

ISMERTETŐK

KÁDÁR IVÁN – BARTHA KAROLINA – NAGY BEÁTA – DR. FÁBIÁN ZSÓFIA

Térstatisztika a Központi Statisztikai Hivatalban

A KSH térstatisztikai rendszere

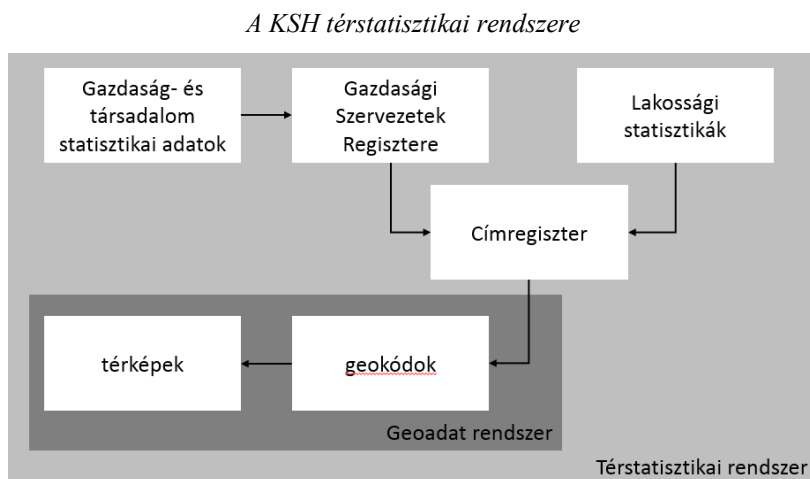
A KSH térstatisztikai rendszerének célja, hogy a megfigyelt egységekhez térképi koordinátákat rendeljen. A koordinátákkal ellátott megfigyelési egységek mind a statisztikai folyamatok előkészítésében, mind a statisztikai elemzésben, tájékoztatásban új lehetőségeket nyitnak meg. Lehetővé válik, hogy az adminisztratív határoktól elszakadva, ne csak települési, járási, megyei stb. adatokat közölhessünk, hanem tetszőleges térbeli lehatárolás szerint aggregálhassuk az elemi adatokat. A koordinátákkal ellátott adatokon összetett térbeli elemzéseket végezhetünk. A statisztikai felvételek előkészítésénél, a mintavétel területén is új lehetőségeket biztosít.

Elsőként az 1 km²-es, illetve a 25 km²-es felbontású rácshálóra aggregált adatok készítésével kísérleteztünk. Ennek előnye, hogy – a településhatárokkal, felsőbb szintű közigazgatási beosztással ellentétben – nem változik sem térben, sem időben. Az adminisztratív határok problémája, hogy a különböző országokban jelentősen eltérhetnek a méreteik, és bármikor megváltozhatnak. A rácshálónak köszönhetően egész Európában összehasonlítható területalapú statisztikák készíthetők, és az ilyen felbontásban készült adatokból idősoroknál nem kell homogenizálni a korábbi adatokat.

A rendszer felépítése

A KSH térstatisztikai rendszerének alapja a címregiszter, amely a közel teljes magyarországi címállományt tartalmazza (~3,4 millió házsám szintű cím). Ezek a címek vannak hozzárendelve a 2011-es népszámlálás adataihoz, illetve a KSH által nyilvántartott gazdasági szervezetek (gazdasági szervezetek regisztere) székhelycímeihez. A címregiszterrel összekapcsolt adatok köre a következő két év folyamán tovább bővül más regiszterekkel is, például a kereskedelmi és a nonprofit szálláshelyek nyilvántartásaival.

A Geoadat-rendszer a címregiszter házsámszintű címeihez rendel koordinátát, így biztosítva, hogy szükség esetén egy függvény segítségével a megfigyelt egység a címazonosítója alapján koordinátát kaphasson.



Forrás: saját szerkesztés.

A Geoadat-rendszer a KSH Oracle adatbázisszerverén működik. Az Oracle az ingyenes Locator bővítményével támogat alapszintű térbeli adattípusokat és hozzájuk kapcsolódó függvényeket, azonban a tesztheink során ennél hatékonyabbnak bizonyult az ESRI által készített ArcGIS for Server részét képező ArcSDE (*Spatial Database Engine*).

