteg-gondozókban kezelt szexuális úton terjedő betegségekben szenvedők gondozásáról, míg a kábítószer fogyasztókról és kezelésükről az Emberi Erőforrások Minisztériuma (EMMI) biztosít adatot.

Az otthoni szakápolás adatforrása szintén az OEP adminisztratív nyilvántartása, amiből éves gyakorisággal érkeznek az adatok.

A mentőszolgálati statisztika is az OEP adatait hasznosítja a KSH saját adatgyűjtései mellett.

A gyógyszerügyi statisztika az EMMI illetékességébe tartozó statisztikai adatokból dolgozik, melyek az OTH-tól és az OEP-től származnak.

3.6.3. Egészségügyi ellátórendszer (személyzet, eszközellátottság, vérellátás)

Az Egészségügyi Nyilvántartási és Képzési Központ (ENKK) szolgáltató saját igazgatási nyilvántartásaiból adatot az orvosok, fogorvosok, gyógyszerészek, pszichológusok egészségügyi szakvizsgáiról, az orvosok, fogorvosok, gyógyszerészek létszámáról, területi, szakképesítés, kor és nemek szerinti megoszlásáról, az Országos Vérellátó Szolgálat pedig a vérkészítmények előállításáról és felhasználásáról.

3.6.4. Megelőzés, védőnők

A szakstatisztika adminisztratív adatforrásai az OTH-tól származó védőnői jelentések összesítése.

3.6.5. Betegségek, balesetek

A bejelentésre kötelezett és a népegészségügyi szempontból kiemelt jelentőségű betegségek, egészségkárosodások előfordulásával kapcsolatos igazgatási adatok nyilvántartásait az OTH, az OEK, az Országos Onkológiai Intézet, az Országos Korányi TBC és Pulmonológiai Intézet, a Nemzeti Rehabilitációs és Szociális Hivatal (NRSZH), valamint az OEP vezeti, és az adatokat statisztikai hasznosításra átadja a KSH-nak.

Az adminisztratív adatok mellett a szakstatisztika felhasználja az EMMI, a Nemzetgazdasági Minisztérium (NGM) és a Belügyminisztérium (BM) OSAP keretében összegyűjtött adatait is.

3.6.6. Egészségügyi elszámolások

Az egészségügyi ellátórendszer pénzügyi erőforrásaira vonatkozó statisztikák előállításához a KSH külső adatforrásokra is támaszkodik. A költségvetési adatok forrása a MÁK, az Egészségbiztosítási Alap adatai az OEP. Az egészségügyi bér- és létszámadatokat az EMMI által elrendelt statisztikai adatgyűjtés biztosítja, míg az önkéntes egészségpénztárak adatai az MNB honlapján érhetők el.

3.7. Szociális védelem

3.7.1. Pénzbeli ellátások, támogatás, segély

A szociális statisztika – az egészségügyihez hasonló fejlődési folyamat után – nagyszámú adminisztratív adatforrást használ fel. A pénzbeli ellátások statisztikai rendszeréhez az ellátásokat biztosító szervezetek, azaz az ONYF, a MÁK, az OEP és a nyugdíjpénztárak adnak át adatokat nyilvántartásaikból.

3.7.2. Szociális szolgáltatások

A szociális és gyermekvédelmi szolgálatokról és intézményekről az NRSZH ad át a KSH részére adminisztratív adatokat.

3.7.3. Szociális védelmi elszámolások (ESSPROS)

A szociális védelmi elszámolások adatforrását részben az állami költségvetésben megjelent szociális célra fordított kiadási tételek jelentik. Ahol a funkcionális bontáshoz részletesebb adatokra van szükség, ott az egyes intézmények: önkormányzatok, ONYF, OEP, az illetékes minisztériumok hivatalosan publikált adatai is felhasználásra kerülnek.

3.8. Oktatás

Az oktatásstatisztikán belül az iskolarendszerű oktatás adatainak összegyűjtéséért az ágazatot felügyelő minisztérium a felelős.

3.8.1. Köznevelés

A KSH átveszi a szakterület statisztikai célú adatgyűjtésének teljes állományát az EMMI-től, és ezek alapján készíti a statisztikai mutatószámokat.

3.8.2. Felsőoktatás

2013-tól a felsőoktatásra vonatkozó adatok a Felsőoktatási Információs Rendszerből (FIR) származnak, amely egy közhiteles központi nyilvántartás az Oktatási Hivatal kezelésében.

Az Oktatási Hivatal bonyolítja a központi felsőoktatási felvételi eljárást, és vezeti a kapcsolódó nyilvántartást, amiből a KSH a felsőoktatásba jelentkezőkre és a felvettekre vonatkozó adatokat veszi át.

3.9. Kultúra, sport

3.9.1. Kulturális intézmények, tevékenységek

A közművelődési intézmények, könyvtárak, levéltárak, múzeumok, színházak stb. adatait magába foglaló szakstatisztikai terület az illetékes minisztérium (EMMI) statisztikai adatgyűjtéseire épül.

3.10. Igazságszolgáltatás

Az igazságszolgáltatási statisztikán belül a bíróságok ügyforgalmára, a regisztrált bűncselekményekre és elkövetőikre, a vádlottakra és a jogerősen elítéltekre, illetve a fogvatartottakra vonatkozó adatgyűjtéseket az Országos Bírósági Hivatal, a BM, a Legfőbb Ügyészség, illetve a Büntetés-végrehajtás Országos Parancsnoksága (BvOP) végzi. A KSH az általuk feldolgozott statisztikai adatokat hasznosítja. A BM felelősségi körébe tartozó szabálysértési adatok adminisztratív nyilvántartását, az egységes szabálysértési nyilvántartó és eljárás támogató rendszert (ESZNYER) a KEKKH üzemelteti, ami egyúttal a statisztika adatforrása is.

4. Általános gazdasági mutatók

4.1. Nemzeti számlák, GDP

A nemzetközileg elfogadott definíció szerint a nemzeti számlák rendszere az ország gazdasági tevékenységét összefoglaló makrogazdasági statisztikai számlarendszer, amely a forrásokat összeveti a felhasználással. Ennek megfelelően, adatforrásrendszere sokrétű, számos igazgatási adatot használ fel statisztikai adatgyűjtések eredményeivel kiegészítve.

4.1.1. A GDP termelése

A GDP termelésére vonatkozó számítások nagymértékben támaszkodnak a gazdasági szervezetek regiszterére (GSZR) kihasználva, hogy az minden adószámmal rendelkező egységet tartalmaz. A nem pénzügyi vállalatok és a kormányzati szektor számláinak összeállítása emellett főleg egy hivatali fejlesztésű harmonizált és konzisztens vállalati adatbázist használ, ami részben a vállalatok társaságiadó-bevallásaira épül, de statisztikai adatforrásokból származó adatokat is integrál. A pénzügyi vállalatok szektor legfontosabb adatforrásai az adóbevallások és a Magyar Nemzeti Bank (MNB) által az ún. egyéb monetáris intézményekről, a biztosítótársaságokról és a nyugdíjalapokról gyűjtött adatok. További külső adatforrásként kerülnek felhasználásra az MNB, a hitelintézetek, a lakás-takarékpénztárak, a biztosító társaságok és egyesületek, a nyugdíjalapok és a garanciaalapok mérlegei és eredménykimutatásai, a Nemzeti Adó- és Vámhivatal (NAV) társaságiadó-bevallások adatai, valamint a központi kormányzat és a helyi önkormányzatok jelentéseiből származó adatok.

A háztartások szektor adatforrása egyrészt a GSZR, másrészt az egyéni vállalkozók személyi jövedelemadója, az egyszerűsített vállalkozói adó és az általános forgalmi adó bevallásainak adatbázisa, illetve az egyéni vállalkozói körben végzett jogerős személyi jövedelemadó- és áfaellenőrzések anonimizált eredménye.

A háztartásokat segítő nonprofit intézmények szektor számításához használt másodlagos adatforrások a politikai pártok által a Cégekülönyben közzétett pénzügyi beszámolók.

4.1.2. A GDP felhasználása

A GDP felhasználása jobbára más szakstatisztikák által előállított adatokat hasznosít, de az adatforrásai közé tartoznak a járóbeteg-ellátás, valamint az aktív és krónikus fekvőbeteg-ellátás OEP által összegyűjtött teljesítményadatai.

4.1.3. Jövedelemszámlák

A jövedelmek keletkezésére, elosztására és felhasználására vonatkozó statisztikai mutatók számítása kizárólag adminisztratív adatforrásokon alapul. Az MNB a tőkebefektetésekre vonatkozó adatokat, valamint a pénzügyi vállalkozások és pénzügyi tevékenységet folytató egyéb szervezetek mérleg- és eredménykimutatásait, az adóhatóság pedig a különböző adófajtákra vonatkozó adatokat biztosítja a jövedelmi számlák összeállításához.

4.1.4. Külföldi működő tőke

Az MNB közvetlen tőkebefektetésekre és tőkekifektetésekre vonatkozó adatai hasznosulnak a közvetlen külföldi tőkebefektetések adatainak meghatározásához.

4.1.5. ÁKM⁶⁴, forrás- és felhasználástáblák

A szakstatisztika szerteágazó adatforrásai között két NAV-nyilvántartást találunk: a vámteheradatokat és a jövedéki termékek termelésére vonatkozó adatokat.

4.2. Gazdasági és nonprofit szervezetek, beruházások

4.2.1. Nonprofit szervezetek

A civil szervezetekről biztosít adminisztratív adatokat az Országos Bírósági Hivatal (OBH). Az egyéb nonprofit szervezetek esetében – a többi gazdálkodó szervezethez hasonlóan – a cégbíróság által vezetett nyilvántartás adatai hasznosulnak a KSH statisztikai rendszerében.

4.2.2. Vállalkozások teljesítménye

Ez a szakstatisztika is zömében a fenti külső adatforrásokat hasznosítja a nem pénzügyi vállalkozások vonatkozásában: a társas és egyéni vállalkozások adatait (összesen 10 féle adóbevallás), melyek közül kiemelt jelentőségű a társasági adóbevallás részét képező és eredménykimutatás.

A teljes éves gazdaságszerkezeti statisztikai mutatórendszer mintegy 97%-a adminisztratív adatforráson alapul.

4.2.3. Beruházás

A beruházás-szakstatisztika éves területi mutatói, valamint a vállalkozások szerkezeti statisztikai esetében a KSH a NAV társaságiadó-bevallásokból származó adatait használja a statisztikai adatok pótlásához, teljeskörűsítéséhez és publikálásához. Adatellenőrzéshez és a meg nem figyelt statisztikai egységek adatainak becsléséhez a miniszterelnökségnek az uniós támogatásokból megvalósult projektekre vonatkozó adminisztratív adatgyűjtéseiből, az építési e-naplóból származó adatok és a MÁK által kezelt pénzforgalmi adatok is felhasználásra kerülnek.

⁶⁴ Ágazati kapcsolatok mérlege.

4.3. Külkereskedelem–statisztika

4.3.1. Külkereskedelmi termékforgalom

A külkereskedelmi termékforgalom statisztikája felhasználja a NAV adminisztratív nyilvántartásait, az EU-n kívüli országokkal folytatott kereskedelem statisztikája az egységes vámokmány adatait, az Európai Közösség területén belül történő közösségi termékértékesítések és termékbeszerzések statisztikája pedig az áfabevallás és az összesítő nyilatkozat adatait.

