

BOZA ISTVÁN–PETŐ RITA–TIR MELINDA

Magyar adminisztratív munkáltatói-munkavállalói adatok

Az Admin4 adatbázis hazai és nemzetközi perspektívában

Az adminisztratív munkáltató-munkavállaló adatbázisok (*Linked Employer-Employee Data, LEED*) az utóbbi évtizedek legfontosabb adatinnovációi közé tartoznak, mivel egyszerre teszik lehetővé az egyének és munkáltatóik hosszú távú vizsgálatát. Tanulmányunk áttekinti a hazai adatforrások fejlődését, majd részletesen bemutatja a legfrissebb magyar kapcsolt munkáltató-munkavállaló adatbázist, a Kapcsolt Államigazgatási Paneladatbázist (rövidítve: Admin4), külön kitérve a hozzá kapcsolható modulokra és kiegészítő információkra. Zárásként nemzetközi kitekintést adunk, amelyben összehasonlító perspektívában értékeljük a magyar adatok sajátosságait, előnyeit és korlátait.*

Journal of Economic Literature (JEL) kód: C81, O57.

Bevezető

Az elmúlt évtizedek adatforradalmának egyik legizgalmasabb fejleménye az adminisztratív munkáltató-munkavállaló adatok (LEED) megjelenése és elterjedése volt. Ma már, ha nemzetközi munkaerőpiaci témákkal foglalkozó konferencián járunk, azt tapasztaljuk, hogy az új kutatások nagyrészt nemcsak empirikusak, hanem egyre gyakrabban támaszkodnak kapcsolt munkáltató-munkavállalói panelekre vagy más adminisztratív forrásokból származó adatokra.

A kapcsolt munkáltató-munkavállaló adatok legfontosabb jellemzője, hogy egyszerre nyújtanak átfogó képet az egyének munkaerőpiaci helyzetéről és a munkáltatóik jellemzőiről. Ez lehetővé teszi a kutatók számára egyrészt a vállalatokon belüli

* A szerzők hálásak Antoine Bertheau, Bernardo Fanfani, Olivier Godechot, Martin Hallsten, Are Skeie Hermansen, Koren Miklós, Alena Křížková, Andrew Penner, Vladimír Peciar, Malte Reichelt, Vera Rocha, Halil Sabanci, Marta Silva, Beatrice Allamand Turner, Andrea Weber hozzájárulásaiért.

Boza István, ELTE Közgazdaság- és Regionális Tudományi Kutatóközpont (e-mail: boza.istvan@rtk.elte.hu).

Pető Rita, ELTE Közgazdaság- és Regionális Tudományi Kutatóközpont (e-mail: peto.rita@rtk.elte.hu).

Tir Melinda ELTE Közgazdaság- és Regionális Tudományi Kutatóközpont (e-mail: tir.melinda@rtk.elte.hu).

A kézirat első változata 2025. szeptember 11-én érkezett szerkesztőségünkbe.

DOI: <https://doi.org/10.18414/KSZ.2025.11.1109>

folyamatok, másrészt a cégek közötti különbségek vizsgálatát. Korábban a kutatók telephelyszintű kérdőíves felmérések segítségével is képesek voltak olyan elemzések összeállítására, amelyekhez egy adott cég összes munkavállalójának megfigyelése szükséges, és az sem számít újdonságnak, hogy longitudinális felmérések révén az egyéneket hosszú időn át követik. Az igazi áttörés azonban az, amikor mindkét dimenzió – az egyéni és a vállalati – több éven keresztül, összekapcsolva figyelhető meg. Az ilyen kapcsolt munkáltató-munkavállaló panelek (LEEP) új kutatási lehetőségeket teremtenek: feltárhatóvá válnak például a vállalatok közötti mobilitási minták, a cégekben belüli változások (például a tulajdonosváltás), valamint ezek az adatok lehetővé teszik a szakpolitikai döntések következményeinek átfogó értékelését is.

Tanulmányunkban a hazai kapcsolt munkáltató-munkavállaló paneladatok elérhetőségének összefoglalása után a legújabb ilyen paneladatbázist, az Admin4-et mutatjuk be. Ezen adatbázis tartalmi leírása – amely az új és a visszatérő felhasználóknak egyaránt érdekes lehet – mellett átfogó képet adunk arról is, hogy a világ más részein hol érhetőek el még hasonló adatbázisok, és miben hasonlítanak, illetve miben térnek el a magyar adatoktól.

