

# Államigazgatási adatgyűjtemény

Az államigazgatási adatgyűjtemény az OEP-ONYF-NAV-NMH-OH<sup>1</sup> – adminisztratív - adatfájlok összekapcsolása által jött létre. Az állomány mintája a megfigyelt időszak kezdetén, 2003-ban 5-74 éves népesség fele. Az állományok a mintába bekerült személyekről demográfiai adatokat, valamint ellátásra és jogviszonyra vonatkozó információkat tartalmaznak, 2011-ig. A feldolgozott, tisztított, harmonizált adatokat több fájlban tároljuk: a törzsállomány tartalmazza a demográfiai adatokat és a státuszváltozókat (kb 26 gb), a kiegészítő állományok pedig a státuszokra vonatkozó egyéb információkat, havi felbontásban. Az állományok 13-es verziójú Stata formátumban érhetők el.

---

<sup>1</sup> OEP: Országos Egészségbiztosítási Pénztár, ONYF: Országos Nyugdíjbiztosítási Főigazgatóság, NAV: Nemzeti Adó- és Vámhivatal, NMH: Nemzeti Munkaügyi Hivatal, OH: Oktatási Hivatal

# Tartalomjegyzék

A tisztított, harmonizált államigazgatási adatgyűjtemény hozzáférhetősége .....	3
A feldolgozott állományok ismertetése .....	3
A mintavétel .....	3
Kapcsolás .....	4
Kiinduló állományok .....	4
OEP .....	4
ONYF .....	4
NAV .....	5
NMH .....	5
OH .....	5
Az államigazgatási adatgyűjtemény felépítése .....	5
Tisztítási folyamat .....	6
Szerkezeti átalakítás .....	6
Folytonos időről diszkrét időre történő átalakítás .....	6
A megfigyelési egység .....	7
Hibás megfigyelések törlése, javítása .....	7
Lezáratlan státuszok kezelése .....	8
Összevont státuszkategóriák kialakítása .....	9
Az államigazgatási adatgyűjtemény változói .....	10

## **A tisztított, harmonizált államigazgatási adatgyűjtemény hozzáférhetősége**

A tisztított, harmonizált adatbázis a MTA KRTK rendes és vendégkutatói számára érhető el.

### **A feldolgozott állományok ismertetése**

Az összekapcsolt OEP-ONYF-NAV-NMH-OH adatfájlokat több állományra bontottuk. A törzsállomány tartalmazza a megfigyelt személyek személyes adatait (születési dátum, nem, lakóhely régiója) valamint munkaerőpiaci szempontból fontos státuszait. A kiegészítő állományok ezen státuszokról szolgáltatnak további információkat.

A törzsállomány 4 601 999 fő – akik 2003-ban 5-74 évesek - életpályáját követi nyomon 2003-2011 között. Az állományban a megfigyelési egység egy személy egy havi jogviszonya vagy ellátása (így tehát egy személynek 108 megfigyelése van).

A kiegészítő állományokban szereplő személyek száma azért lehet kevesebb a törzsállományhoz képest, mert nem minden személy szerepel azokban az állományokban, amelyekből a kiegészítő változók származnak, illetve ha szerepelnek is, egyik kiegészítő változó sem vesz fel értéket esetükben.

Az egyes kiinduló állományok 2011-ig vagy 2012-ig állnak rendelkezésre. A törzsállomány javarészt az ONYF adatokra épül, így az alap adatok 2011-ig tartalmazzak megfigyeléseket. Egyes kiegészítő állományok szintén 2011-ig állnak rendelkezésre emiatt, azonban ahol lehetett, és önmagában is hasznos kiegészítő állományok állhattak elő, ott megtartottuk a 2012-es adatokat is.

### **A mintavétel**

A kapcsolás során az OEP képezte a mintát.

A kiválasztott személyekhez a 2003-2011/2012 években (összesen 9/10 év) megfigyelt adatai lettek hozzárendelve (mindaddig, amíg a kiválasztott személyek az adatbázisban szerepeltek). A minta elemszáma időben előre haladva nem lett pótolva, aminek következménye, hogy az folyamatosan csökken az elhalálozások végett.

Nem minden megfigyelt személy esetében látjuk a 2003-2011 közötti évekre a teljes életpályáját. Az állományban szerepelnek olyan hónapok – esetleg évek – amikor semmilyen információnk nincs adott személyről.