4.3.2. Szolgáltatás–külkereskedelem

A szakstatisztika adatforrásai között szerepelnek az MNB nyilvántartásai a viszonzatlan átutalásokról és az eszmei, szellemi javak szolgáltatásáról, a rezidens nem banki ügyfelek külföldi fizetési forgalom miatti jóváírásairól és terheléseiről, valamint a pénzügyi közvetítési szolgáltatások közvetett módon mért díjáról (FISIM). Emellett a KSH felhasználja a diplomáciai és konzuli képviselők, nemzetközi szervezetek fogyasztásával kapcsolatos adóalap-nyilvántartást, és a vámbeszédési költségek beszámolóit, az áfa bevallás és az összesítő nyilatkozat adatait, melyeket a NAV-tól vesz át. A Nemzeti Média- és Hírközlési Hatóság (NMKH) biztosít adatokat a közvetlen magyarországi filmgyártási költségekről. A nemzetközi üzleti és szállítási szolgáltatások statisztikai számbavételéhez az elsődleges felvételi keretet a NAV szolgáltatásokra vonatkozó áfa- és társaságiadóbevallás-adatai biztosítják. Az NMKH állami támogatásra jogosult és támogatásban nem részesülő filmalkotások, valamint a költségellenőrzési eljárásban hozott jogerős határozatok nyilvántartásai, továbbá a Nemzeti Közlekedési Hatóság (NKH) Közösségi Árufuvarozói Engedély nyilvántartása is a felvételi keret része. Ezt eseti jelleggel más adatforrások is kiegészíthetik, közöttük az MNB nyilvántartásai és az EMMI egyes statisztikái

4.4. Energia és környezet

Az energiagazdálkodással kapcsolatban a Magyar Energetikai és Közmű-szabályozási Hivatal (MEKH) biztosít adatokat az ipari szektor energiamérlegéről, az energiahordozók felhasználási mérlegéről és az energiahordozók értékesítési mérlegéről, továbbá a távhőtermelők és távhőszolgáltatók távfűtéséről és távhűtéséről, illetve az energiahordozók termeléséről, importjáról és exportjáról. Az adatokat a MEKH az OSAP keretében gyűjti.

4.5. Árak

Az építőipari termelői árak statisztikai rendszerének forrásául a NAV társasági adóadatai szolgálnak. A Mezőgazdasági ráfordítási árak területén hasznosul az Agrárgazdasági Kutató Intézet (AKI) statisztikai adatgyűjtése a műtrágya-értékesítésről, valamint a Vetőmag Szövetség és TermékTanácsnak (VSZT) a vetőmagtermelésre vonatkozó adatai.

A szakstatisztika céljai között szerepel a vásárlóerő-paritás számítása és a nemzeti fizetőeszközök vásárlóerejének mérése az árak nemzetközi összehasonlítása alapján, és így a felhasznált adatforrások összessége szerteágazó. Az egészségügyi árak statisztikai becsléséhez az OEP szolgáltató adatot a kórházi szolgáltatásokról.

5. Gazdasági ágazatok

5.1. Mezőgazdaság

5.1.1. Földhasználat

Az agrártámogatásban részesülő gazdálkodók nyilvántartása alapján az Mezőgazdasági és Vidékfejlesztési Hivatal (MVH) évenként ad információt a támogatásra kifizetett összegekről jogcímeik, terület, a hasznosítás jellege és a támogatott gazdálkodók szerint.

5.1.2. Mezőgazdasági számlarendszer

A mezőgazdasági számlarendszer adatforrás készlete – a szakterület jellegéből fakadóan – igen gazdag. Számos, a KSH által végzett statisztikai adatgyűjtés és más statisztikai eredmények mellett adminisztratív nyilvántartások adatait is felhasználja. Közöttük az egyik a földhasználat kapcsán már ismertetett adatforrás. A szakstatisztikai terület felhasználja a Hegyközségek Nemzeti Tanácsa (HNT) adatait is a szőlőtermés, borkészletek mennyiségéről, minőségéről és a fajtacsoportok szerinti megoszlásukról. További információt biztosít az adóhatóság nyilvántartása is a jövedéki termékek termeléséről.

A Mezőgazdasági ráfordítási árak számításához az Agrárgazdasági Kutató Intézet (AKI) szolgáltat adatot a műtrágya-értékesítésre vonatkozó statisztikai adatgyűjtéséből, továbbá felhasználásra kerül a VSZT már említett nyilvántartása a vetőmagtermelésről.

5.1.3. Állatállomány és állati termékek

A szakstatisztika számos statisztikai célú adatgyűjtésből építkezik, amikhez felvételi keretként szolgál az MVH nyilvántartása az agrártámogatásban részesülő gazdálkodókról, továbbá az Nemzeti Élelmiszerlánc-biztonsági Hivatal (NÉBiH) ún. BIR-rendszere a baromfikeltető állomásokról és az ENAR rendszere a szarvasmarha-állományról. Az utóbbi az állatállomány számának meghatározásához is forrásul szolgál.

5.1.4. Ültetvénytisztika

Az ültetvények statisztikai összeírását a gyümölcsös ültetvények esetében 5 évente, a szőlőültetvények esetében 10 évente kell végrehajtani.

A 2012. évi gyümölcsös ültetvények összeírása során a felvételre kijelölt ültetvények kétféle forrásból álltak elő, a korábbi statisztikai adatgyűjtések mellett a zöldség-gyümölcs terméktanács (FRUITVEB) új telepítések bejelentésére vonatkozó nyilvántartása képezte.

A 2009. évi szőlőültetvény-összeírás felvételi keretének alapját a HNT által vezetett és naprakészen nyilvántartott szőlőkataszter biztosította

5.1.5. Erdőgazdálkodás

A szakterület a NÉBiH Erdészeti Igazgatóságától átvett adatait hasznosítja, amelyek alapja az FM által az OSAP-ban elrendelt statisztikai adatgyűjtések összessége az erdősítésekről, a fakitermelésről, az erdei fatermékek, a faipari alapanyagok, félkész- és késztermékek termeléséről és készletváltozásáról, valamint az Országos Erdőállomány Adattár.

5.1.6. Halászat

A halászati indikátorok számítása az AKI statisztikai adatgyűjtése alapján történik. A „lehalászás” elnevezésű felvétel adatszolgáltatója valamennyi halastavat vagy intenzív halgazdaságot üzemeltető halgazdaság vagy jogi személy.

5.1.7. Mezőgazdasági gazdaság szerkezet

A mezőgazdaság strukturális változásainak megfigyelése két, a KSH által lebonyolított statisztikai adatgyűjtésen és néhány adminisztratív adatforráson alapul.

Az utóbbi csoportba tartoznak a biogazdálkodásra vonatkozó adatok, amiket a KSH a Biokontroll Hungária Kft. és a Hungária Öko Garancia Kft. adatbázisából veszi át, továbbá az MVH Integrált Igazgatási és Ellenőrzési Rendszer adatai, a szőlőregiszter adatai a HNT-től és a geokódok a Földmérési és Távérzékelési Intézettől.

5.1.8. Építőipari termelői árak

A szakstatisztika a társaságiadó-bevallások adatait veszi át az adóhatóságtól.

5.2. Belkereskedelem

5.2.1. Belkereskedelmi egységek, üzletek

A szakstatisztika célja a KSH kereskedelmi egységeket és tevékenységeiket nyilvántartó regiszterének karbantartása, amihez az önkormányzatok szolgáltatnak adatot jogszabályi kötelezettség mellett a KSH-val egyeztetett egységes szerkezetű nyilvántartásaik alapján a kereskedelmi tevékenységet folytató egységek tárgyidőszakban bekövetkezett változásairól.

5.2.2. Kiskereskedelmi forgalom üzlettípusonként

A KSH az áfaadatok mellett felhasználja a Nemzeti Dohánykereskedelmi Nonprofit Zrt. adatait is a dohányboltok dohányforgalmáról. Emellett vizsgálatokat folytat a NAV online pénztárgépeken keresztül gyűjtött kiskereskedelmi forgalmi adatok hasznosítási lehetőségeiről is.

5.3. Turizmus, vendéglátás

5.3.1. Szálláshely-szolgáltatás

A szakstatisztika főként közvetlen adatgyűjtéseken és statisztikai regisztereken alapul, de ezek mellett a települési önkormányzatok is adnak információt a területükön működő kereskedelmi és nem üzleti céllal üzemeltetett szálláshelyekről.

5.3.2. Utazásszervezés

Az utazásszervezők és közvetítők tevékenységének és árbevételének, illetve az utazási céloknek a megfigyelésére végzett statisztikai adatgyűjtéshez a felvételi keret meghatározása a Magyar Kereskedelmi Engedélyezési Hivatal (MKEH) utazási vállalkozók hatósági nyilvántartása alapján történik.

5.3.3. Turisztikai kereslet

A külföldiek magyarországi, valamint a magyar lakosság belföldi és külföldi utazási szokásainak statisztikája alapvetően a KSH adatgyűjtéseit hasznosítja, de felhasználja a Belügyminisztérium

által elrendelt, és a hivatal által végrehajtott teljes körű statisztikai adatgyűjtést is, az ún. Személy- és járműforgalmi jelentést.

Mindemellett a KSH a turisztikai keresleti adatfelvételeinek alapsokasági becsléseihez átveszi a Nemzeti Útdíjfizetési Szolgáltató Zrt. kamerarendszerében keletkező gépjárműforgalmi adatokat is.

5.3.4. Vendéglátóhelyek

Ez a szakstatisztika is annak az önkormányzatoknál folyó adatgyűjtésnek az eredményeit hasznosítja, amit a belkereskedelmi egységek, üzletek regisztere is.

5.4. Szállítás, közlekedés

5.4.1. Közúthálózat

Az országos és helyi közút-állomány fejlesztésének, műszaki állapotának megfigyelése a Nemzeti Fejlesztési Minisztérium (NFM) statisztikai adatgyűjtése és a Magyar Közút Nonprofit Zrt. adminisztratív nyilvántartása alapján történik.

5.4.2. Járműállomány

Az országos közúti járműállomány és a Magyarországon első alkalommal forgalomba helyezett járművek adatainak számbavétele kizárólag a KEKKH igazgatási nyilvántartásai alapján történik. Az országos közúti járműállomány, illetve a Magyarországon első alkalommal forgalomba helyezett járműállomány regisztereiből jut a KSH a járműveket és az üzemeltetőiket, illetve a tulajdonosaikat jellemző néhány adathoz.

5.4.3. Szállítási módok

Közúti áruszállítás

A közúti áruszállítás teljesítményét a KSH reprezentatív adatgyűjtés keretében méri, amihez a KEKKH Országos Járműnyilvántartása biztosítja a felvételi keretet.

Vasúti szállítás

A vasúti szállítási teljesítmények mérésére a KSH saját adatgyűjtéseivel felvételi keretként használja a NKH kezelésében lévő országos hálózaton történő vasúti szállításhoz szükséges engedéllyel rendelkező vállalkozások listáját és ugyanez az adatgazda biztosítja a vasúti, vízi és légi balesetekre vonatkozó adatait. Az adatforrások között szerepelnek az NFM adatai is a szállítási tevékenységgel összefüggő gazdálkodásra vonatkozóan.