Kapcsolt munkáltatói-munkavállalói adatok forrásai és jellemzői

A kapcsolt munkáltató-munkavállaló adatbázisok két leggyakoribb forrása a kérdőíves adatfelvétel és az adminisztratív adatforrás. A fejlődést időrendben áttekintve, először az 1960-as években jelentek meg a munkáltatók által kitöltött bérstruktúra-felmérések – ezek közül az egyik legkorábbi példa a francia Enquête sur la Structure des Salaires (ESS, 1966), amelyet a hetvenes évek végére számos hasonló adatfelvétel követett Európában (Abowd–Kramarz [1999]).

A kérdőíves adatfelvétel legnagyobb előnye a tartalmi rugalmasság: a kérdések köre az aktuális kutatói és szakpolitikai igényekhez igazítható. Így ezek az adatbázisok nemcsak a szokásos munkaerőpiaci mutatókat (például a ledolgozott munkaórákat, a túlórák számát, a bérelemeket) rögzítik, hanem a munkavállalók háttérére (például képzettségére, készségeire és az általuk használt eszközökre) vonatkozó részletes információkat is. A rugalmas tartalom lehetővé teszi, hogy a döntéshozók és a kutatók mélyebb betekintést nyerjenek az aktuális gazdasági tendenciák mögött meghúzódó folyamatokba.

Annak érdekében, hogy a létrejövő adatbázisban egyidejűleg lehessen látni az egyéni és a céges adatokat, az adatokat a munkáltatók szolgáltatják. Ez azonban előrevetíti az ilyen adatbázistípus legnagyobb hátrányát is. A mintavétel alapjául az alkalmazottak köre szolgál, és nem a teljes népesség. E miatt az eljárás miatt kimaradnak a munkanélküliek, az inaktívak, valamint a nem alkalmazotti viszonyban dolgozók (például az egyéni vállalkozók). Ráadásul a felmérés lebonyolításának magas költségei miatt a legtöbb esetben a minta alapjául szolgáló alapsokaságnak csak egy meghatározott részét kérdezik le.

E korlátok egy részére megoldást kínálnak az adminisztratív adatforrásokból épített kapcsolt munkáltató-munkavállaló adatbázisok. Ezek jellemzően egy vagy több

közigazgatási szerv – például adó- vagy nyugdíjhatóságok, központi bankok – különböző nyilvántartásainak rendszerezésével és összekapcsolásával jönnek létre. Előnyeik éppen a keletkezésük mikéntjéből fakadnak. Költséghatékonyak abból a szempontból, hogy míg egy kérdőíves felmérés lebonyolítása jelentős összegeket emészt fel, addig az adminisztratív forrásokból építkező adatbázisok valamely állami szerv által a saját feladatainak elvégzése céljából gyűjtött adatokból épülnek fel. Az adminisztratív adatok esetén azonban az adatok folyamatos frissítése is kevésbé költséges, mint egy újabb kérdőíves hullám lekérdezése. Az adatbázisok gondozása és tárolása természetesen mindkét esetben további költségekkel járhat.

A kutatók számára további előnyt jelenthet, hogy a több közigazgatási nyilvántartás összekapcsolása lehetővé teszi a teljes népesség (foglalkoztatottak, inaktívak, munkanélküliek) nyomon követését és viselkedésük vizsgálatát, szemben a kérdőíves felmérésekkel, amelyek jellemzően a népesség egy jól körülhatárolható szeletére koncentrálnak (például a foglalkoztatottakra). Ráadásul az adminisztratív adatokból származó információkat megbízhatóbbnak tekintjük.

Ugyanakkor az a hátrányuk, hogy csak az állami szervek feladatainak ellátásához szükséges információkat rögzítik. A jövedelmi adatok rendszerint például csak a közigazgatás számára releváns jövedelmeket tartalmazzák. Az adóhatóságot például csak az adózás vagy a társadalombiztosítási járulékok alá tartozó teljes munkajövedelem érdekli, és nem tesz különbséget a rendszeres vagy a túlórából származó jövedelem között. A jövedelem részletes megbontásának hiánya egy fontos korlátja ezeknek az adatbázisoknak, hiszen így például a túlórák vagy a teljesítményhez kapcsolódó bónuszok elemzésének lehetősége elvész. Emellett a munkakörülményekre, a családi háttérre, a szociális jellemzőkre vagy a részletes iskolázottságra vonatkozó információk is hiányoznak, és az informális foglalkoztatás sem jelenik meg ezekben az adatbázisokban.