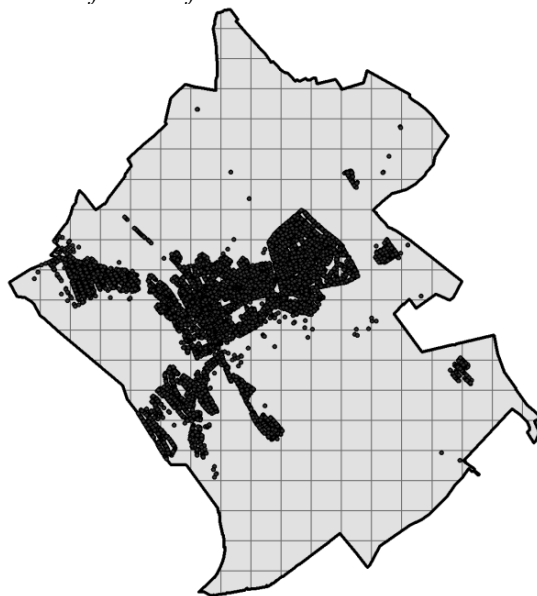
A rendszer használata

Adatbázis szinten az ArcSDE biztosít térbeli műveletek elvégzéséhez SQL-függvényeket. Ezekre a későbbiekben a KSH adatfeldolgozó rendszeréből meghívható műveletek építhetők, ezzel egyszerűsítve az adatok térbeli feldolgozását. A hivatal munkatársai így a megszokott környezetből kezdeményezhetnek akár térbeli elemzéseket is.

Asztali térinformatikai szoftverként az ArcGIS érhető el a hivatalban, ami grafikus kezelőfelületen biztosít hozzáférést az adatokhoz és teszi lehetővé megjelenítésüket, feldolgozásukat.

2. ábra

Székesfehérvár felosztása 1 km²-es rácshálóval



Forrás: saját szerkesztés.

3. ábra

Székesfehérvári geokódok



Forrás: saját szerkesztés.

A rácshálós adatszolgáltatás mellett a térstatistikai rendszer lehetőséget biztosít egyedi adatkérések teljesítésére is. A megrendelő igényeinek megfelelő területre aggregálhatunk adatokat, természetesen az adatvédelmi szabályok figyelembevételével.

Térképek tárolása

A statisztikai adatok geokódolásán túl a térstatistikai rendszer célja, hogy összegyűjtse, központosítva tárolja a KSH térképkészletét. Ennek érdekében megkezdjük a térképek feltöltését éves bontásban. A rendszerbe nemcsak a nyilvános térképek kerülhetnek be a jövőben, hanem egyénileg szabályozott hozzáférési jogokkal bármely főosztály saját adatai is.

Hosszabb távon a térképekkel foglalkozó kollégák egy helyen elérhetik majd az összes térkép legfrissebb vagy éppen adott időpontra érvényes változatát.

Alapfogalmak problémái

A fogalmak használatában a KSH igyekezett a nemzetközi gyakorlatot követni, ahol azonban fogalmi ütközés keletkezett. A kezdetektől használt geostatistika (*geostatistics*) szó alatt egyrészt érthetjük a társadalmi és gazdasági adatok térben való elhelyezését, statisztikai kutatását, másrészt viszont ezt a szót használják az alkalmazott földtudományokban valószínűség-számítási és matematikai-statisztikai kutatásokra is (pl. leőhely méretének becslése mintafúrások alapján). Ezért célszerűbb a térstatistika (*spatial statistics*) szóhasználat, melyet az Eurostat GISCO (*Geographical Information System at the COMmission*) munkacsoportja is szorgalmaz. Alapos körülírásként használható még az *Integrated statistical-geospatial information*, ami azonban hossza és bonyolultsága miatt a közbeszédben várhatóan nem fog elterjedni.

Geostat-projekt

A hivatal, azon belül is a Népszámlálási és a Módszertani főosztály *grant projekt* keretében részt vett egy európai szintű projektben, melynek célja egy rácsháló-adatbázis létrehozása Európa területére a 2011 körüli népszámlálás adataiból. Az ESSnet-projekt neve GEOSTAT, beágyazódik az Eurostat és az ESS hosszú távú közös stratégiájába.

A GEOSTAT IC grant projekt feltételei

A GEOSTAT-hrojektnél a hivatal feladata egy 1 km²-es cellaméretű nemzeti népesség rácsháló-adatbázis előállítás volt a 2011. évi népszámlálás adataiból. Ezt térképen megjelenítve egy grides népességtérképet kaphatunk.

Az egységesség kedvéért szigorú szabályokat szabtak meg, alkalmazkodni kellett a Geostat-projekt előállítási módszereihez és adatspecifikációjához.

Minimális követelményként az adatbázisnak tartalmaznia kellett a teljes népszámlálási lakónépességet, hozzárendelve ahhoz az 1 km²-es cellához (más néven gridhez), amelybe helyileg tartozik. A cellához rendelést a személy geokódolt címe alapján végeztük el.