Vízi szállítás

A KSH adatgyűjtései mellett az NFM azon statisztikai felméréseinek eredményei hasznosulnak a belvízi áru- és személyszállítási teljesítmények, a kikötői, valamint a tranzit hajó- és áruforgalom jellemzésére szolgáló statisztikákban, melyek a kikötői áru-, és hajóforgalmat, valamint a kikötők és úszólétesítmények infrastruktúráját írják le.

Légi szállítás

A szakstatisztika a repülőterek forgalmát, a légi szállítással foglalkozó gazdasági szervezetek teljesítményeit, repülőgép- és személyzeti állományát és a repült útvonalait vizsgálja, amihez a KSH saját adatgyűjtésein túl felhasználja a NKH-tól átvett adminisztratív adatokat a polgári

légi járművekről, valamint az NFM légi járművekre, légi szállításra, légitársaságokra, repülőterekre és gazdálkodásra vonatkozó, statisztikai célból gyűjtött adatait.

5.4.4. Közlekedési balesetek

A szakstatisztika alapvető adatforrásai az Országos Rendőr-főkapitányságtól átvett igazgatási adatok a személysérüléssel közúti közlekedési balesetekről, az NKH nyilvántartásából származó vasúti, vízi és légi balesetek adatai, valamint az NFM statisztikai adatai a kötélpálya, távolsági szalagpálya, sikló és sífelvonó használatával történő szállításról és az ezzel összefüggésben történt balesetekről.

5.5. Információ, kommunikáció

5.5.1. Posta és távközlés

A szolgáltatások különböző szegmenseinek bemutatásához, a GDP-n belül a távközlés és internet szerepének becsléséhez néhány KSH adatgyűjtés mellett az NFM egy adminisztratív és három statisztikai, valamint a Miniszterelnökség két statisztikai adatgyűjtése szolgál adatforrással. Az átvett adatok a postai szolgáltatás minőségi mutatóira, a hírközlési szervezetek gazdálkodására, a küldeményekre, bevételre, a postai szolgáltató helyeire, mobil postával való ellátottságra, a távközlési tevékenységre és annak feltételeire, a telefonellátottságra és a szolgáltatás jellemzőire vonatkoznak.

5.5.2. Internet-szolgáltatás

A szakstatisztika az internet-előfizetések jellemzőit és az internet-szolgáltatók erőforrásait leíró mutatókat állít elő, amihez a KSH saját adatgyűjtései mellett felhasználja az NFM idevonatkozó statisztikai felvételeinek eredményeit is.

6. Környezet

6.1. Élővilág és természetvédelem

A védett területek, védett természeti értékek, a bioszféra rezervátumok és nemzetközi jelentőségű hazai vizes élőhelyek bemutatásának adatforrása az FM „Tájékoztató az országos jelentőségű védett természeti területekről” című éves gyakoriságú, teljes körű statisztikai adatgyűjtése, ami felhasználja a minisztérium felügyelete alá tartozó Természetvédelmi Információs Rendszer egyes adatait is.

6.2. Erdő és környezet

Az erdőállomány és erdőgazdálkodás jellemzőinek bemutatása a Nébih statisztikai adatgyűjtéseire és adminisztratív adatforrásaira alapul.

6.3. Élővilág és természetvédelem

A szakstatisztika a biológiai sokféleség megőrzését szolgáló természetvédelem bemutatását az FM statisztikai adatgyűjtésére alapozza, amely a védett természeti területekről és a "Natura 2000" területekről szolgáltat adatot a KSH-nak.

6.4. Levegő

A légszennyezőanyag-kibocsátás és a levegőminőség statisztikai eszközökkel történő jellemzéséhez a KSH az FM-től „levegőtisztaság-védelmi adatok” címen évente átvett adminisztratív adatokat használja.

6.5. Agrárkörnyezet

A mezőgazdaság és a környezet kapcsolatának, illetve a mezőgazdaság környezetre gyakorolt hatásának statisztikai mérése kizárólag másodlagos adatforrások felhasználásával történik. Az FM, az AKI, a Földmérési és Távérzékelési Intézet (FÖMI), az MVH, a NÉBiH, a MEKH és az Országos Meteorológiai Szolgálatnak (OMSZ) az adatgyűjtéseit, adminisztratív adatbázisait, mérési, tudományos eredményeit, illetve szakértői becsléseit hasznosítja a szakstatisztika. Ezek között az AKI statisztikai adatgyűjtése a műtrágyák értékesítésére vonatkozik. A KSH az ökológiai termelést, feldolgozást, forgalmazást tanúsító szervezetek nyilvántartásaiból veszi át a biogazdálkodást folytató gazdálkodók, valamint a mezőgazdasági és feldolgozott biotermékek, a NÉBiH igazgatási állományából pedig a mezőgazdasági területen felhasznált szennyvíziszap adatait.

6.6. Energia és környezet

A szakstatisztika adatforrásai a MEKH statisztikai adatgyűjtései az energiamérlegről, energiafelhasználásról, energiagazdálkodásról, a távhőtermelőkről és távhőszolgáltatókról.

6.7. Közlekedés és környezet

A szakstatisztika a közlekedés és környezet kapcsolatrendszerét írja le, amihez a KEKKH és a Magyar Közút Nonprofit Zrt. közlekedéssel kapcsolatos, valamint az FM levegőszennyezésre vonatkozó adatait használja fel.

7. Területi statisztika

A szakstatisztika, jellegéből következően az adatforrások szerteágazó sokaságából állítja elő az indikátorait. Az OSAP-ban előírt közvetlen KSH adatgyűjtések mellett külső szervezetektől származó adatokat is felhasznál: a helyi önkormányzatok mérleg és költségvetési adatait a MÁK-tól, a meteorológiai adatokat az OMSZ-től, a nyilvántartott álláskeresőket a NGM-től, a nyugdíjban, nyugdíjszerű ellátásban részesülőket az ONYF-től veszi át, és a személyi jövedelem adatokat a NAV biztosítja.

2. melléklet: A KSH által használt másodlagos adatforrások

A KSH által használt másodlagos adatforrások legfontosabb jellemzői megtekinthetők a hivatal honlapján a Módszertani információk (metaadatok)⁶⁵ cím alatt. A hivatalos statisztikai szolgálat tagjai között a statisztikai felmérések adataira vonatkozó átvételi igényeket a KSH az OSAP mellékletében⁶⁶ évente közzé teszi. Az alábbiakban összefoglaló jelleggel felsoroljuk a másodlagos adatforrásokat jellegük szerint két külön táblázatban a KSH adatforrás-nyilvántartásának 2016. márciusi állapotának megfelelően. A 196 adminisztratív⁶⁷ adatforrást adatgazdánként, illetve adatgazda-csoportonként rendeztük, vagyis azon szervezetek, vagy azok típusa szerint, amelyek az adatállományokért jogszabály szerint felelősek. A 2015-re bejelentett 80 külső hivatalos statisztikai adatforrást azonban azok szerint az intézmények szerint soroljuk, fel, amelyek a tényleges adatgazdákat képviselik a hivatalos statisztikai szolgálatban.

A nem statisztikai másodlagos adatforrások között csak a hazaiakat soroljuk fel, de ezek mellett a hivatali hasznosításban léteznek olyanok is, amiket nemzetközi szervezetek állítanak össze, és tesznek közzé. Ilyenek például az ENSZ országcódjai, az Eurostat Geonómenklatúrája, vagy az egyes nemzetközi szervezetek által közzé tett tagsági listák.

A táblázatokban az adatgazdák, illetve a felelős szervezetek a KSH-nak átadott adatállományok száma szerint vannak rendezve.

1. tábla

A KSH által használt adminisztratív és egyéb nem az OSAP keretében gyűjtött adatok

Adatgazda – az adatállomány megnevezése

Nemzeti Adó- és Vámhivatal

A mezőgazdaságba sorolt gazdasági szervezetek és egyéni vállalkozások összesített áfaadatai

A Nemzeti Adó- és Vámhivatal által gyűjtött adók éves adatai

A vámteher adatok gazdálkodó szervezetenkénti és KN bontásban

Adóadatok (a naptári évtől eltérő üzleti évek cégek adatai)

Adóadatok (áfabevallás)

Adóadatok (társaságiadó-bevallások)

Adóhatóságtól átvett gazdasági szervezetek elemi szintű adatai

Áfa 65M Összesítő jelentés kereskedelmi partnereknél

Áfaadónemben végzett adóellenőrzések eredménye

Ágazati különadók, gyógyszeradók, játékadó és egyéb adók

Bevallás pénzügyi szervezetek különadójáról (P90)

Családi adókedvezmény adatok

Diplomáciai és konzuli képviseletek, nemzetközi szervezetek fogyasztásával kapcsolatos adóalap-nyilvántartás

Egyéni vállalkozók adó és járulék adatai

⁶⁵ <http://www.ksh.hu/apps/meta.main>

⁶⁶ Lásd: http://www.ksh.hu/docs/hun/info/adatgyujtes/2015/HSSZ_egymuk_adatvetelei.pdf

⁶⁷ Itt soroljuk fel mindazokat az adatátvételeket, amik klasszikusan adminisztratív adatforrásból származnak, valamint azokat is, amik egyéb, de nem az OSAP keretében végzett adatgyűjtés eredményei.

Egyéni vállalkozók adóbevallásai
Egyes adókötelezettségek, járulékok és támogatások
Egységes vámokmány (export)
Egységes vámokmány (import)
Háztartási alkalmazottak száma
Hozzáadott-értékadó visszatérítése (ELEKÁFA)
Illetékbevétel képződése
Járulékadatok
Jelentés a lakás- és lakótelek-forgalom alakulásáról
Jövedéki termékek termelésének alakulása
Kimutatás az áfa-visszatérítések teljesítéséről
Környezetvédelmitermékdíj-adatok
Külföldi vállalkozások kapcsolattartóinak elérhetősége
Mezőgazdasági őstermelők
Munkavállalók és egyéni vállalkozók járulékadatai
NAV A60 uniós külkereskedelmi termékforgalmi összesítő
Onlinepénztárgép-adatok
Regisztrációs- és energiaadóadatok
Személyi jövedelemadó bevallás egyedi azonosításra alkalmatlan egyedi adatai
Személyi jövedelemadó külföldiekre vonatkozó adatbázisa
Szia adónemben végzett adóellenőrzések eredménye
Tájékoztató egyes jövedelemtulajdonosok bizonyos költségvetési kötelezettségeinek és juttatásainak pénzforgalmáról
Településsoros személyijövedelemadó-adatok
Transzferáradatok
Vámbeszedési költségek
VIES-(VIES=VAT Information Exchange System) áfa összesítő nyilatkozat

Országos Egészségbiztosítási Pénztár

A bőr és nemibeteg gondozók működési mutatói
A migrációban résztvevő TAJ-számmal rendelkező magyar és külföldi állampolgárok adatai
A pszichiátriai gondozók jelentése
Bevándorlási és Állampolgársági Hivatal részére elszámolt kiadások
Egészségbiztosítási Alap éves költségvetési beszámolója
Egészségbiztosítási ellátások éves területi adatai
Egészségbiztosítási ellátások havi adatai
Háziorvosi és házi gyermekorvosi praxisok betegforgalmi és morbiditási adatai
Járóbeteg-szakellátás betegforgalmi kimutatása