Magyarországon kérdőíves forrásból származó kapcsolt munkavállalói-munkáltatói adatbázis a Bértarifa-felmérés, amelynek első hulláma 1986-ban zajlott, amit 1989-ben megismételtek. 1992-től kezdődően viszont évente elvégzik a felmérést.¹ A költségvetési szektorban a felvétel minden intézményre kiterjed. A vállalati szektorban a vizsgálat a nagyobb méretű vállalatok teljes körére – azon belül viszont a dolgozók véletlen mintájára –, a kisebbek esetében pedig egy véletlen mintára terjedt ki (viszont minden munkavállalóra vonatkozóan). Az 1995-ig tartó időszakban a kritikus mérethatár alatti vállalatoknál, valamint a költségvetési szektorban a felvétel minden dolgozóra kiterjedt. 1995-től kezdődően azonban a mérethatár feletti cégeknél és a költségvetési intézményeknél már csak a dolgozók véletlenszerűen kiválasztott mintájáról gyűjtenek egyéni adatokat.

A fentiekkel szemben adminisztratív forrásokból építkezik a KRTK Adatbankja (későbbiekben: Adatbank) által gondozott Kapcsolt Államigazgatási Paneladatbázis (Admin), amelynek több hulláma is elérhető már. Míg az első hullám adatbázisa (Admin1) csak a munkavállalói oldalt tartalmazta, és a cégeket nem, addig a 2015-ben létrejövő következő hullám (Admin2) már teljes értékű kapcsolt munkáltató-munkavállaló panelnek tekinthető. Az ezt követő hullám (Admin3) – amely 2020-ban

¹ A korábbi években április–május hónapokban, majd az elmúlt években október folyamán.

jött létre – újítása, hogy egyrészt hosszabb időhorizontot ölelt fel (2003–2017), másrészt a kutatható témakörök tárházát is jelentősen bővítette. Az adatbázisokat már közel 100 publikáció használta fel – ennek több mint fele folyóiratban –, és az Admin3 esetében jelenleg is több mint 160 folyamatban lévő kutatást tart számon az Adatbank. A folytatásban a legfrissebb hullám adatbázisát, az Admin4-et mutatjuk be. Az adatbázis technikai létrehozásának hátterét *Sebők* [2022] tárgyalja. Mi igyekszünk a kutatói szempontból releváns tartalmi elemeket számba venni és így bemutatni az adatban rejlő lehetőségek széles tárházát.

A Kapcsolt Államigazgatási Paneladatbázis (Admin4)

Az Adatbank munkája nyomán negyedik alkalommal jön létre a Kapcsolt Államigazgatási Paneladatbázis. Az adatbázis legújabb verziója (Admin4) az előző hullámokhoz képest kibővült mind az adatforrások, mind az időtáv tekintetében, és immár a 2003–2021 közötti éveket fedi le.²

Az adatbázis a Nemzeti Egészségbiztosítási Alapkezelő (NEAK), a Magyar Államkincstár (MÁK), az Oktatási Hivatal (OH), a Nemzeti Szakképzési és Felnőttképzési Hivatal (NSZFH), az akkori Technológiai és Ipari Minisztérium (TIM) és a Nemzeti Adó- és Vámhivatal (NAV) egyéni és vállalati szintű adatainak összekötésével jön létre. A minta alapsokaságát és így a kapcsolás alapját a NEAK adja a TAJ-nyilvántartás egy 50 százalékos egyéni valószínűségi mintájának kiválasztásával. Az adatkapcsolás így a teljes magyar lakosságra reprezentatív. A minta a belépések (születések, bevándorlás) és a kilépések (halálozás, TAJ-törlés) folyamatos követésével frissül.

Az anonimizált rekordok – szigorú adatvédelmi eljárásokat figyelembe véve – egy központi adatbázisba kerülnek. Az adatfeldolgozást a Nemzeti Infokommunikációs Szolgáltató Zrt. (NISZ) végzi, mielőtt az adatok az Adatbankba kerülnek, ahol az adatokat tisztítják, és havi bontású formátumba alakítják a kutatók számára. A teljes adatbázis legfeljebb 228 hónapon keresztül követi az egyéneket. Az egyének mellett minden foglalkoztató – a verseny- és a közsférából egyaránt –, aki legalább egy embert alkalmazott a mintában szereplők közül, a teljes történetével megjelenik az adatbázisban (kb. 1,7 millió vállalat és intézmény).