Ezen túlmenően a következő népszámlálási témakörökre kellett előállítani a rácsháló adatbázisokat:

- nem,
- kor, három korcsoportra bontva (0–14, 15–64, 65 éves és idősebb),
- nem és kor, három korcsoportra bontva (nő/férfi, 0–14, 15–64, 65 éves és idősebb).

A népességbontások összegeinek ki kellett adniuk a teljes népességet minden rácscellában.

A rácscellában lévő népességek összegeinek ki kellett adniuk a teljes népességet, amit a népszámláláskor a hivatal közzétett. Minden eltérést külön kellett dokumentálni a metaadatok között.

Lehetőség volt adatvédelmi intézkedéseket alkalmazni, amennyiben a nemzeti adatvédelmi szabályok miatt valamelyik cellában a megjelenő népesség alacsonyabb volt, mint a meghatározott küszöb, és ezért nem volt nyilvánosságra hozható. Az adatvédelmi intézkedéseket a metaadatok között kellett leírni. Mindazonáltal az alábbi feltételeket be kell tartani:

1. A fent leírt nemzeti rácsháló-adatbázisok teljes népessége vagy népességbontása meg kellett, hogy egyezzen a népszámláláskor közzétett azonos témájú népességgel.

2. Ha ez a feltétel az alkalmazott adatvédelmi eljárások miatt nem volt teljesíthető, akkor a hiányzó vagy hozzáadott személyek számát, a módszert és az okokat dokumentálni kellett a metaadatokban.

3. Az adatvédelem miatt más cellába helyezett személyeket és általánosságban az adatvédelem alá tartozó személyek számát is dokumentálni kellett.

4. A lakott cellákat meg kellett tartani, és nem szabadott lakatlannak jelölni. Ez vonatkozott a teljes népességre, és a népességbontásokra is.

5. A nem lakott cellákat meg kellett tartani, és nem szabadott lakottnak jelölni. Ez vonatkozott a teljes népességre és a népességbontásokra is.

6. Integertípusú adatként kellett jelenteni a népességadatokat. Ez azt jelentette, hogy az adatvédelmi intézkedések nem használhattak string vagy alfanumerikus típusú adatokat.

7. A cellarácsoknak meg kellett maradniuk 1 km²-esnek és négyszög alakúnak. Ez azt jelentette, hogy az egyes lakossági küszöbértékek elérése céljából történő cellaegyesítések nem voltak megengedettek.

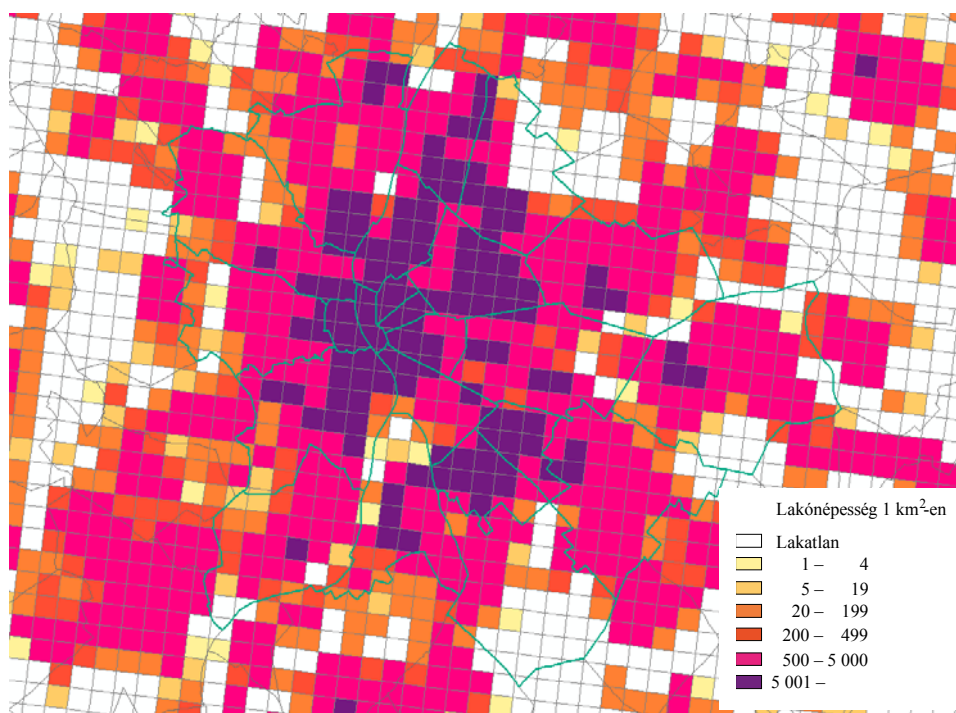
A rácsadatok előállítására többféle lehetőség van attól függően, hogy egy adott területre milyen pontos és részletes térbeli adatok állnak rendelkezésre. A legpontosabb és a valósághoz leginkább közelítő eredményeket a geokódolt címpontok aggregációjából előállított rácsadatok adják. Ennél a módszernél alulról felfelé építkezünk, és a pontszerű adatokból épülnek fel a rács poligonjainak adatai. Azonban vannak olyan esetek, amikor nem állnak rendelkezésre geokódolt címek, és a legkisebb egységek, ami alapján térben el tudjuk helyezni az adatainkat – valamiféle poligonok: pl. számlálókörzetek. Ezek területe azonban sok esetben több cellát is érint, ilyenkor az adott poligonba tartozó adatokat szét kell bontani, és meg kell osztani az adott cellák között.