Jelentés a járóbeteg-ellátás valós teljesítményadatairól

Jelentés a kórházi ápolási esetekről

Jelentés a közforgalmú gyógyszerárak támogatott és nem támogatott gyógyszerforgalmáról

Jelentés az aktív és a krónikus fekvőbeteg-ellátás valós teljesítményadatairól

Jelentés az iskola-egészségügyi ellátásról

Jelentés az otthoni szakápolási tevékenységről

Kórházi szolgáltatások mutatói

Magzati veszteség adatai

Védőnői jelentés összesítője

Magyar Nemzeti Bank

Adatszolgáltatás a FISIM felhasználói szektorokra való felosztásához

Adatszolgáltatás a külföldi közvetlen tőkebefektetések (részvény és egyéb részesedések) teljeskörűsített

Adatszolgáltatás a teljeskörűsített közvetlen tőkebefektetésekről, tőkekifektetésekről (Outward FATS R01, R29)

Az MNB halmozott eredménykimutatása, banküzem működési költségei, ráfordításai

Befektetési alapok jövedelem mátrixa

Biztosítók statisztikai mérlege

Egészség- és önszegélyező pénztárak éves beszámolója

Éves fizetési mérleg adatok a nemzeti számlákhoz

Felügyelt pénzügyi szervezetek és nyilvántartott biztosításközvetítők listája

FISIM (pénzügyi közvetítési szolgáltatások közvetett módon mért díja)

Hitelintézetek, hitelintézeti fióktelepek adatai, pénzügyi vállalkozások és pénzügyi tevékenységet folytató egyéb szervezetek jelentései

Kamatmátrix

Külfölddel kapcsolatos kiugró (szuper-) osztalékadatok a nemzeti számlákhoz

Külfölddel kapcsolatos osztalék és visszaforgatott jövedelem adatok a nemzeti számlákhoz

Negyedéves fizetési mérleg adatok a nemzeti számlákhoz

Nem pénzügyi vállalatok tájékoztató mérlegadatai

Rezidens nem banki ügyfelek külföldi (forint, deviza) fizetési forgalom miatti jóváírásai és terhelései

Viszonyatlan átutalások és eszmei, szellemi javak adatszolgáltatása

Magyar Államkincstár

A helyi önkormányzatok mérleg és költségvetés adatai

Alapadatok a fizikai és szellemi foglalkozásúak munkaidőmérlegéhez a Központosított

Éves munkaügyi jelentés a Központosított illetményszámfejtési rendszerben lévő szervezetekről

Havi jelentés a központi költségvetés kiadásainak és bevételeinek alakulásáról

Havi munkaügyi jelentés a központosított illetményszámfejtési rendszerben lévő szervezetekről

Helyi önkormányzatok éves költségvetési beszámolója
Időközi költségvetési jelentések beruházás adatai
Jelentés az önkormányzatok tulajdonában lévő ingatlanvagyonról
Költségvetési szervek adatai
Központi költségvetési szervek és alkalmazottainak adatai – az EU Személyzeti Szabályzat 65. cikke szerinti mutatók számításához
Központi költségvetési szervek és helyi önkormányzatok költségvetése
Központi költségvetési szervek éves költségvetési beszámolója
Központi költségvetési szervek mérlegjelentése
Központi költségvetési szervek, önkormányzatok és önkormányzati költségvetési szervek törzsadattára
Munkaerőköltség-felvétel a központosított illetményszámfejtési rendszerben lévő szervezetekről
Önkormányzati költségvetési szervek mérlegjelentése és költségvetési jelentése
Üres álláshely a központosított illetményszámfejtési rendszerben lévő szervezetekről

Nemzetgazdasági Minisztérium

A kormányzati szektorba sorolt vállalkozások negyedéves adatai
Aktív foglalkoztatáspolitikai eszközök megyesoros adatai
Bejelentett álláshelyek adatai
Éves zárszámadási törvényjavaslat
Foglalkoztatási hivatalok munkanélküli ellátásokra fordított kiadásai
Költségvetési zárszámadás közgazdasági/funkcionális adatai
MNV Zrt., MFB Zrt. és Nemzeti Földalapkezelő Szervezet beszámoló jelentése, rábízott vagyonnal kapcsolatos tranzakciói
Munkaerőpiaci politikák, kiadási adatok
Munkaerőpiaci politikák, résztvevő adatok
Nyilvántartott álláskeresők adatai
Nyilvántartott álláskeresők megyesoros adatai
Nyilvántartott álláskeresők településsoros adatai
Uniós igények szerinti egyéni kereseti adatok

Nemzeti Élelmiszerlánc-biztonsági Hivatal

Beszámoló az erdősítésekről és a fakitermelésről
BIR (Baromfi Információs Rendszer) - baromfikeltető állomások
ENAR (Egységes Nyilvántartási és Azonosítási Rendszer)
Erdőstatisztikai adatok
Jelentés az erdészeti szaporítóanyagokról
Mezőgazdasági területen felhasznált szennyvíziszap

Növényvédő szerek mezőgazdasági felhasználása

Országos erdőállomány adattár

Őstermelői nyilvántartás

Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala

A KEKKH személyi és lakcímnnyilvántartása

A közúti járműállományból a 3,5 tonnás és feletti teherbírású tehergépjárművek és vontatók névvel és címmel kiegészítve

Állandó népesség településenkénti adatai (nemenkénti és koréves bontásban)

Közúti járműállomány

Magyarországon anyakönyvezett népmozgalmi események

Magyarországon első alkalommal forgalomba helyezett járműállomány

Nemzetközi vándorlásban résztvevő állampolgárok adatai

Országos Nyugdíjbiztosítási Főigazgatóság

Állománystatisztikai adatok a nyugdíjban, ellátásban, járadékban és egyéb járandóságban részesülőkről (január, emelés után)

Családtámogatási ellátások éves és területi adatai

Családtámogatási ellátások havi adatai

Kimutatás a fogyatékosági támogatás adatairól

Nyugdíjban, ellátásban, járadékban és egyéb járandóságban részesülők

Nyugdíjbiztosítási Alap éves költségvetési beszámolója

Országos Meteorológiai Szolgálat

Éves meteorológiai adatok

Havi meteorológiai adatok

Légszennyező anyagok kibocsátási leltára (CLRTAP)

Levegőtisztaság-védelmi adatok

Üvegházhatású gázok kibocsátási leltára (UNFCCC)

Országos Tisztifőorvosi Hivatal

Egészségügyi szolgáltatók és működési engedélyük nyilvántartása

Jelentés a kábítószerfogyasztókról és kezelésükről

Jelentés az addiktológiai betegek gondozásáról

Az Országos Tisztiorvosi Hivatal címlistája a fürdőkről

Összesítő jelentés a veleszületett rendellenességgel sújtott újszülöttekről (csecsemőkről)

Nemzeti Közlekedési Hatóság

A Magyarországon bejegyzett polgári légi járművek száma

Az országos vasúti pályahálózaton történő vasúti szállításhoz szükséges engedéllyel rendelkező gazdasági szervezetek listája

Közösségi áru fuvarozói engedély nyilvántartása

Légiközlekedési szakszemélyzet adatai

Bevándorlási és Állampolgársági Hivatal

Bevándorlási és Állampolgársági Hivatal éves költségvetési beszámoló

Külföldi állampolgárok adatai

Menekültügyi nyilvántartás (MIR)

Egészségügyi Nyilvántartási és Képzési Központ

Oklevelet, diplomát szerzett orvosok, fogorvosok, gyógyszerészek létszámkimutatása

Jelentés az orvosok, fogorvosok, gyógyszerészek, pszichológusok egészségügyi szakvizsgáiról

Kimutatás az orvosok fogorvosok területi szakképesítés kor és nemek szerinti megoszlásáról

Igazságügyi Minisztérium

Cégbejegyzésre kötelezett gazdasági szervezetek adatai

Egyéni vállalkozói nyilvántartásba vett vállalkozások adatai

Vállalkozások éves elektronikus beszámoló adatai

Magyar Energetikai és Közmű-szabályozási Hivatal

Az egyes villamosenergia-kereskedők által lebonyolított havi import és export forgalom határmetszékeként mennyiségben (MWh) és értékben kifejezve a V610 számú MEKH adatszolgáltatás alapján

Földgáz exportjára és importjára vonatkozó havi adatok határmetszékeként mennyiségben kifejezve

Energiaadatok

Nemzeti Média- és Hírközlési Hatóság

Közvetlen magyarországi filmgyártási költségek összesített nyilvántartása

Nyilvántartás a vezetékes műsorelosztókról

A postai, hírközlési, média és műsorszórási hatósági tevékenység

Oktatási Hivatal

Jelentés a tanév I. évfolyamára jelentkezett pályázókról

Jelentés a tanév I. évfolyamára felvett pályázókról

Felsőoktatási információs rendszer

Hegyközségek Nemzeti Tanácsa

Hegyközségi adatok

Szőlő- és borkészítés adatai, borkészletek

Miniszterelnökség

Az európai uniós támogatásokból finanszírozott projektek egyedi adatai (EMIR, FAIR adatbázis)

Az építési e-napló egyedi adatai (számított építményérték)

Nemzeti Rehabilitációs és Szociális Hivatal

Működési engedéllyel rendelkező szociális és gyermekvédelmi szolgálatok és intézmények adatai
Statisztikai jelentés a Nemzeti Rehabilitációs és Szociális Hivatal munkájáról

Országos Onkológiai Intézet

Jelentés az onkológiai tevékenységről
Összesítő jelentés a daganatos megbetegedésekről

Települési önkormányzatok

Jelentés a kereskedelmi tevékenységet folytatókról
Jelentés a működést kezdő, megszűnt, üzemeltetési kört módosított kereskedelmi és nem üzleti céllal üzemeltetett szálláshelyekről, az egyéb szálláshelyek kivételével

Belügyminisztérium

Egységes szabálysértési nyilvántartó és eljárástámogató rendszer
Munkaerőpiaci politikák, közfoglalkoztatásban résztvevők adatai

Mezőgazdasági és Vidékfejlesztési Hivatal

Agrártámogatásban részesülő gazdálkodók nyilvántartása
Integrált igazgatási és ellenőrzési rendszer

Budapest Főváros Kormányhivatala

Külföldön történt népmozgalmi események

Büntetés-végrehajtás Országos Parancsnoksága

Fogvatartotti jelentés

ÉGSZI Senior Kft.

Építőipari anyagárak

Erzsébet Utalványforgalmazó és Gazdasági Szolgáltató Zrt.