## **Kapcsolás**

### ***Kiinduló állományok***

#### **OEP**

Az OEP-től öt állományt kaptunk 2003-2012-re vonatkozólag:

- OEP TAJ-BSZJ: demográfiai adatokat nyilvántartó adatbázis (4 602 489 személy), olyan személyekről is, akik az időszak során semmilyen jogvisztonnyal sem rendelkeznek
- OEP jogviszony: a bejelentett személyek jogviszonyait (4 512 894 személy esetén) mutatják, annak kezdő és befejező dátumával együtt
- OEP pénzbeli ellátás (1 457 616 fő): táppénz, TGYÁS, GYED ellátásokról ad információt, annak kezdő és befejező dátumával együtt
- OEP betegellátás (4 520 968 fő): éves adatbázis különböző betegellátási adatokról
- OEP közgyógyellátás (329 921 fő): a közgyógyellátásra való jogosultság jogcímét tartalmazza, annak kezdő és befejező dátumával együtt

#### **ONYF**

Az állomány 2003-2011 közötti időszakban 3 131 344 nyugdíjjogosult személy éves, a KELEN adatbázisból származó nyugdíjjogosultság-biztosítási adatait tartalmazza időszakonkénti bontásban. Tehát minden olyan jogviszonyt és ellátást, amely alapján az egyén nyugdíj-jogosultságot szerzett. Egy jogosultnál a különböző jogcímen, különböző időszakban nyilvántartott adatok egymás alatti sorokban szerepelnek.

A NYUFIG adatbázisból a 2003-2011-es időszakra vonatkozó nyugdíj és nyugdíj jellegű ellátásokra vonatkozó adatok szerepelnek. Ez 1 729 474 személy időszaki adatait tartalmazza.

## **NAV**

Az állomány megfigyelési egységei egy vállalat éves (2003-tól 2011-ig), egyes mérleg és eredménykimutatásból származó adatai. Az adatbázis a foglalkoztatói adószámból képzett HASH kódok segítségével az ONYF egyéni, munkavállalói adataihoz kapcsolható. Összesen 426 738 vállalatról van – legalább egy évben – adat.

## **NMH**

Az NMH adatok 4 állományban találhatóak meg a 2004-2012 közötti időszakra vonatkozólag.

- A nyilvántartási regiszterből származó adatokban egy személy (összesen: 1 383 373 fő) egyszeri munkaerőpiaci részvétele, szakrendszerei kódja szerepel.
- Az álláskeresési járadék regiszterből származó adatok fájljában a megfigyelési egység egy személy (összesen: 996 758 fő) egyszeri ellátása.
- A munkaerőpiaci programokra vonatkozó adatbázisban 368 109 fő programban/szolgáltatásban való részvétele szerepel, annak kezdő és befejező dátumával együtt.
- A regisztrált munkanélküliek közvetítési adatainak megfigyelési egysége egy személy (összesen: 659 729 fő) egyszeri kiközvetítése.

Mindegyik adatbázisra igaz, hogy amennyiben a megfigyelt többször volt nyilvántartva vagy részesült adott személy ellátásban, vett részt programban, akkor az külön sorokon jelenik meg.

## **OH**

Az oktatási adatok a köznevelési és felsőoktatási információs rendszerből származnak. 1-1 adatbázis a tanulói státuszokról, a végzettségekről és a képzőhelyekről tartalmaz éves, 2009-2012-es adatokat.

### ***Az államigazgatási adatgyűjtemény felépítése***

A státuszok (ellátások és jogviszonyok) egy része egynél több adatbázisban is megjelenik. A különböző forrásból származó állományok összevetésekor azonban azt tapasztaltuk, hogy az azonos tartalmú státuszok időtartama gyakran nem egyezik meg teljesen ugyanazokat az egyéneket vizsgálva. Továbbá, gyakran egymást kizáró státuszok ugyanabban az időszakban jelennek meg a megfigyeléseknél. Ezen okok miatt az államigazgatási

adatgyűjteményt úgy építettük fel, hogy minden státuszt csak egy állományból jelenítettünk meg, mégpedig abból az állományból, amelyet – illetékességi, pénzügyi transzfer stb. szempontok alapján – a legmegbízhatóbbnak ítéltünk.