Szerkezetét tekintve az Admin4 egy alapfájlból és több kiegészítő modulból áll. Míg az alapfájl elsősorban a foglalkoztatással kapcsolatos alapvető információkat (például a foglalkoztatási és béradatokat) tartalmazza, addig a kiegészítő fájlok részletes egyéni (vagy vállalati) szintű adatokat tartalmaznak az egészségügy, az oktatás és a munkanélküliség területéről, valamint a gazdasági társaságok pénzügyi adatairól is. Az információk egyéni azonosítók segítségével kapcsolhatók össze. Ez a szerkezet lehetővé teszi, hogy a felhasználók maguk alakítsák ki azoknak az információknak a körét, amelyekre ténylegesen szükségük van az adott kutatási kérdés megválaszolásához. Fontos azonban megjegyeznünk, hogy a legtöbb kiegészítő modul csak 2009 utáni adatokat tartalmaz.

² Az információknak egy szűk része elérhető a 2022-es évre is. Erről később részletesen írunk.

Az információk érzékenysége alapján az elérhető adatokat az Adatbank három csoportba sorolja, ez látható az 1. táblázatban. Míg a legalacsonyabb szintre (H1) sorolt adatok az Adatbank szerverén keresztül érhetők el engedélyeztetés után, addig a legmagasabb szintre sorolt adatok (H3) egy külön erre a célra létesített, zárt adatszobai környezetben kutathatók csak.

1. táblázat

A Kapcsolt Államigazgatási Paneladatbázis (Admin4) hozzáférési szintjei

Szint	Adatkörök	Hozzáférési jogosultak	Hozzáférés módja
H1	Törzsadatok, egyszerű kiegészítő modulok	kutatói célú jelentkezők, adatbanki engedélyezés után	KRTK-adatbanki szerver
H2	Részletesebb egészségügyi és oktatási adatok	KRTK-s kutatók, a terület szakértői, engedélyezés után	KRTK-adatbanki szerver
H3	Teljes részletességű kiegészítő modulok (például BNO-, OENO-információk)	KRTK-s kutatók, a terület szakértői, részletes projektterv engedélyezése után	KRTK-adatszoba

Törzsadatok

Az adatbázis a teljes lakosság egy reprezentatív (50 százalékos) mintájára vonatkozóan tartalmaz információkat, így a munkaerőpiacon aktív szereplők mellett megjelennek az inaktívak (többek között tanulók, háztartásbeliek) és a munkanélküliek is. Minden esetben ismert az egyén kora és neme. Az egyes kiegészítő modulok egy része pedig a nem aktív lakossághoz is kapcsolható.

MUNKAVISZONY ÉS JELLEMZŐI • A rendelkezésre álló információk alapját a nyugdíj jogosultságot megalapozó jogviszonyok adják, amelyek jellemzően a munkáltató által benyújtott bevételekből származnak. Mind a közsféra, mind a magánszektor munkáltatói által foglalkoztatottak megfigyelhetők az adatbázisban, valamint látható a foglalkoztatási viszonyuk is (például munkaviszony, közfoglalkoztatás, egyszerűsített foglalkoztatás stb.). Minden munkáltató egyedi, anonim azonosítót kap (beleértve az egyéni vállalkozásokat is), ami lehetővé teszi a munkáltatók és a hozzájuk kapcsolódó munkavállalók időbeli követését.

A főbb foglalkoztatási jellemzők között szerepel a szerződés szerinti heti munkaidő, a havi biztosított munkanapok száma, valamint minden nyugdíj jogosultságot eredményező kereset. A munkavállalók foglalkozásait a Foglalkozások Egységes Osztályozási Rendszere (FEOR) szerint rögzítik, amely az Európai Unióban használt ISCO rendszer magyar megfelelője. Így lehetővé válik a munkakörök pontos megkülönböztetése egy adott munkáltató esetén. Az egyidejűleg több munkahelyen dolgozó személyek összes munkaviszonya nyomon követhető az adatbázisban. Bármely időpillanatban az összes munkavállaló körülbelül 10 százaléka rendelkezik párhuzamosan több munkahellyel.