Jelentés az Erzsébetutalvány-rendszer adatairól

EU-transzferek lebonyolítói

Az Európai Uniótól kapott, illetve az Európai Uniónak fizetett transzferek

Földmérési és Távérzékelési Intézet (FÖMI)

Kimutatás a városok, községek területnagyságáról

Földművelésügyi Minisztérium

Hulladékgazdálkodási információs rendszer

Honvédelmi Minisztérium

Katonai célú eszközök adatai

Kompenzációt fizető költségvetési szervezetek

Kompenzációban részesülő teljes munkaidősök átlagos létszáma és a kifizetett kompenzáció összege

Közlekedési balesetek adatgazdái

A vasúti, vízi és légi balesetek adatai

Magánnyugdíjpénztárak és önkéntes nyugdíjpénztárak

Jelentés az önkéntes- és magánnyugdíjpénztárak tagjairól

Magyar Kereskedelmi Engedélyezési Hivatal

Utazásszervező és -közvetítő tevékenységi engedéllyel rendelkező gazdálkodási szervezetek listája

Magyar Közút Nonprofit Zrt.

Országos közutak adatai

Magyar posta

Postai irányítószámok

Magyar Tudományos Akadémia

MTA hazai és külföldi tagjai, tudományos fokozatosok

Magyar Villamosenergia-ipari Átviteli Rendszerirányító Zrt.

Villamos energia külkereskedelmi forgalom mennyiségi adatai

Munkavédelmi hivatalok

A munkabalesetek adatai az európai munkabaleseti statisztika (ESAW) szerint

Nemzeti Dohánykereskedelmi Nonprofit Zrt.

Nemzeti dohányboltok havi rendszerességű dohányforgalmi adata

Nemzeti Útdíjfizetési Szolgáltató Zrt.

Az útdíjrendszerben üzemeltetett, adott mérőpontokról származó járműadatok átvétele

Nyugdíj- és egyéb ellátás adatok adatgazdái

Új nyugdíj és egyéb ellátás megállapítások

Országos Bírósági Hivatal

Nonprofit szervezetek adatai

Országos Epidemiológiai Központ

Kimutatás a bőr - és nemibeteg-gondozókban kezelt szexuális úton terjedő betegségekről

Országos Korányi TBC és Pulmonológiai Intézet

Tüdőbeteg-gondozók éves jelentése

Országos Rendőr-főkapitányság

Személy sérüléses közúti közlekedési baleset

Országos Vérellátó Szolgálat

Országos vérkészítmény-előállítási és-felhasználási jelentés

Ökológiai termelést, feldolgozást, forgalmazást tanúsító szervezetek

Biogazdálkodást folytató gazdálkodók adatállománya, mezőgazdasági és feldolgozott biotermékek statisztikai adatai

Pénzügyi felügyeleti szerv és garanciaalapok

A pénzügyi vállalkozások és pénzügyi tevékenységet folytató egyéb szervezetek éves jelentései

Szellemi Tulajdon Nemzeti Hivatala

Szabadalmi adatok

Vetőmag Szövetség és Termék tanács

Vetőmagtermelés adatai

*A KSH által az OSAP keretében átvett statisztikai adatforrások**A hivatalos statisztikai szolgálat felelőse – az adatállomány megnevezése***Emberi Erőforrások Minisztériuma**

Adatgyűjtés az egészségügyi ágazati munkaügyi és bérnyilvántartási rendszeréhez

Adatszolgáltatás a köznevelési intézmények tevékenységéről

Adatszolgáltatás a nem állami oktatási intézmények oktatási kiadásairól

Az egészségügyi intézmények árinformációi

Bejelentett fertőző megbetegedések

Beszámoló jelentés a foglalkozás-egészségügyi alapszolgálatok munkájáról

Influenza és influenzaszerű megbetegedések miatt orvoshoz fordult betegek becsült száma járványok idején

Jelentés a bruttó 200 ezer forint értékhatár feletti tárgyi eszköznek minősülő egészségügyi gépek és műszerek állományának változásáról

Jelentés a felsőoktatási intézmények tanév eleji helyzetéről

Jelentés a foglalkozási megbetegedésekről és fokozott expozíciós esetekről

Jelentés a gyógyszertárakról

Jelentés a hangversenyekről

Jelentés a HIV-fertőzésekről, AIDS megbetegedésekről

Jelentés a kiállításokról

Jelentés a könyvtárak tevékenységéről

Jelentés a közművelődési tevékenységekről

Jelentés a levéltárak működéséről

Jelentés a magyarországi állatkertek, vadsparkok és kultúrparkok tevékenységéről

Jelentés a sporttal kapcsolatos ágazati szegmensről

Jelentés a szabadtéri színházak tevékenységéről

Jelentés a színházak (opera) tevékenységéről

Jelentés a táncegyüttesekről

Jelentés az egészségügyi szakképzésben végzettekről

Jelentés az emberi mérgezési esetekről

Jelentés az év folyamán oklevelet, tudományos fokozatot, címet és felsőfokú, illetve szakirányú szakképzettséget szerzett hallgatókról

Muzeális intézmények adatai

Összesítő az élelmiszer eredetű megbetegedési eseményekről

Védőoltási jelentés

Nemzeti Fejlesztési Minisztérium

A gazdálkodást jellemző fontosabb adatok

A hajózási engedélyesek közforgalmú kikötők saját használatú kikötők és vízi úszólétesítmények adatai

A hálózati közlekedési infrastruktúra felügyeleti adatai

A helyi közutak és hidak adatai

A hírközlési gazdasági szervezetek gazdálkodását jellemző adatok

A kikötői áru- és hajóforgalom

A közlekedésrendészet ráfordításadatai

A közlekedési hálózati infrastruktúra központilag nyilvántartott hitelszolgálati adatai

A távközlési tevékenység berendezései és szolgáltatásai

A távközlési tevékenység évközi adatai

A vasút infrastruktúraadatai

A vízi út infrastruktúraadatai

A vízi utak fenntartási munkái és költségei

Az internet- és a televíziószolgáltatás berendezései és szolgáltatásai

Az internet- és a televíziószolgáltatás évközi adatai

Az út infrastruktúraadatai

Jelentés a közúti közlekedéshez kapcsolódó tevékenységekről

Kötélpályasíkló, távolsági szalagpálya és sífelvonó szállítási és baleseti adatai

Légi járművek és az alkalmazott személyzet adatai

Légi szállítás a repülés szakaszai szerint

Légi szállítás az indulási és célállomások szerint

Légi szállítási teljesítmények adatai

Légitársaságok pénzügyi adatai

Magyarországi repülőterek forgalmi adatai

Vasúti infrastruktúra közlekedési eszközök üzemi és szállítási teljesítmények adatai

Belügyminisztérium

A felszín alatti vizet kitermelő vízkivételek, valamint megfigyelőkutak üzemi figyelési tevékenysége

A felszíni vízkivételek és a felszíni vízbe történő vízbevezetések adatai

A közműves vízellátási és csatornázási tevékenységek főbb műszaki-gazdasági adatai

A tűzmegeelőzési tevékenység éves értékelő statisztikája

A mezőgazdasági vízhasználat és vízszolgáltatás

Az 5 m³/h teljes vízforgalmat, illetve 80 m³/d frissvíz használatot elérő ipari jellegű vízhasználók víztermelési és vízkezelési adatai

Beépített tűzvédelmi berendezések

Egységes nyomozó hatósági és ügyészségi bűnügyi statisztika (azonos a Legfőbb Ügyészség 2341. számú adatgyűjtésével)

Személy- és járműforgalmi jelentés

Tűzeseti és műszaki mentési statisztika

Agrárgazdasági Kutató Intézet

Árumérleg

Dísznövénytermesztés

Főbb termények és termékek készletalakulása

Műtrágya-értékesítés

Országos Bírósági Hivatal

Járásbíróságok ügyforgalmi statisztikai jelentése

Jelentőlap a jogerős határozattal befejezett büntető eljárás fiatalokú vádlottairól

Jelentőlap a jogerős határozattal befejezett büntető eljárás felnőtt korú vádlottairól

Közigazgatási és Munkaügyi bíróságok ügyforgalmi statisztikai jelentése

Magyar Energetikai és Közmű-szabályozási Hivatal

Energiafelhasználási beszámoló

Energiamérleg, energiaszektor, energiahordozók

Energiamérleg, ipari szektor

Nemzetgazdasági Minisztérium

Adatszolgáltatás az egyéni bérekről és keresetekről

Adatszolgáltatás az egyéni illetményekről és keresetekről

Családi pótlék ellátásban részesülők részletezése 20.. évi márciusi állapot szerint

Földművelésügyi Minisztérium

Nettó fakitermelés

Tájékoztató az országos jelentőségű védett természeti területekről és a Natura 2000 területekről

Miniszterelnökség

A postai szolgáltatás és eszközei

A postai szolgáltatás negyedéves adatai

3. melléklet: Együttműködési megállapodás

(másodlagos adatátvétel esetén)

Iktató számok:		
----------------	--	--

I. A megállapodás címe:	Együttműködési megállapodás a -val-ról
-------------------------	--

II. Jelen együttműködési megállapodás az alábbi szervezetek (továbbiakban együttműködő felek) között jött létre:		
A szervezet	Egyrésztől	Másrésztől
1. Megnevezése:	Központi Statisztikai Hivatal	
2. Székhelye:	1024 Budapest, Keleti K. u. 5-7	
3. Rövid neve:	KSH	
4. Képviseli:		
5. Az együttműködési megállapodás összeállításáért felelős munkatárs		
6. Funkcionális központi e-mail címe		

III. Együttműködő felek az együttműködés tapasztalatait évente rendszeresen értékelik, melynek során áttekintik az együttműködési megállapodásban foglaltak teljesülését, szükség esetén a változtatás és a továbbfejlesztés lehetőségét.

1. Jelen megállapodás hatálya	Hatálybalépés napja																
	<table border="1"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td colspan="4">év</td> <td colspan="2">hó</td> <td colspan="2">nap</td> </tr> </table>									év				hó		nap	
	év				hó		nap										
Megszűnés napja:																	
<table border="1"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr> <td colspan="4">év</td> <td colspan="2">hó</td> <td colspan="2">nap</td> </tr> </table>									év				hó		nap		
év				hó		nap											
Határozatlan időre szól <input type="checkbox"/>																	

IV. Az együttműködés alapelvei, célja, tartalma, jogszabályi háttere

V. Az együttműködés területei, azok rövid leírása

VI. A jelen megállapodás keretében a KSH a-tól (*partnerszervezet megnevezése*) az 1. mellékletben részletezett állomány(oka)t veszi át.