Mivel a 2011. évi népszámlálás címállományához rendelkezésre állnak a geokódok, a hivatal az első, pontosabb eredményt adó módszert alkalmazta.

Minden adatbázist, ami a projekt során előállításra került, ingyenesen és szabadon hozzáférhetővé és használhatóvá kellett tenni az Eurostat számára. Az adatokat összeállítják egy európai rácsháló adatbázissá (GEOSTAT 2011), ami nyilvánosan közzéteszik az Eurostat honlapján.

4. ábra

Budapest és környéke lakónépességének 1 km²-es gridtérképe



Forrás: saját szerkesztés.

A munkafolyamat lépései

Először az összes munkafolyamatot teszteltük egy megyén, majd a módszer teljes kidolgozása után hajtottuk végre a folyamat összes lépését az egész országon. Mivel az adatvédelem során mikroadatokkal kellett dolgozni, a nagy rekordszám miatt külön kezeltük a megyéket, így 20 különböző fájlal dolgoztunk, amelyek a végén az aggregálásnál kerültek egybe.

A lépések:

1. Áttanulmányoztuk a témával kapcsolatos nemzetközi szakirodalmat, különös figyelmet fordítva az alkalmazható adatvédelmi eljárásokra. Úgy láttuk, a *grant* kiírás szigorú feltételeinek egyedül a célzottadatsere-módszer felel meg. A módszerről részletesebben a későbbiekben lesz szó.

2. Leválogattuk az összes lakott címet és hozzájuk tartozó népességadatokat a 2011-es népszámlálási címállományból.
3. Letöltöttük a rácsháló(*grid*)térképet Magyarország területére. A rácshálót Európa területére hozták létre, és darabolták fel az országok területére, így a határok mentén gond nélkül össze lehet azokat illeszteni.
http://www.efgs.info/data/eurogrid/Grid_ETRS89_LAEA_HU_1K.zip/view
4. Az adatbázis előállításához a címpontok rácshálókhoz való rendelését az ArcMap térinformatikai szoftverben végeztük el. Ezzel minden címpont megkapta *grid*-azonosítóját.
5. A fedél nélküli hajléktalanokat elhelyeztük az adott település legnépesebb *grid*jében.
6. Kidolgoztuk a magyarországi adatállomány esetén alkalmazandó részletes adatvédelmi eljárást. Az adatvédelmet még az aggregálás előtt a mikroadatokon kellett alkalmazni.
7. Egy megye adatállományán futtattuk az adatvédelmi programot. A kapott eredményt értékeltük, ahol szükséges volt az adatvédelmi eljárás módosítása, ott ezt elvégeztük.
8. Az előállt mikroadat-állományt adatvédelemnek vetettük alá.
9. Az adatvédelem mikroadat-állományt SAS-ban aggregáltuk, kiszámítottuk a pályázatban előírt mutatókat, így létrejött az országos adatbázis a már említett 12 mutatóval.
10. A végső adatbázisban minden *grid* minden mutatójánál feltüntettük, hogy történt-e adatvédelem.

Adatvédelem

A hivatalos statisztikai adatok térstatisztikai módszerekkel történő megjelenítésének egyik nagy kihívása az adatvédelem¹ kérdése. Az adatvédelem fő feladata, hogy különböző matematikai-statisztikai módszerek és eljárások alkalmazásával minimalizálja a statisztikai egységek² azonosításának³, illetve felfedésének⁴ a kockázatát úgy, hogy közben a lehető legkisebb információvesztést okozza. Azonban nemcsak a statisztikai felfedés elleni védelmet⁵ szükséges figyelembe venni az adatvédelem során, hanem a hazai irányadó jogszabályokhoz⁶ is minden esetben alkalmazkodni kell.

1 Adatvédelem: minden olyan jogi, módszertani, fizikai, informatikai és egyéb eljárás, módszer, tevékenység, amelyet annak érdekében alkalmaznak, hogy a rendelkezésre álló adatokat a jogosulatlan, céltól eltérő felhasználástól megóvják.