Az adatbázis szerkezeti sajátosságainak vannak azonban korlátai is. Ilyen például, hogy az időskorú foglalkoztatás csak azon időszakokban jelenik meg az adatbázisban, amikor az járulékköteles volt. A foglalkoztatott nyugdíjasok (illetve azok munkáltatói) azonban mentesültek a járulékfizetés alól 2007 előtt és 2019 után, így ezen időszakokban nem látjuk a nyugdíj mellett dolgozók munkaerőpiaci információit (munkáltatójuk, keresetük, foglalkozásuk). 2012 előtt pedig a fegyveres testületek tagjai (rendőrség és katonaság) külön nyilvántartásban szerepeltek, így az ő foglalkoztatásuk sem szerepel az adatbázisban. Végül, ahogy az adminisztratív adatbázisoknál elkerülhetetlen, a be nem jelentett „feketefoglalkoztatás” sem követhető nyomon.

A kereseti adatok mind az alkalmazottak, mind az egyéni vállalkozók esetében elérhetők, de kizárólag a nyugdíjjárulék-alapot képező jövedelem kerül rögzítésre. Emiatt egyrészt a vállalkozók tényleges keresetét az adatbázisban szereplő információk alulbecsülik. Másrészt, mivel a nyugdíjjárulék alapját 2012-ig felső plafon korlátozta, a rögzített jövedelmek az alkalmazottak esetében is cenzorálva vannak ezen a felső korláton. Az érintett években a jövedelemeloszlásnak körülbelül a felső 2 százalékát érintheti ez a probléma. A legfelső kereseti kategóriákban dolgozók így ugyan megfigyelhetők maradnak, de pontos bérszintjük nem ismert.

2012 előtt a legtöbb munkáltató éves bontásban nyújtott be járulékbevallásokat, míg a havi adatszolgáltatás 2013-tól vált kötelezővé. Ennek következtében a korábbi időszakokban nem lehet pontosan követni az évközi bértváltozásokat. Az időszak első felében az Adatbank osztotta el a jövedelmeket a releváns hónapokra. Magyarországon, főleg az érintett időszak elején, a „szürkefoglalkoztatás” is elterjedt gyakorlat volt, amely során a munkáltatók a dolgozóikat hivatalosan minimálbéren jelentették be, miközben a fennmaradó bért készpénzben fizették ki. Mivel ezek után nem fizettek nyugdíjjárulékot, a kereseti adatok sokszor alulbecsültek lehetnek – különösen bizonyos ágazatokban (például építőipar, kereskedelem), illetve mikrovállalatok körében (*Benedek és szerzőtársai [2012]*).

Bár a teljes adatbázis a 2003 és 2021 közötti időszakra érhető el, a munkaviszonyra vonatkozó adatok a 2022-es évre is a kutatók rendelkezésére állnak. Fontos azonban megjegyezni, hogy ez azt is jelenti, hogy a 2022-es évre vonatkozó mintában csak az ebben az évben a munkaerőpiacon aktívak szerepelnek, más információ nem áll rendelkezésre az inaktívokról, és nem látjuk az ebben az évben belépő egyéneket sem.

A TELJES POPULÁCIÓRA VONATKOZÓ INFORMÁCIÓK • Az adatbázis nemcsak a foglalkoztatottak, hanem a teljes lakosság egy reprezentatív mintájára vonatkozóan is tartalmaz információkat. Egyrészt 2003-tól kezdődően a teljes lakosság esetében látjuk, hogy valamilyen formában dolgozik-e, vagy sem. Másrészt a regisztrált munkanélküliekhez kapcsolódó adatok 2009-től kezdődően érhetőek el, ideértve a regisztráció, járadékfolyósítás, programrészvétel, közvetítés, valamint a közfoglalkoztatási részvétel információit (utóbbit 2011-től).³ A koronavírus-járvány idején nyújtott bértámogatási programok is szerepelnek az adatbázisban.

³ Fontos azonban, hogy csupán a munkaügyi hivataloknál regisztrált álláskeresőket látjuk, ami a teljes munkanélküli-populáció egy (szelektált) almintája. A nem regisztráló, de ettől függetlenül állást kereső munkanélküliek és a ténylegesen inaktívak megkülönböztetésére az adminisztratív adatok sajnos nem nyújtanak lehetőséget.