- VII. A (partnerszervezet megnevezése)⁶⁸ vállalja, hogy a jelen megállapodás tárgyát képező adatforrásokat érintő bármilyen átalakítást, módosítást csak a KSH illetékes szakmai vezetőinek bevonásával, a velük való egyeztetést követően hajt végre.
Mindazokat a váratlan eseményeket, melyek az átadott adatok minőségét érintik, ... munkanapon belül közlik a KSH-val.
- VIII. A (partnerszervezet megnevezése) vállalja, hogy a KSH visszajelzései alapján a jelen megállapodás tárgyát képező adatforrások adatminőségének javítása érdekében megteszi a szükséges intézkedéseket.
- IX. A..... (partnerszervezet megnevezése) (az adatbenyújtó bevonásával) azokat a módszertani és minőségi információkat, amelyekről a 2. melléklet szerint a KSH-val megállapodott, évente egyszer, legkésőbb az adatok első ízben történő átadásával együtt, valamint változás esetén ismételten a KSH rendelkezésére bocsátja. Az adatgazda az átadásra vállalt tételeket az adatgazdák kérdőívén jelöli be.
- X. A felek a jelen együttműködési megállapodásban foglalt feladatok végrehajtásával kapcsolatban egymással szemben semmilyen jogcímen nem támasztanak költségtérítés vagy egyéb díj iránti követelést.
- XI. Amennyiben a felek nem tartják be a jelen megállapodásban foglalt kötelezettségeiket – különös tekintettel az adatátadás határidejére, illetve a VII. pontban foglaltakra – a másik fél jogosult a jelen megállapodást azonnali hatállyal felmondani.
- XII. A jelen megállapodást bármelyik fél jogosult 30 napos felmondási idővel indokolás nélkül felmondani.
- XIII. A KSH az átvett adatokat a hatályos jogszabályoknak megfelelően kezeli.
- XIV. Az adatgazda a kitöltött 2. melléklethez annak benyújtásakor az alábbi további dokumentumokat csatolja (pl. adatkezelési szabályzat, értékkészletek):
1.
 2.
- XV. A KSH az alábbi kiegészítő dokumentumokat csatolja a megállapodáshoz (pl. adatvédelmi szabályzat, minőségi szabványok):
1.
 2.
- XVI. A felek a jelen megállapodásból eredő – egyeztetéssel nem megoldható – vitáik rendezésére a Budai Központi Kerületi Bíróság kizárólagos illetékességét kötik ki.
- XVII. Jelen megállapodás hatálybalépésével egyidejűleg a felek között érvényben lévő, valamennyi azonos tárgyú megállapodás hatályát veszti.
- XVIII. A felek a jelen megállapodást, mint akaratukkal mindenben megegyezőt, jóváhagyólag aláírták.

Budapest, 20.....

.....

név

.....

részéről

.....

név

a Központi Statisztikai Hivatal

részéről

⁶⁸ Adminisztratív adat átvétele esetén „Az európai statisztika Gyakorlati Kódex” 8.7. ismérvével (az adminisztratív adatok statisztikai céloknak való jobb megfelelése érdekében a statisztikai szervezetek részt vesznek az adminisztratív adatok tervezésében).

AZ ADATÁLLOMÁNY LEÍRÁSA

I. A KSH adatátvételének alapadatai	
1. Megnevezése magyarul	
2. Megnevezése angolul	
3. KSH-azonosítója	
4. Az adatok felhasználásának célja, hasznosulása a KSH-ban	

II. Az eredeti adatforrás, amiből az átvett adatok származnak	
1. Megnevezése	
2. Típusa	2.1 adminisztratív adat statisztikai célú átvétele 2.2 nem adminisztratív adat statisztikai célú átvétele 2.3 egyéb adatátvétel
3. (2.1 esetén) Az eredeti adatgyűjtést/nyilvántartást elrendelő jogszabály(ok) neve és címe	
4. (2.1 esetén) A KSH-t az adatok felhasználására felhatalmazó jogszabály neve és címe	
5. A nyilvántartási/statisztikai egység leírása	
6. A nyilvántartási/statisztikai egységek száma az eredeti adatforrásban	
7. Az adatok frissítésének gyakorisága	

III. Az adatátvétel módja	
1. Eszköze	1.1 ELEKTRA 1.2 KARÁT 1.3 Egyéb, (kérem, nevezze meg!).....
2. KARÁT esetén az adatküldés csatornája	2.1 KARÁT-felület 2.2 Ügyfélkapu/hivatali kapu 2.3 Webservice 2.4 MQSeries
3. Adatállomány szerkezetének leírása KARÁT-adatátadásnál	3.1 Adatgazda adja meg KARÁT-rendszeren kívül 3.2 Adatgazda adja át KARÁT-rendszerben 3.3 KSH írja le KARÁT-adatátvétel tervezésekor 3.4

4. KSH szerkezetleírása esetén az együttműködő fél kér-e a KSH-tól mintafájlt	4.1 igen 4.2 nem
5. Az adatállomány típusa	5.1 txt 5.2 csv 5.3 xml 5.4 Egyéb, <i>(kérem, nevezze meg!)</i>

IV. Az adatátvétel jellemzői	
1. Az adatátvétel rendszeressége	1.1 évenként ismétlődő 1.2 többévenként ismétlődő 1.3 egyszeri, nem ismétlődő
2. Az adatátvétel gyakorisága	
3. A beérkezés határideje	
4. Az adatok vonatkozási éve – kezdőév	
5. Az adatok vonatkozási éve – befejező év	
6. A vonatkozási idő típusa	
7. Az adatok gyakorisága	
8. Az adatok feldolgozottsági foka időszakonként	
9. Kiterjedési köre	9.1 teljes körű adatátvétel 9.2 nem teljes körű (mintavételes) adatátvétel 9.3 kombinált adatátvétel
10. A kiválasztás módja, ha az előző kérdésnél a 9.2 vagy a 9.3 választ adta	
11. Az átvett állományok száma egy időszakban	
12. Az átvett állomány nagysága MB-ban	
13. Az átvett állomány adathelyeinek száma	
14. Az átvett adatállomány szerkezetét a mellékelt sablon szerint (1.a. melléklet, 1. sablon) a partnerszervezet	14.1 a KARÁT-rendszerben tölti fel az adatállománnyal együtt 14.2 jelen megállapodáshoz csatolt dokumentumban adja meg
15. Az átvett adatállomány értékkészleteire vonatkozó leíró adatokat a mellékelt sablon szerint (1.a. melléklet, 2. sablon) a partnerszervezet (több válasz is lehetséges)	15.1 az elrendelő jogszabályra hivatkozással adja meg 15.2 a KARÁT-rendszerben tölti fel az adatállománnyal együtt 15.3 jelen megállapodáshoz csatolt dokumentumban adja meg

16. Az adatállomány tartalmaz-e egyedi (az adatszolgáltatóval kapcsolatba hozható) adatot?	16.1 igen 16.2 nem
17. 16.1 esetén, tartalmaz-e különleges adatot vagy egyéb személyes adatot?	17.1 különleges adatot tartalmaz 17.2 egyéb személyes adatot tartalmaz 17.3 nem tartalmaz különleges vagy egyéb személyes adatot

V. A kapcsolattartó	KSH	partner
1. szervezeti egység (főosztály)		
2. személy neve		
3. telefonszáma		
4. e-mail címe		

VI. Az informatikus		
1. neve		
2. telefonszáma		
3. e-mail címe		

VII. Az adatátvitel végrehajtásáért felelős szervezet (ha az nem a KSH)⁶⁹	
1. neve	
2. a szervezeten belül a szakmai egység (főosztály) megnevezése	
3. felelős személy neve	
4. telefonszáma	
5. e-mail címe	
4. az informatikus neve	
6. telefonszáma	
7. e-mail címe	

VIII. Az adatbenyújtó⁷⁰ szervezet (ha az nem a partnerszervezet)	
5. neve	
6. a szervezeten belül a szakmai egység (főosztály) megnevezése	
7. felelős személy neve	
8. telefonszáma	
9. e-mail címe	
10. az informatikus neve	
11. telefonszáma	
12. e-mail címe	

⁶⁹ A végrehajtásért felelős szervezet a KSH megbízásából annak nevében veszi át és kezeli az adatokat.

⁷⁰ Adatbenyújtó: az a szervezet, amelyik ténylegesen végzi az adatok gyűjtését, az adatállomány karbantartását.

IX. Az adatgazda, illetve az adatbenyújtó a KSH rendelkezésére bocsátja az adatok minőségére vonatkozó információkat (A 2. melléklet „Adatgazdák kérdőívé”-n a 33–42. kérdésekben megfogalmazottak közül azokat, melyeknek átadását az adatgazda vállalja.)		
1. minden adatátadással együtt	1.1	igen
	1.2	nem
2. 1.2 esetén a következő megállapodás szerint:		
X. Visszajelzés a KSH-tól az adatok minőségéről		
1. Az adatgazda kér visszajelzést a KSH-tól az adatok minőségére vonatkozóan?	1.1	igen
	1.2	nem
2. 1.1 esetén a visszajelzés rendszeressége, ütemezése		

SABLONOK

„Az adatállomány leírása” kérdőívhez

A KARÁT felhasználói kézikönyve 4.3-as fejezete szerint kell a szerkezetet és az értékkészleteket megadni.

1. Az adatállomány-szerkezet leírása

A	B	C	D	E	F	G	H	I	J	K	12.	M
FILE-LOG-NEV	SOR-REND	OSZ-LOP-NEV	MEG-NEV	MC57	HOSSZ	TIZE-DES	KO-TELE-ZO	A_MEG-NEV	NKOD	MEGYSEGC59	VONIDOC60	LE-IRAS
Adatfájl logikai neve	Oszlop sorrendje	Oszlop-név	Oszlop megnevezése (comment)	Oszlop típusa (Ismérv típusa)	Oszlop hossza	Tizedesjegyek száma	Kötelezőség	Oszlop megnevezése angolul (comment)	Értékkészlet kódja	Mértékegység	Vonatkozási időszak	Definíció

A fenti fejléc szerint annyi sort kell megadni, ahány oszlopa van az átadásra kerülő adatállománynak. Több adatállomány esetén folyamatosan egymás után kell leírni az adatállományokat, amelyeket a logikai fájlnevében szereplő sorszám különböztet meg.

A – Adatfájl logikai neve – kötelező névkonvenció szerint kell megadni

(VK_[TORZSSZAM]_[MC01]_[Típus]_ Sorsz]_[MEVHONAP]_[DB])

B – Sorrend – Meghatározza, hogy a szerkezetleírásban megadott oszlopok milyen sorrendben követik egymást. Célszerű nagyobb léptéket megadni (pl. tízesével számozni az oszlopokat), hogy szükség esetén könnyebben változtatható legyen az oszlopok sorrendje.

C – Oszlopnév – Az adatbázistábla oszlopának neve, rövidnév, legfeljebb 30 karakteren, kötelező kitölteni, ékezet nélkül kell megadni.

D – Oszlop megnevezése – Az adatbázistábla oszlopának neve, legfeljebb 80 karakteren, ékezhelyesen kell megadni.

E – Oszlop típusa – Kötelező kitölteni az alábbi 3 érték egyikével:

- alfanumerikus
- szám
- dátum – a dátum típusú mezőben időpontot lehet megadni, amelyet a rendszer a „YYYY.MM.DD HH:MM:SS” formátum szerint dolgozza fel, ahol:
 - YYYY – év
 - MM – hónap
 - DD – nap
 - HH – óra
 - MM – perc
 - SS – másodperc

Az év, hónap, nap kitöltése kötelező, míg az óra, perc, másodperc kitöltése nem kötelező.

- F – Oszlop hossza – kötelező kitölteni, pozitív egész szám.
 G – Tizedesjegyek száma – Szám típusú adat esetén kötelező kitölteni.
 H – Kötelezőség – értéke I/N (Igen / Nem), kötelező kitölteni.
 I – Oszlop megnevezése angolul (comment), nem kötelező adat.
 J – Értékkészlet – Nem kötelező adat, itt adhatja meg az értékkészlet nevét, ha az oszlophoz értékkészlet tartozik.
 K – Mértékegység – nem kötelező adat.
 L – Vonatkozási időszak – tájékoztató adat, nem kötelező kitölteni.
 M – Definíció.