Ezek alapján az államigazgatási adatgyűjteményben megjelenő státuszok forrása:

<b>Státusz</b>	<b>A státusz forrása (állomány)</b>	<b>Fellelhető még</b>
GYET, GYES	ONYF	OEP jogviszony
GYED, táppénz, GYED	OEP ellátás	OEP jogviszony, ONYF
TGYÁS	ONYF	OEP jogviszony, OEP ellátás
nyugdíj és nyugdíj jellegű ellátások	ONYF	OEP jogviszony
munkanélküli ellátások	NMH nyilvántartási és járadékos valamint OEP jogviszony (utóbbiból csak a szociális segélyek)	OEP jogviszony, ONYF
munkanélküli - ellátás nélküli	NMH nyilvántartási	
a fentiekén kívül minden státusz	ONYF	az ONYF kódok nagy része megtalálható az OEP jogviszonyban is

## **Tisztítási folyamat**

### **Szerkezeti átalakítás**

#### **Folytonos időről diszkrét időre történő átalakítás**

Az OEP-től, ONYF-től, NMH-től beérkező adatok folytonos időben voltak megadva, azaz időszakokra vonatkoztak. Ezeket időpontra alakítottuk át, úgy hogy az adott egyén jogviszonyának típusát vagy ellátásban való részesülését adott hónapra vonatkozva is lehessen látni.

Módszer:

- A jogviszonyokat és ellátásokat mindig adott hónap 15-én vizsgáltuk, és ezt az adott hónap egészére vonatkoztattuk. Egy napon több jogviszonya/ellátása is lehet egy adott embernek.
- Ha egy jogviszony adott hónap 15-én vagy előtte kezdődik, akkor már az adott hónap esetén megjelenik ezen státusz az állományban, kivéve ha még adott hó 15-dike előtt lezárt is lesz.

- Ha egy jogviszony vagy ellátás adott hónap 16-án (vagy később) indult, akkor az csak a következő hónapnál jelenik meg, mint jogviszony vagy ellátás, amennyiben a következő hónap 15-én még fennáll az a státusz. Amennyiben az adott jogviszony vagy ellátás adott hónap 16-án vagy utóbb kezdődik, és a következő hónap 15-dike előtt be is fejeződik, úgy ezen megfigyelések az átalakított állományban nem szerepelnek.
- Az 1 hónapnál rövidebb jogviszonyok és ellátások tehát csak abban az esetben jelennek meg az adatok között, amennyiben adott hónap 15-én is fennálltak. (Például egy január 16-tól február 14-éig tartó megfigyelés nem szerepel az átalakított adatbázisban).

Ez alól az alapfájlban két jogviszonyra vonatkozó változó („nap”, „w”) jelent kivételt. Részletesen lásd a változókatalógust.

## **A megfigyelési egység**

A diszkrét időre történő átalakítás után az állomány szerkezetét úgy alakítottuk ki, hogy egy megfigyelési egység egy ember egy havi státusza (jogviszonya, ellátása) legyen. Így egy ember 108 soron szerepel. Ha egy személynek ugyanabban a hónapban több státusza is volt, akkor azok egymás melletti oszlopokban jelennek meg sorszámozva (pl: status1, status2, status3).

Az állomány nagy mérete nem engedte meg azt, hogy minden változóból annyi variációt hozzunk létre, ahány a legtöbb különböző értéket felvevő változó mennyisége. Adott foglalkoztatási jogviszonyra vonatkozó változóból így csak kettő – a legnagyobb keresetet jelentő – szerepel az állományban. Viszont az összevont transzferre vonatkozó változókból minden személy esetén annyi az előfordulások száma, ahány a legtöbb különböző értéket felvevő ember esetében az előfordulás

## **Hibás megfigyelések törlése, javítása**

Az állományok tisztítása során töröltük azokat a személyeket – minden megfigyelési egységükkel együtt - :

- akik ugyanolyan egyéni azonosító kóddal rendelkeznek, de az összes állományt nézve legalább 2 különböző születési idővel vagy nemmel szerepeltek – ez 3 239 személyt érintett;

Továbbá törlésre kerültek az alábbi megfigyelések (egy személy egy időtartamra vonatkozó státusza) is:

- ahol a jogviszony időtartama feltételezhetően hibás volt (pl. előbb volt vége, mint a kezdete).