2 Statisztikai egység: a sokaság meghatározott tulajdonságokkal, jellemzőkkel, más néven ismérvekkel rendelkező egyede.

3 Azonosítási kockázat: annak valószínűsége, hogy a támadó legalább egy statisztikai egységet azonosít a közzétett adatok alapján.

4 Felfedési kockázat: az azonosítási kockázat, valamint a statisztikai egységre vonatkozó új információ nyilvánosságra kerülésének valószínűsége.

5 Statisztikai felfedés elleni védelem: a statisztikai célra összeállított adatállományok módosítását célzó olyan módszerek összessége, melyek a lehető legnagyobb mértékben csökkentik a felfedési kockázatot. Az adatállományok statisztikai felfedés elleni védelmének legfontosabb célja annak megakadályozása, hogy egyedi adatok jelenjenek meg bennük.

6 1993. évi XLVI. törvény a statisztikáról, amelyben a 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”. Ehhez kapcsolódik a 170/1993.(XII.3.) számú kormányrendelet a statisztikáról szóló 1993. évi XLVI törvény végrehajtásáról, amelyben a 19. § kimondja, hogy „összesítve sem lehet nyilvánosságra hozni olyan adatot, amelynél az adatszolgáltatók száma háromnál kevesebb”.

A területi változó mindig kiemelt figyelmet igényel a felfedés elleni védelem végrehajtása során, hiszen minél kisebb a vizsgált területi egység, és minél kisebb esetszámok fordulnak elő (melyet az adatvédelem területén abszolút gyakoriságnak nevezünk), az adatvédelmileg annál nagyobb kockázatot jelent. Alacsony gyakoriságú cellaértékeknel ugyanis megnő a kockázata az egyedek beazonosíthatóságának, így jelentősen megnövekszik a felfedés lehetősége is. Ez a jelenség a térstatiszti területén hatványozottan megjelenik, ahol az adatok speciális megjelenítése tovább nehezíti az adatvédő dolgot.

A felfedés elleni védelmi módszereknek többféle csoportosítása létezik. Megkülönböztetünk perturbatív és nem perturbatív eljárásokat az alapján, hogy azok módosítják-e a változók eredeti értékeit. Perturbatív adatvédelmi eljárások közé tartozik pl. a kerekítés, rekordcsere, míg a cellaelnyomás,⁷ amely a KSH-ban az egyik legáltalánosabban alkalmazott felfedés elleni védelmi módszer, a nem perturbatív eljárások közé tartozik. Egy másik csoportosítási lehetőség azon alapul, hogy a felfedés elleni védelmet még a mikroadatokon⁸ vagy már a mikroadatokból képzett aggregátumokon (táblázatos adat⁹) hajtjuk-e végre.

A GEOSTAT-projekt kapcsán a célzottadatcsere-módszert (*targeted record swapping*) használtuk felfedés elleni védelemként. A kiválasztott módszert mikroadatokon szükséges alkalmazni. Ez a módszer elegendő védelmet biztosít ahhoz, hogy a mikroadatból keletkező aggregátumokban ne kelljen további felfedés elleni védelmet alkalmazni.

A KSH által használt cellaelnyomás módszerét a GEOSTAT 1C esetében nem lehetett alkalmazni, mert a projekt olyan követelményeket állított a felfedés elleni védelemre vonatkozóan, melyet a cellaelnyomás nem tudott volna teljesíteni. A fentebb felsorolt feltételekből az 1., a 4., a 6. és a 7. pontok kapcsolódtak szorosan a felfedés elleni védelem kialakításához.

Célzottadatcsere-módszere

Első lépésként a felfedés elleni védelemmel ellátandó, azaz kockázatos rekordokat kell meghatározni a mikroadatban. Ezek azok a rekordok, amelyek a *nem*, *korcsoport* és *grid*, illetve ezen jellemzők kombinációi (például *nem*–*korcsoport*; *nem*–*korcsoport*–*grid* stb.) alapján 3 alatti gyakoriságban fordulnak elő. A felfedés ellen védendő rekordokat megjelöltük a mikroadat állományban. A rekordok megjelölése azért fontos, mert minden egyes felfedés elleni védelemmel ellátandó rekordot adatcsere alá vontunk. A célzott adatcsere során minden egyes felfedés elleni védelemmel ellátandó (*A*) rekordhoz különböző feltételeknek megfelelő (*B*) „partnert” kerestünk.

A feltételek a következők voltak:

- a pároknak eltérő korcsoporttal kellett rendelkezniük,
- a pároknak eltérő gridben kellett elhelyezkedniük,
- a vizsgált településen belül kellett tartózkodniuk.