2. Az adatállomány értékkészleteinek leíró adatai

A	B	C	D	E	F	G	H
NKOD	NK_MEG- NEV	A_MEG- NEV	NEKODH	NEKOD	NE_MEG- NEV	NE_A_ MEGNEV	NE_LE- IRAS
Nómen- klatúra kódja	Nómen- klatúra megneve- zése	Nómen- klatúra megneve- zése angolul	Elemkód hossza	Elem kódja	Elem megneve- zése	Elem megneve- zése angolul	Elem leírása/ definíciója

Az értékkészletek leíró adatait a fenti fejléc szerint az adatállomány valamennyi nómenklatúrájára elemenként kell megadni.

- A – Nómenklatúra kódja – az átadandó értékkészlet kódja legfeljebb 20 karakter.
 B – Nómenklatúra megnevezése – az átadandó értékkészlet megnevezése legfeljebb 255 karakter.
 C – Nómenklatúra megnevezése angolul – az átadandó értékkészlet megnevezése angolul legfeljebb 255 karakter.
 D – Elemkód hossza – Az értékkészlet elemkódjának hossza legfeljebb 5 hosszú egész szám.
 E – Elem kódja – Az értékkészlet elemeinek kódja legfeljebb 255 karakter.
 F – Elem megnevezése – Az értékkészlet elemeinek megnevezése legfeljebb 255 karakter.
 G – Elem megnevezése angolul – Az értékkészlet elemeinek megnevezése angolul legfeljebb 255 karakter.

ADATGAZDÁK KÉRDŐÍVE

„X”-szel jelölje azokat a kérdéseket a jobb szélső, dupla keretes oszlopban, amelyeknek megválaszolását határidőre⁷¹ vállalják.

Az adatátvétel:	
KSH-azonosítója:	megnevezése

Azonosító adatok		
1.	Az adatforrás megnevezése:	
2.	Az adatgyűjtés célja, tartalma (magyarul):	
3.	Az adatgyűjtés célja, tartalma (angolul):	

Alap meghatározások		
4.	A rendelet vagy más szabály alapján mi az adatgyűjtés célsokasága, milyen körre terjed ki? (mi az a kör, amely tagjairól információt gyűjtenek)	
5.	Használja-e az adatgazda az adatállomány adatait statisztikai célra, készítene-e az adatokból statisztikákat, hol jelentetik meg?	
6.	Az adatállomány adatszolgáltatói: azon személyek, szervezetek meghatározása, akik/amik az adatgyűjtés számára az elsődleges adatokat szolgáltatják.	

Az adatgyűjtés, adatfeldolgozás módszertana		
7.	Van-e olyan jogszabály vagy egyéb hivatalos dokumentum, ami rögzíti az adatgyűjtés fogalmi kereteit, a használt definíciókat és osztályozásokat? Nevezze meg (és csatolja a dokumentumot)!	
8.	Létezik-e nemzetközi ajánlás, szabvány a nyilvántartási/statisztikai egységekre, a használandó definíciókra és osztályozásokra vonatkozóan? Ha van, nevezze meg!	
8.a	Ha létezik, összhangban vannak-e a nyilvántartásban/adatgyűjtésben alkalmazott fogalmak az ajánlással? Igen - nem	
8.b	Ha vannak különbségek, adja meg, hogy mik azok?	
9.	Létezik-e az adatgazda szervezet által készített dokumentum az adatgyűjtés módjáról, rendjéről, tartalmáról, a szereplők feladatairól, az adatellenőrzés, adatrevízió szabályairól, a hibás adatok javítási eljárásáról és az adatok minőségéről? Ha igen, csatolja.	
10.	Az adatállomány idősorának hossza. Mióta működik a nyilvántartás/adatgyűjtés tartalmi, módszertani változtatás nélkül? Ha volt már tartalmilag vagy módszertanilag módosítva, mikor történt, miért, hogyan befolyásolta az adatokat?	
11.	<u>Adminisztratív nyilvántartás esetén</u> áll-e az adminisztratív eljárás ügyfele, az elsődleges adatszolgáltató rendelkezésére olyan számítógépes program, ami az adatszolgáltatói kötelessége teljesítéséhez saját nyilvántartásaiból generálja az adatokat?	
11.a	A számítógépes program megnevezése. Több ilyen esetén van-e az adat-előállítási módszerükben különbség?	
12.	Használ-e az adatgazda nyilvántartást (regisztert) az adatszolgáltatók kiválasztásához?	
13.	Ha igen, milyen személyeket, szervezeteket, területi egységeket stb. tartalmaz ez a nyilvántartás, amire az adatgyűjtést alapozza, és azoknak milyen időpontra vonatkozó állapotát használja?	
14.	Milyen úton (csatornán), milyen állomásokon keresztül kerülnek az adatok az adatállományba? Írja le az adatok útvonalát!	
15.	<u>Nem statisztikai adatgyűjtések</u> esetén mi az adatok aktualizálását, frissítését kiváltó ok, mi az aktualizálás rendje, gyakorisága?	

⁷¹ Együttműködési megállapodás IX. pontja.

16.	<u>Nem statisztikai adatgyűjtések</u> esetén milyen gyorsan – a változást követően mennyi idővel – kerülnek be az új és módosuló információk?	
17.	Hogyan tartják a kapcsolatot az adatszolgáltatókkal? Mikor, milyen okból keresik meg az egyes adatszolgáltatót és milyen formában?	
18.	Késlekedő adatszolgáltatás esetén sürgetik-e az adatszolgáltatót? Ennek jellemzői.	
19.	Ellenőrzik-e rendszeresen a mikroszintű adatokat az alábbi szempontok szerint?	
19.a	a teljességet a célsokaság egységeire vonatkozóan (mindenki válaszolt, akinek kellett, illetve nincsenek többszörösen rögzített rekordok),	
19.b	a változókat (pl. lehetséges értékek, kiugró, szokatlan értékek),	
19.c	a változók közötti összefüggéseket	
20.	Mérik-e a mikroszintű adatok minőségét erre a célra előállított, számított mutatókkal? (Például hiányzó egységek, tételek száma, érvénytelen kódok száma, lehetséges, de hibás adatok száma.)	
21.	Használ-e a beérkezett adatok ellenőrzéséhez, feldolgozásához valamilyen más nyilvántartást, adatforrást? Ha igen, nevezze meg!	
22.	Ha ellenőrzik az adatok minőségét, készítenek-e jelentést, ami összefoglalja a vizsgálat eredményét?	
22.a	Ha igen, a KSH számára elérhető-e a jelentés? Ha igen, csatolja!	
23.	Hogyan javítják a hibás adatokat?	
23.a	Végeznek automatikus javítást?	
23.b	Végeznek manuális javítást?	
23.c	Újra kéri az adatot az adatszolgáltatótól?	
24.	Kérjük, adja meg az adatellenőrzési szempontrendszert, a vizsgált összefüggéseket, elfogadási, elutasítási kritériumokat. Ha van erre vonatkozó dokumentum, mellékelje!	
25.	Végeznek-e adatpótlást, azaz más adatforrásból vagy becsléssel állítják elő az adatot, ha az eredeti hibás volt, vagy hiányzott?	
25.a	Ha igen, pótolnak-e egész rekordokat?	
25.b	Milyen módszerekkel pótolják a hiányzó vagy hibás adatokat?	
26.	Dokumentálják az adatjavítást? Miről mire és milyen módszerrel javították az adatokat?	
27.	Megjelölik-e az állományban a javított adatokat?	

Minőségbiztosítási politika

28.	Van-e az adatállomány vonatkozásában is érvényes minőségbiztosításra vonatkozó szabályzat?	
28.a	A KSH számára hozzáférhető-e? Ha igen, kérjük mellékelni.	
29.	Van-e olyan szervezeti egység, munkakör, aminek felelősségi körébe tartozik az adatok minőségének monitorozása, karbantartása?	
30.	Van-e olyan dokumentum, amelyik leírja, hogy kinek és hogyan kell az adatok minőségét ellenőrizni, javítani?	
31.	Szerveznek-e képzéseket az adatkezelő munkatársak részére, ahol felkészítik őket a minőségbiztosítási feladatokra?	
32.	Az adatok minőségjavítása érdekében kapnak-e visszacsatolást az adatokat felhasználók részéről?	

Az adatok minősége az aktuális adatállományban

Az alábbi adatokat minden adatátadással egy időben vagy megállapodás szerint bocsátják a KSH rendelkezésére.

33.	Az adatellenőrzések mennyisége. Egy-egy adatot milyen alkalommal és módszerekkel ellenőriznek?	
34.	Értékelje az adatok pontosságát a tapasztalatai alapján! Milyen típusú hibák fordulnak elő az adatállományban, és milyen gyakorisággal?	
35.	Intervallumhiba mértéke (tételenként). Változónként/mezőnként adja meg, hogy az esetek hány százalékában esik az adatszolgáltató által megadott adat a lehetséges értékhatárokon kívül.	
36.	Vannak-e az állományban lejárt érvényű adatok?	
36.a	Ha igen, mely változók vagy mely egységek esetén fordulnak elő?	
37.	Adatjavítási arány tételenként. Egy-egy változót, mezőt tekintve az adatok hány százalékát javították?	

38.	Adatszolgáltatók megkeresésével javított hibák aránya tételenként. Az előző kérdés megválaszolásán jelzett javítások hány százaléka történt az adatszolgáltatók megkeresésével?	
39.	Pótolt rekordok aránya. A rekordok hány százalékát pótolták az adatszolgáltatók megkeresése nélkül?	
40.	Regiszterrel vagy más adatállománnyal történő összekapcsolás esetén a nem összekapcsolható rekordok száma, aránya.	
41.	Összehasonlították-e az adatokat a nyilvántartáson belül időben és térben? Értékeltek-e az eredményt? Ha igen, mik a fő megállapítások az adatok konzisztenciájáról?	
42.	Összehasonlították-e az adatokat a nyilvántartáson kívüli adatokkal? Értékeltek-e az eredményt? Ha igen, mik a fő megállapítások az adatok konzisztenciájáról?	