Ezen esetekben nem adott személy minden sorát töröltük, csak azokat a státuszait (jogviszonyait és ellátásait) amelyekhez hibás időtartam tartozott.

Azokra a megfigyelésekre (egy személy egy időtartamra vonatkozó státusza) vonatkozóan, amelyek 2003 előtt kezdődnek a következő szabályt alkalmaztuk:

- OEP ellátás, NYUFIG, NMH érintett – minden esetben úgy lett véve, mintha helyes lenne, és 2003.01.01-től lett figyelembe véve.

Azokra a megfigyelésekre (egy személy egy időtartamra vonatkozó státusza) vonatkozóan, amelyek 2011 után végződnek a következő szabályt alkalmaztuk:

- minden 2011 után végződő státuszról feltételezzük, hogy helyes, s úgy tekintjük őket, hogy 2011.12.31-én érvényben vannak.

### **Lezáratlan státuszok kezelése**

Azon jogviszonyok és ellátások kezelése, melyek esetében az időszak végét mutató érték hiányzik.

OEP jogviszony és ellátás:

- A jogviszony adatoknál nem lehetett megállapítani, hogy a vége dátum azért hiányzik, mert még jelenleg is tart az adott ellátás vagy adatrögzítési hiba. Ezért ezeket a státuszokat végigvezettük 2012. decemberig, de jelölve van, ha ezen ellátás lezáratlan. Az ellátás adatok esetén ezek olyan státuszok voltak, melyek nem sokkal 2012.12.31. előtt kezdődtek, s valószínűsíthető, hogy ezek még folyamatban voltak ekkor. Itt is a jogviszony fájlokhoz hasonlóan jártunk el, a státuszokat végigvezettük 2012. decemberéig, és jelöltük, hogy lezáratlan.

ONYF:

- A KELEN adatbázisban nincsenek missing végű jogviszonyok.
- A NYUFIG állományban nincsenek a nyugdíjknak és nyugdíjszerű ellátásoknak folyósítására vonatkozó hiányzó vége dátumok.

NMH:



- Nem lehetett megállapítani, hogy a vége dátum azért hiányzik, mert még jelenleg is tart az adott ellátás vagy adatrögzítési hiba. Ezért ezeket a státuszokat végigvezettük 2011. decemberig, de jelölve van, ha ezen ellátás lezáratlan.

## Összevont státuszkategóriák kialakítása

Az egyes állományokból származó ellátások és jogviszonyok alapján összevont státuszkategóriákat alakítottunk ki:

Foglalkoztatási viszony (fogvisz)	201 „munkaviszonnal foglalkoztatott”
	202 „köztisztviselő”
	203 „közalkalmazott”
	204 „bíró, ügyész, katona, hivatásos szolg. viszony”
	205 „polgármester, választott tisztségviselő, képviselő, egyházi személy”
	206 „szövetkezet tagja”
	207 „társas vállalkozó”
	208 „egyéni vállalkozó”
	209 „segítő családtag”
	210 „tanulás mellett foglalkoztatott”
	211 „bedolgozói, megbízási, vállalkozási jellegű jogviszony, felhasználási szerződés”
	212 „mezőgazdasági őstermelő”
	213 „alkalmi munkavállalói könyvvel foglalkoztatott”
	214 „egyéb foglalkozási viszony”
	215 „közfoglalkoztatott”
Transzfer státusz (transzfer)	301 „gyes”
	302 „gyed”
	303 „gyet”
	304 „tgyás”
	305 „öregségi nyugdíj”
	306 „rokkantsági nyugdíj”
	307 „baleseti rokkantsági nyugdíj”
	308 „egyéb nyugdíjas”
	309 „öregségi járadékok”
	310 „rokkantsági járadékok”
	311 „özvegyi járadék”
	312 „szociális járadékok”
	313 „egyéb járadékok”
	314 „ápolási díj”
	315 „hozzátartozói ellátások”
	316 „munkanélküli járadék és járadékszerű ellátások”
	317 „munkanélküliek szociális segélyei”
	320 „fogvatartott”
	321 „passzív táppénz”
	322 „egyéb nyugdíjszerű ellátás”
402 „lezáratlan: gyed”	
420 „lezáratlan: fogvatartott”	

## **Az államigazgatási adatgyűjtemény változói**

Az állományban megtalálható változókról lásd a változókatalógust.