⁷ Cellaelnyomás: táblázatos adatok védelmére alkalmazott eljárás, amelynek lényege, hogy valamilyen táblázatos adatok védelmére alkalmazandó módszer(ek) eredményeként érzékenynek ítélt cella értéke nem közölhető, hanem egyezményes jellel helyettesítendő.

⁸ Mikroadat: olyan, rekordok sorozatából álló állomány, amely megfigyelési egységek adatait tartalmazza. Az aggregált adatok előállításának elsődleges forrása.

⁹ Táblázatos adat: olyan táblázatos formába rendezett adatállomány, amely aggregált adatokat tartalmaz.

A lehető legkisebb információvesztés céljából elsősorban olyan gridekben kerestünk cseréhez alkalmas párokat, amelyek minél kisebb gyakoriságúak, és amelyekben legalább egy felfedés elleni védelemmel ellátandó rekord szerepelt.

Miután a „párkeresés” folyamata lezárult, kialakultak a cserére alkalmas rekordpárok, akkor az egymáshoz rendelt rekordok „nem” és „korcsoport” szerinti értékeit kicseréltük. A csere után a cserében résztvevő rekordok „korcsoport” változójának értékei minden esetben megváltoztak, (mivel az „partner” keresési feltétel volt), míg a „nem” változóé csak azon esetekben, mikor két ellentétes nemű rekord került egy párba.

A fenti módszer biztosítja azt, hogy a teljes népességre és különböző népességbontásokra vonatkozó adatok megegyezzenek a népszámlálásból az ország összesenre vonatkozó népességszámokkal, továbbá a gridekre vonatkozó gyakoriságok is változatlanok maradnak.

A grideken belüli „korcsoport” és „nem” szerinti megoszlás az eredetitől eltér, biztosítva azt, hogy minimális kockázatú legyen az egyedek azonosíthatósága, valamint felfedése.

1. táblázat

*Gridenkénti aggregált fiktív adatok korcsoportos bontásban,
célzott adatsere előtt és után*

Grid	Célzott adatsere előtti értékek				adat- véde- lemmel ellátott	Célzott adatsere utáni értékek			
	0–14	15–64	65+	összesen		0–14	15–64	65+	összesen
1kmN2640E5003	2	15	4	21	adv	4	14	3	21
1kmN2640E5004	3	17	3	23	adv	1	17	5	23
1kmN2640E5005	3	17	7	27	adv	0	20	7	27
1kmN2640E5006	0	4	4	8	adv	2	2	4	8
1kmN2640E5007	1	5	0	6	adv	3	2	1	6
1kmN2640E5011	1	2	0	3	adv	1	1	1	3
1kmN2640E5012	47	254	109	410		47	254	109	410
1kmN2640E5013	61	327	121	509		61	327	121	509
1kmN2640E5015		10	3	13	adv	2	11		13
1kmN2640E5017	123	308	124	555		123	308	124	555
1kmN2640E5030	69	307	98	474		69	307	98	474
1kmN2640E5090		1	2	3	adv	1	2		3
<i>Összesen</i>	<i>7 000</i>	<i>34 000</i>	<i>8 000</i>	<i>49 000</i>		<i>7 000</i>	<i>34 000</i>	<i>8 000</i>	<i>49 000</i>

Forrás: saját számítás.

A példa alapján összehasonlítható, hogy a célzott adatsere módszere során mely cellaértékek változtak, illetve melyek változatlanok.

Több ország alkalmazza az alacsony gyakoriságú cellák helyettesítését bizonyos értékekkel pl. „3”; „9999”; „...”, de számos esetben a cellák ilyen fajta elnyomása, helyettesítése túl nagy információvesztést okoz. A célzott adatcsere alkalmazásával nem csak alacsony információvesztés érhető el, de az adatok speciális megjelenítése sem akadály, így hasznos felfedés elleni védelmi módszer a térstatistikában is. A fent bemutatott módszert első alkalommal használta a KSH, ami kitűnő példája annak, hogyan lehet egy általános felfedés elleni védelmi módszert a saját igényeinkre szabni.

A KSH hozzájárulása az európai térinformációs rendszerhez

Az Európai Parlament és a Tanács az Európai Közösségen belüli térinformációs infrastruktúra (INSPIRE) kialakításáról szóló irányelvének (2007/2/EK) célja, térinformáció biztosítása a közösségi politikák kialakításához, végrehajtásához. Elsődlegesen a környezet állapotával szoros összefüggésben lévő adatok elérhetőségét, felhasználhatóságát hivatott biztosítani.