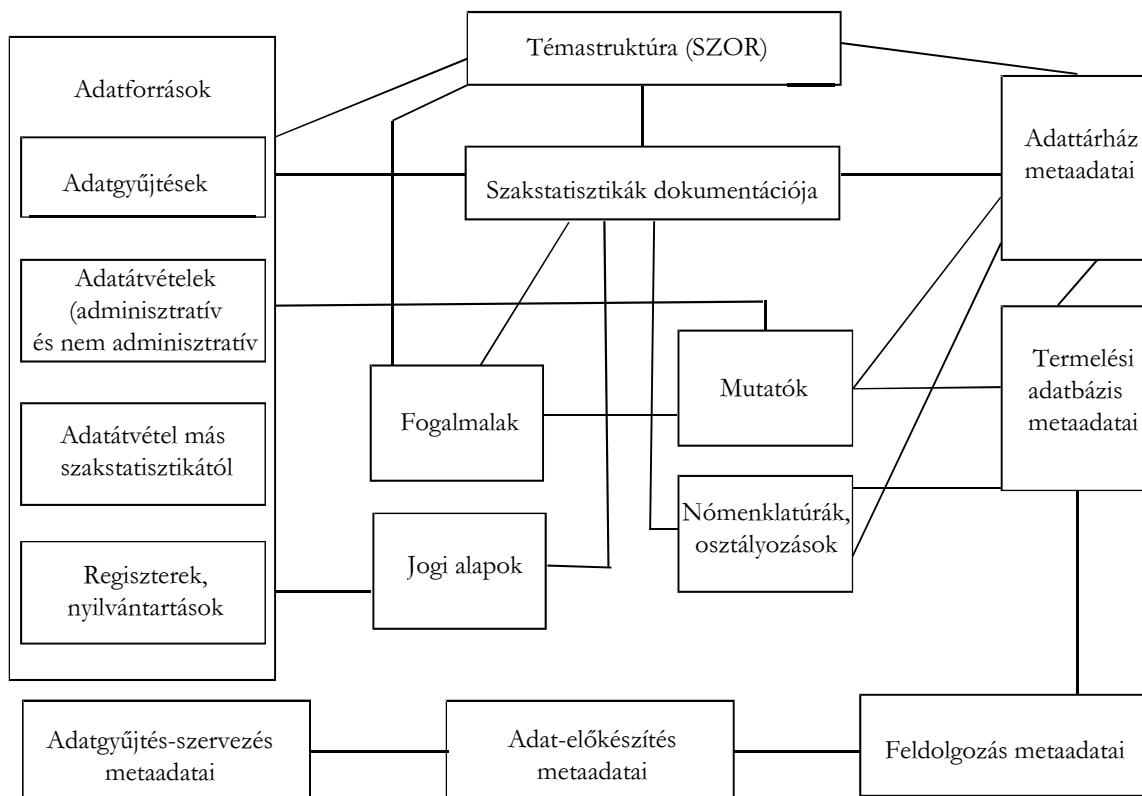
4. melléklet: A KSH metainformációs rendszere⁷², különös tekintettel az adatátvételekre

Az adminisztratív adatok, valamint az átvett nem adminisztratív adatok szerves részét képezik a KSH munkájának, így a metarendszerben is megjelennek.

A metaadatoknak csak egy része szolgálja a felhasználók tájékoztatását, a részletes adatbázis a KSH statisztikusainak munkáját támogatja, míg más részei ún. vezérlő metaadatok, amelyeket a KSH más alkalmazásai használnak. A meta felhasználóknak szóló része elérhető egy egységes rendszerben közvetlenül a honlapról, vagy pedig egy-egy területre vonatkozó rész a konkrét adatokhoz kapcsolva. Az alábbi ábra a KSH teljes metaadatrendszerét mutatja be.

14. ábra

A KSH metainformációs rendszere és összefüggései



A KSH külső honlapján elérhető a szakstatisztikák módszertani leírása, és ehhez kapcsolódnak a jogi alapok, a szakstatisztika által felhasznált adatforrások (statisztikai adatgyűjtések, átvétel más statisztikákból, adminisztratív, valamint nem adminisztratív adatátvételek, továbbá regiszterek, nyilvántartások) metaadatai, a szakstatisztika fontosabb fogalmai, mutatói és a használt osztályozások. A fogalmak, az osztályozások, valamint az adatforrások a honlapon közvetlenül is elérhetők, így ezekből átfogó képet nyerhetünk, de jelenleg csak a KSH szakstatisztikáihoz kapcsolódóan. Középtávú cél a metainformációknak a teljes hivatalos statisztikára való kiterjesztése, valamint a statisztikai célra használt

⁷² A melléklet Almádi Ágnes és Hajósné Ender Éva anyagai alapján készült.

adminisztratív források fogalmainak és jogi alapjának integrálása is. A honlap kétnyelvűségének fenntartása érdekében a nyilvános metaadatokat magyarul és angolul is tárolja a hivatal.

A másodlagos adatforrások értelemszerűen az adatforrások között kerülnek számbavételre. Az adatforrások alrendszer a 2013-as év során került egységesítésre, addig külön szerkezetben voltak letárolva az Országos Statisztikai Adatgyűjtési Program (OSAP), valamint az adminisztratív adatforrások metaadatai. Az egységesítés során mind az elsődleges (adatgyűjtések), mind a másodlagos (adminisztratív - és nem adminisztratív adatátvételek, továbbá regiszterek, nyilvántartások) adatforrások metaadatai rögzítésre kerülnek ugyanazon adatbázisban. Az egységes nyilvántartási szerkezetben történő tárolás és az ezzel párhuzamosan bevezetett másodlagos adatokat kezelő informatikai rendszerek (KARÁT és ADAMES) új igényei következtében az adatforrások metaadatainak köre bővült.

A KSH jövőbeli törekvése, hogy az adatforrások nyilvántartását teljessé tegye, a KSH-ban használt valamennyi adatforrás bekerüljön, valamint ezek metaadatai teljeskörűen le legyenek írva. Az adatforrások jellemzőinek azonos struktúra szerinti begyűjtését egy egységes szerkezetű, az adatgazdákkal kötött együttműködési megállapodás részeként történő leírás, és az adatátvétel során a metaadatok átadása is támogatná.

A 2013 előtti adatforrások hiányzó metaadatainak pótlása nem biztosítható a megfelelő információk rögzítése hiányában, ezért a korábban letárolt metaadatok köre az aktuálistól továbbra is eltérő marad.

Adminisztratív és nem adminisztratív adatátvételek

Az *adminisztratív és nem adminisztratív adatátvételekből* származó adatforrásokról a metainformációs rendszerben nyilvántartott adatok:

- Országos Statisztikai Adatgyűjtési Programban szerepel-e az adatátadás?
- Alapadatok.
- Adatforrás*
 - azonosítója és megnevezése,
 - jellege,
 - állapota,
 - típusa.
- Az adatátvételért felelős szervezet, szakmai egység és felelős munkatársak adatai.
- Az adatátvétel végrehajtásáért felelős szervezet, szakmai egység és felelős munkatársak adatai.
- Az adatgazda szervezet*, szakmai egység és felelős munkatársak adatai.
- Az adatbenyújtó szervezet, szakmai egység és felelős munkatársak adatai.
- Az adatátvétel jellemzői
 - kapcsolódó szakstatisztika,
 - az adminisztratív adatgyűjtést elrendelő magyar jogszabály,
 - nyilvántartási egység leírása, száma,
 - adathelyek száma,
 - az eredeti nyilvántartás frissítésének gyakorisága,
 - az adatátvétel beérkezésének határideje,
 - az adatátvétel rendszeressége, gyakorisága,

- az adatátvétel kiterjedési köre az alapsokaságra vonatkoztatva,
- az adatátvétel lebonyolításának módja.

□ *Adatkör*

- átvett adatkör megnevezése,
- részletes leírás.

A fenti listában a vastag dőlt betűvel jelöltek nyilvánosak, a honlapon is hozzáférhetők. A KSH metainformációs rendszerében az összes információ rendelkezésre áll.

Más szakstatisztikától történő átvétel

A más szakstatisztikától történő átvételek nem képezik szerves részét a jelen kiadvány tematikájának, a szakstatisztikák szempontjából értelmezve azonban a másodlagos adatok körébe ezek is beletartoznak, leírásuk a KSH metainformációs rendszerének része, ezért tájékoztató jelleggel ezekről is szólnunk.

A KSH külső honlapján elérhető a szakstatisztikák módszertani leírásában a felhasznált adatforrások között a más szakstatisztikától történő átvételek ismertetése is szerepel.

A más szakstatisztikától történő átvételek metainformációs rendszerben nyilvántartott adatai a következők:

□ Alapadatok:

- adott adatkör átvételének éve,
- az átadó szakstatisztika megnevezése,
- az átvevő szakstatisztika megnevezése.

□ Az átvett adatkör metaadatai:

- az átvett adatkör megnevezése,
- az átvett adatkör tartalma,
- az adatátvétel gyakorisága.

A fenti listában felsorolt metainformációk nyilvánosak, a honlapon is hozzáférhetők.

A szakstatisztikák dokumentációját évi rendszerességgel aktualizálják, és ennek során az egyes szakstatisztikák valamennyi adatforrásának metaadatait frissítik.

5. melléklet: A KARÁT-rendszerben képezhető indikátorok

I. Az adatátvétel indikátorai

Az indikátorok képezhetőek:

idő szerint

- a) éven belül havonként,
- b) vagy egy adott hónapra,
- c) vagy egy évre összesítve, illetve

szervezet szerint

- a) az egész KSH-ra,
- b) főosztályi felelősség szerint részletezve,
- c) adatgazdánként részletezve.

Az alábbi indikátorok képzése lehetséges:

1. Az adatátvételek dokumentáltságának indikátorai:
 - A. Az összes nyilvántartásba vett adatátvétel száma: a metaadatforrás nyilvántartásában szereplő KSH adatátvételek száma.
 - B. A KARÁT-ba bevont adatátvételek száma: a KARÁT-rendszerben az adott évre megtervezett, véglegesített adatátvételek száma (egynek számolva az adatátvételek adatgazda csoportjaira képzett külön tervváltozatokat).
 - C. A KARÁT-ba bevont adatátvételek aránya $(A/B*100)$.
2. Az adatátvételi feladat összetettségének indikátorai:
 - D. Adat-előkészítésre átadható adatbázisba töltött adatátvételek száma.
 - E. Az adatátvételek során egy időszakban átveendő adatállományok száma.
 - F. Az átvett adatállományok file típusonkénti megoszlása.
3. Az adatátvétel végrehajtás pontosságának indikátorai:
 - G. A várt adatátvételek száma: leszervezett adatátvételek száma adatátvételek, időszakok és adatgazdák figyelembevételével.
 - H. A beérkezett adatátvételek száma, csak a sikeresen feldolgozott adatátvételek figyelembevételével.
 - I. A beérkezési arány $H/G*100$.
 - J. A hibás (átvett, azonosított, de fel nem dolgozott) adatátvételek száma.
 - K. Az átvett (hibás vagy sikeresen feldolgozott) adatátvételek aránya $(H+K)/G*100$.
 - L. A hiányzó (be nem küldött) adatátvételek száma: $G - (H+K)$.
 - M. A hiányzó adatátvételek aránya $((G-H-K)/G*100)$.
 - N. A beérkezett adatátvétel verziók száma az adat hibája miatti újraküldések figyelembevételével.
 - O. Az újraküldés indikátora N/H .
4. Az adatátvétel időszerűségének indikátorai:
 - P. A határidőre beérkezett adatátvételek száma.
 - Q. A határidőre beérkezett adatátvételek aránya $I/G*100$.
 - R. Az első sürgetésre beérkezett adatátvételek száma.
 - S. Az első sürgetésre beérkezett adatátvételek aránya $R/G*100$.

T. A második sürgetésre beérkezett adatátvételek száma.

U. A második sürgetésre beérkezett adatátvételek aránya $T/G*100$.

II. Adatszolgáltatói terhelésmérés indikátorai

1. Adatgazdánként a megtervezett adatátvételekre
 - a. az átadandó állományok száma egy időszakra,
 - b. az átadandó állományok száma az évközi időszakok figyelembevételével.
2. Az adatátvételek a küldés csatornája szerinti megoszlása (KARÁT-felületen, hivatali kapun stb. keresztüli átadás).
3. Az adatátvételekhez elvárt dokumentáció (szerkezetleírás) igénye szerinti megoszlás.