A nyugdíj- és nyugdíjszerű ellátásokkal kapcsolatos adatok a teljes időszakra elérhetők. Az ellátástípusok csoportosítva szerepelnek, saját jogú nyugdíj esetén az elismert szolgálati idő is ismert. A pénzbeli szociális transzferek közül 2012-től szerepelnek az olyan ellátások, mint a csed, a tgyás/gyed vagy a táppénz. Korábbi időszakokra ezek csak részlegesen vagy státusszinten érhetők el. A havi szintű adatfeldolgozás lehetővé teszi ezen ellátások időbeli alakulásának részletes követését.

FÖLDRAJZI DIMENZIÓ • A törvényi szabályozás értelmében az adatbázis legrészletesebb térbeli felbontása a járási szint. Egyéni szinten ismert a mintába került személyek adott hónapban érvényes állandó és ideiglenes lakcíméhez tartozó járása. A születési hely magyar állampolgárok esetében szintén járási szinten, külföldi születés esetén aggregált országcsoportok szerint áll rendelkezésre.

Kiegészítő modulok

MUNKÁLTATÓI INFORMÁCIÓK • A munkáltatók azonosítása egyedi, anonimizált azonosítók révén történik. A vállalati szintű információk közül ismert a TEÁOR-besorolás, a Bértarifa-felvételből származó egyes jellemzők (2018-ig), illetve a mérlegbevallásra kötelezett cégekre vonatkozóan a NAV-mérlegbevallási adatok, ideértve az eredménykimutatások fő sorait, mint például árbevétel, anyagköltség, mérlegfőösszeg, tőkeösszetétel és létszámadatok. Az információk éves bontásban érhetők el. A TEÁOR-besorolás és a mérlegadatokat teljes körűen csak a 2011 és 2020 közötti időszakra kapcsolhatók közvetlenül, míg a korábbi évekre vonatkozóan az Adatbank valószínűségi illesztési eljárást alkalmazott. A 2021 és 2022 közötti adatok az adatbázis jelen verziójában még nem szerepelnek.

RÉSZLETES TRANSZFERADATOK • A 2003–2021 közötti időszakra részletes nyugdíj-adatok állnak a kutatók rendelkezésére havi bontásban. Az adatbázis tartalmazza a folyósított nyugdíjak, illetve nyugdíjszerű ellátások és juttatások megállapítási idejét, az adott ellátáshoz tartozó megállapított szolgálati időt, a havi rendszeres folyósítási összegét és az eseti kifizetéseket is. A feldolgozás során az ellátásokat kategóriánként összevonták, a legfontosabb nyugdíjtípusok azonban külön-külön is vizsgálhatók (öregségi nyugdíj, korhatár előtti nyugdíj, rokkantsági nyugdíj).

EGÉSZSÉGÜGY • Az egészségügyi adatok a közfinanszírozott ellátásokra vonatkoznak, és 2009-től állnak rendelkezésre.⁴ A főbb ellátástípusok: fekvő- és járóbeteg-ellátás, gyógyszerkiváltás, valamint diagnosztikai vizsgálatok (például MRI, CT). Az ellátásokat a BNO- (Betegségek Nemzetközi Osztályozása) és az OENO- (Orvosi Eljárások Nemzeti Osztályozása) kódok azonosítják, az adatok

⁴ A háziorvosi ellátásra, fekvő- és járóbeteg-ellátására, valamint a vénnyírástokra és CT/MRI vizsgálatokra vonatkozó információk 2009-től érhetők el. A közgyógyellátási adatok azonban már 2005-től, míg a fogászati ellátásra vonatkozó információk csak 2016-tól állnak a kutatók rendelkezésére.

eseményszinten követhetők, ismertek az adott ellátáshoz kapcsolódó elszámolt finanszírozási költségek. Ezen kódok alapján aggregált mutatók is elérhetők az egészségi állapot jellemzésére.

OKTATÁS • Az oktatási adatok elsősorban 2009-től állnak teljes körűen rendelkezésre, de egyes esetekben korábbi évekre is visszavezethetők. Az adatbázis tartalmazza a közoktatásban való részvételt, az érettségi vizsgák eredményeit (2005-től), valamint az országos kompetenciamérés adatait (2008-tól), beleértve a teszteredményeket és a háttérkérdőívben szereplő információkat is. A felsőoktatási adatok elérhetősége korlátozott, de az intézményi részvétel és a végzettség bizonyos esetekben ismert. Az iskolai végzettségre vonatkozó információ az 1991 után születetteknek áll rendelkezésre, részben visszavezetett módon.

