

FÜGGELÉK

ADMIN3 – A KRTK KAPCSOLT ÁLLAMIGAZGATÁSI PANELADATBÁZISA

SEBŐK ANNA

2019 nyarán a Közgazdaság- és Regionális Tudományi Kutatóközpont (KRTK) Adatbankja harmadik alkalommal hozta létre a Kapcsolt Államigazgatási Paneladatbázist, amelyre kutatói körökben röviden csak Admin3-ként hivatkoznak. Az Admin3 a Nemzeti Egészségbiztosítási Alapkezelő (NEAK), a Magyar Államkincstár (MÁK), az Oktatási Hivatal (OH), a Pénzügyminisztérium (PM) és a Nemzeti Adó- és Vámhivatal (NAV) egyéni és vállalati szintű adatai összekapcsolásával jött létre. Az adatbázis tehát anonimizált módon, ugyanakkor egyéni szinten tartalmazza a magyar lakosság 50 százalékos véletlen mintájának többéves egészségügyi, oktatási, munkaerőpiaci és munkanélküliségi adatait, valamint a magyarországi vállalatok számos jellemzőjét.

Az Admin3 (2003–2017) adatintegrációs eljárással készült. Az adatösszekötés alapját jelentő minta a NEAK mint a magyar népességet majdnem teljes egészében lefedő nyilvántartó adatbázisából származik. A leválogatás során 50 százalékos véletlen mintát vettek a 2003-ban TAJ-számmal rendelkezők közül, így jött létre a többi adatgazda számára is ismeretes kapcsolati kódok (TAJ-szám és foglalkoztatói adószám) listája. Az anonim adatkapcsolásért felelős Nemzeti Infokommunikációs Szolgáltató (NISZ) Zrt. által, kifejezetten az adott összekötéshez generált *hash*-algoritmus segítségével az eredeti kódokhoz egyedi, technikai azonosítókat rendelt a NEAK. A következő lépésben a többi adatközlő is leválogatta az alapsokasághoz kapcsolódó – az egészségügy kivételével a 2003–2017 közötti időszakra vonatkozó – adatait, és *hashed* állapotban továbbadta a NISZ Zrt.-nek, amely végül egyesítette és anonimizálta az adatbázist. Az egyesítés után az eredeti azonosítókat (például a TAJ-számot és a foglalkoztatói adószámot) elvesztve, ugyanakkor a személyek adatait összekötve kapta meg az adatállományt a KRTK Adatbank nyers, kutatásra még alkalmatlan formában. Az adatbázis tisztítását együtt végezték és végzik folyamatosan az adott nyilvántartással régóta dolgozó kutatók és az Adatbank munkatársai. Az anonimizált adatok védett szerveren, illetve kutatószobában elemezhető kutatási célra.

Az egészségügyi területen az alábbi információk érhetők el a 2009–2017 évekre vonatkozóan: TAJ-regiszter (nem, születési év és hó, valamint TAJ-érvényességi információk), lakóhely járása, társadalombiztosítási jogviszonyra vonatkozó, közgyógyellátási, háziorvosi, járóbeteg- és fekvőbeteg-ellátási, halálozási, vénykiváltási, pénzügyi ellátási (táppénz-, tgyás-, csed-, gyed- és betegszabadság-) adatok, egyaránt egyéni szinten. A Munkaerőpiaci tükrőben

ezek közül elsősorban a lakóhely járását, a háziorvosi megjelenések számát, a járóbeteg-szakellátási esetszámot és kiadást,¹ a kórházi tartózkodási napok számát és kiadást,² a kiváltott vények számát és kiadást³ – összesen és a főbb hatóanyagkategória- (*Anatomical Therapeutic Category, ATC*) csoportok szerint –, valamint a halálozást használjuk. Figyelembe kell azonban venni, hogy az adatok aggregálásával – a némiképp eltérő adatkör és a mintavétel jellege miatt – nem pontosan kaphatók vissza az aggregált egészségügyi statisztikák.

A munkaerőpiaci területről munkavállalói, közfoglalkoztatási és munkaerő-kiközvetítési adatkörök szerepelnek, egyéni szinten. A Munkaerőpiaci tükrőben ezek közül elsősorban a munkapiaci státust és a bérjövédelmet használjuk.

A szociális transzferek területén az adatbázis nyugdíjfolyósítási, pénzbeli ellátási, munkanélküliségi és a munkaerőpiaci programokhoz kapcsolódó adatokat tartalmaz egyéni szinten.

Az oktatási területhez kapcsolódó adatkörök: a felsőoktatási képzési részvétel, a felsőoktatási hallgatói jogviszony, a köznevelési tanulói jogviszony, az érettségi, illetve az országos kompetenciamérés egyéni szintű adatai.

Az adatbázisban szereplő vállalati információk a társaságiadó- (tao) bevéletről és a Bértarifa-felvételről származó adatok, vállalati szinten, ugyanakkor munkavállalókhoz köthetően.

Az adatbázisról részletesen *Sebők* (2019) cikkében lehet olvasni.

Hivatkozás

SEBŐK ANNA (2019): A KRTK Adatbank Kapcsolt Államigazgatási Paneladatbázisa. *Közgazdasági Szemle*, 66. évf. 11. sz. 1230–1236. o. <https://doi.org/10.18414/KSZ.2019.11.1230>.

¹ Az itt használt járóbeteg-szakellátási adatok tartalmazzák a laboratóriumi ellátást, de nem tartalmazzák a CT-és MRI-diagnosztikát.

² Beleértve az aktív és krónikus ellátást is.

³ Külön rendelkezésre áll a társadalombiztosítás által finanszírozott és a beteg által fizetett rész is.