Az irányelv a tagállamok téradat-infrastruktúrájának összehangolását, összekapcsolását és szabványosítását tűzi ki célul, szabályozza a metaadatokat, a téradat készletek és szolgáltatások szabványosításának és interoperabilitásának, a hálózati szolgáltatásoknak és az adatok megosztásának kérdését az Európai Unió szintjén és a tagállamokra érvényesen. Fontos szempont, hogy a tagállamok saját nemzeti szintű megoldásait az INSPIRE-höz igazítva hozzák létre.

A jogszabály a téradatokat téradattémákba sorolja, s ezek az irányelv I., II. mellékletében (alap téradatok) és III. mellékletében (tematikus téradatok) szerepelnek.

A KSH a *statisztikai egységek, a népesség eloszlása – demográfia*, illetve *mezőgazdasági és akvakultúra-ágazati létesítmények* téradattémakon¹⁰ belül közölt adatokat. A *statisztikai egységek* téradattémán belül 1 km²-es és 25 km²-es gridterképeket publikáltunk (shp formátumban letölthetők).

A *népesség eloszlása – demográfia* téradattémában a 2011-es népszámlálásból származó népesség- és lakásszámadatokat tüntettük fel 1 km²-es griden.

A *mezőgazdasági és akvakultúra-ágazati létesítményekhez* sorolható a 2010-ben végrehajtott általános mezőgazdasági összeírásból származó gazdaságok száma, mezőgazdasági munkaerőadatok 25 km²-es griden történt megjelenítése.

A megjeleníthető adatok köre bővíthető például korcsoportokat, illetve az állatállomány sűrűségét feltüntető térképekkel, azonban ezek esetében adatvédelmi kérdéseket is figyelembe kell venni.

10 Statisztikai egységek: statisztikai információk terjesztésére vagy felhasználására szolgáló egységek.

Népesség eloszlása – demográfia: a népesség földrajzi eloszlása – beleértve a népességi jellemzőket és a tevékenységi szinteket is – földrajzi rácsrendszer, régió, közigazgatási egység vagy más elemzési egység alapján összesítve.

Mezőgazdasági és akvakultúra-ágazati létesítmények: mezőgazdasági eszközök és termelőlétesítmények (beleértve az öntözőrendszereket, üvegházakat és istállókat).

Térinformatikai módszerek alkalmazása a mezőgazdasági statisztikában

A mezőgazdasági statisztikai összeírások egyik alapkérdése a gazdaságok, a gazdálkodók és a mezőgazdasági létesítmények elérése, megtalálása. Ennek többféle lehetősége is van, a módszer kiválasztását nagyban befolyásolja az összeírás célsokasága, témája. Az összeírásoknak elérési szempontból két nagy csoportja létezik, az egyik az, amikor az összeírás tárgyát képező helyre mennek ki az összeírók, a másik, amikor az összeírás célját képező terület használóját keresik fel. Az első csoportba a szőlő-gyümölcs összeírások tartoznak, ebben az esetben az összeírók a gyümölcsösökben végzik el az adatfelvételt. A második csoportba tartoznak az általános mezőgazdasági, illetve gazdaságszerkezeti összeírások, amikor az összeírók a földhasználókat, gazdálkodókat keresik fel. Ilyenkor az egyéni gazdaságok esetében a gazdát a lakóhelyén érik el, a gazdasági szervezetek képviselőit pedig a szervezet székhelyén, illetve újabban a gazdasági szervezetek adatszolgáltatása elektronikus úton, önkitöltéssel történik.

A szőlő-gyümölcs összeírások esetében az ültetvények helyének meghatározása a 2000-es évek elejéig helyrajzi számmal történt, a legutóbbi összeírás során áttért a hivatal a geokódok használatára. Az összeírók ebben az esetben ortofotókat kaptak, ahol az ültetvényt geokódja alapján tüntették fel. Az ültetvények egy részénél csak helyrajzi számmal rendelkezünk, ebben az esetben az összeírók feladata volt a geokód felvétele.

Az általános mezőgazdasági, illetve a gazdaságszerkezeti összeírások idején az összeírók a gazdálkodókat keresik fel, ehhez címadatokat használnak. Ebben az esetben is meghatározásra kerül a tevékenység helye, a gazdáknak meg kell adniuk a legfontosabb földterületük/telephelyük helyrajzi számát vagy MEPAR-blokkazonosítóját.¹¹

A legfontosabb földterület/telephely azonosítása a következőképpen történik:

A gazdaság elhelyezkedését az alapján kell meghatározni, hogy jellemzően (egészében vagy nagyobb részben) hol történik a mezőgazdasági termelés. A meghatározás kétféleképpen valósulhat meg. Ha a termelés a gazdálkodó lakhelyétől/gazdaság székhelyétől legfeljebb 5 km-re folyik, a laccím/székhely címe (telek) vagy a parcella adatai egyaránt szolgálhatnak azonosítóként. Ha a termelés a lakhelytől/székhelytől 5 km-en kívül folyik, akkor a legfontosabb földterület (parcella) vagy a legfontosabbnak ítélt mezőgazdasági rendeltetésű épület vagy építmény (istálló, üvegház, gépszín stb.) elhelyezkedése szolgál azonosításra. A földterület fontosságát a következő sorrend szerint kell megállapítani: szántóföldi intenzív termelés > ültetvények > gyep; míg az egyes kategórián belül a terület nagyságát kell figyelembe venni. Az így kapott információkhoz történik a geokódok meghatározása.

¹¹ MEPAR-blokk: a mezőgazdasági táblák nagyobb tömbökben, ún. fizikai blokkokban helyezkednek el. A fizikai blokk a mezőgazdasági művelés szempontjából időben állandó, a terepen azonosítható határokkal (pl.: utak, vasutak, csatorna) rendelkezik, és többnyire azonos típusú művelés alatt lévő földterület (pl. szántó, gyep, erdő). Egy fizikai blokkban általában több mezőgazdasági tábla van, területét több gazdálkodó is művelheti. A blokkok országosan egyedi azonosítóval vannak ellátva.

A fizikai blokkokat a Földmérési és Távérzékelési Intézet (FÖMI) alakította ki az egész országra vonatkozóan. A fizikai blokkon belül a különböző mezőgazdasági hasznosítások (pl. szántó, gyep, halastó, mozaikos művelés), valamint a beépített és infrastruktúrának használt területek vannak elkülönítve.

(http://www.fomi.hu/honlap/magyar/projektek/leirasok/MePAR_kiadvany_akt_v06.htm)

A gazdaságok esetében a termelés helye legtöbbször a lakóhelytől/székhelytől 5 km-en belüli területre esik, így a lakóhelyük/székhelyük koordinátaival reprezentálhatjuk termelési helyüket is. Ez elfogadható közelítés, így saját településük területére fog „számítani” a termelési helyük. Természetesen itt az egyes termelési típusokat nézve is különbségek tapasztalhatók, hiszen például az állattenyésztést végzők biztosan a gazdálkodás helyéhez közel laknak, míg a szőlőtermesztés nagyobb eséllyel történik lakóhelyüktől távolabbi területen.

A gazdaság szintű tér adatok rendelkezésre állása azért fontos, mert lehetővé teszik a közigazgatási határoktól eltérő tájékoztatást. Adatainkat például megjeleníthetjük, sűrűség- vagy más néven grid térképeken, illetve készíthetünk térképeket a borvidékekre vonatkozóan is.

A mezőgazdasági összeírások mellett a felszínborítás, földhasználat témakörében van jelentős szerepe a térinformatikai alkalmazásoknak, elsősorban a műholdas távérzékelésnek és az ortofotóknak. A hivatal erre vonatkozó adatgyűjtéséből származó adatait jól kiegészítik a térinformatikai forrásból származó információk. A fő forrásaink a CORINE-¹², a MEPAR- és a LUCAS-adatbázis.

A jövőben nagyobb figyelmet fordítunk ezek ötvözési lehetőségeire, egymást kiegészítő voltokra. Ebben már előrelépések is történtek, ugyanis 2014 decemberében indult egy, a Földmérési és Távérzékelési Intézet (FÖMI) közös projekt, amelynek keretében az Eurostat által koordinált LUCAS-felmérés 2012-es felszínborítási és földhasználati adatait hasonlítjuk a hazánkban elérhető adatforrásokhoz. A projekt fő célja a hazai földhasználat és felszínborítás adatforrások összegyűjtése és ezekből a LUCAS kategóriáinak megfelelő adatok összeállítása.

A projekt kezdeti szakaszban tart most, jelenleg az adatforrásokat összegyűjtő „leltár” van készen. Ezt az adatforrások minőségének, használhatóságának értékelése fogja követni. Végezetül pedig módszertant fogunk kidolgozni a LUCAS-kategóriák más adatforrásokból történő létrehozására.

12 A CORINE Land Cover az EU által az 1980-as években indított projekt, amelynek célja a kvantitatív, megbízható és összehasonlítható felszínborítási információ biztosítása az EU területére, amelyeken keresztül a felszínborítás és annak változásának ismerete segíti az összehangolt európai környezeti politika kialakítását. Az alapadatokat műholdfelvételek biztosítják, amelyek alapján öt csoportba rendezett felszínborítási kategóriákat hoztak létre.