CSALÁD • A mintába került egyének egy részére elérhetők a családi kapcsolatokra vonatkozó információk is. A szülő-gyermek kapcsolatok azonosítása a családipótlék-igénylések és -folyósítások alapján történik, és akkor áll rendelkezésre, ha mind a szülő, mind a gyermek bekerült az 50 százalékos mintába. A kapcsolatoknak tehát körülbelül az egynegyede rekonstruálható. Szintén a családi pótlék adatain alapul a gyermekszám meghatározása. Ez az információ azokban az esetekben érhető el, amikor a mintába került személy gyermeke után családi pótlékot igényelt vagy kapott. Az igénylésekre és folyósításokra vonatkozó információk 2008-tól érhetőek el.

Nemzetközi összehasonlítás

A kapcsolt munkáltató-munkavállaló adatok (LEE-adatok) az elmúlt két évtizedben világszerte a munkaerőpiaci kutatások egyik központi adatforrásává váltak. Ma már több mint harminc országban érhető el ilyen típusú adatbázisok, és egyre több országban épülnek adminisztratív nyilvántartásokból származó panelek. Egy nemzetközi felmérés segítségével 34 ország – ebből 21 európai – kapcsolt munkáltató-munkavállalói adatairól gyűjtöttünk információt (*Boza-Pető* [2025]). Ezeket a 2. és 3. táblázat foglalja össze.

Az Admin4 egyike azon adatbázisoknak, amelyek egy véletlenszerűen kiválasztott egyéni alapú népességmintára épülnek, így nem csupán a foglalkoztatottakat vagy a magánszektor fedik le. Ez különösen fontos abból a szempontból, hogy nemcsak a munkapiac aktív szereplőiről, hanem az inaktívokról, munkanélküliekről és egyéni vállalkozókról is rendelkezésre állnak adatok, lehetővé téve például a munkaerőpiaci belépések és kilépések pontos követését vagy a foglalkoztatási státuszok közötti átmenetek vizsgálatát. Egyes kutatási témák tekintetében azonban továbbra is hátrányt jelent, hogy csupán egy 50 százalékos mintán dolgozhatnak a kutatók – még ha egyes országokban ennél szigorúbb korlátozásokkal is találkozhatunk.

Az időbeli felbontás szempontjából is versenyképes a magyar adatbázis: míg sok nemzetközi adat csupán évente rögzíti a munkaviszonyokat, az Admin4 havi vagy akár folytonos (*spell*) szintű adatai lehetővé teszik a munkaerőpiaci mozgások finom

2. táblázat

Munkáltató-munkavállalói paneladatok elérhetősége – Európa

Ország	Alapsokaság ^a	Mintavétel ^b	Felbontás ^c	Időablak ^d
Ausztria	foglalkoztatottak	teljes	folytonos	1972–2020
Belgium	foglalkoztatottak	minta (10%)	negyedéves	2003–2022
Dánia	minden személy	teljes	éves	1980–2024
Egyesült Királyság	foglalkoztatottak	minta (1%)	éves	1975–2019
Észtország	foglalkoztatottak	teljes	éves	2001–2017
Finnország	foglalkoztatottak	teljes	éves	2004–2018
Franciaország	foglalkoztatottak	teljes	éves	1994–2020
Görögország	magánszféra	teljes	havi	2002–2023
Hollandia	foglalkoztatottak	teljes	havi	2006–2022
Írország	foglalkoztatottak	minta (10%)	éves	2011–2022
Lengyelország	foglalkoztatottak	teljes	éves	2000–2019
Litvánia	minden személy	minta (25%)	havi	2000–2020
Magyarország	minden személy	minta (50%)	folytonos	2003–2022
Németország	minden személy	teljes	folytonos	1993–2019
Norvégia	foglalkoztatottak	teljes	havi	1995–2021
Olaszország	magánszféra	teljes	havi	1974–2021
Portugália	magánszféra	teljes	éves	1985–2019
Spanyolország	minden személy	minta (5%)	éves	2006–2018
Svédország	minden személy	teljes	éves	1990–2021
Szlovákia	foglalkoztatottak	teljes	havi	2014–2021
Szlovénia	foglalkoztatottak	teljes	éves	1992–2014

^a A „foglalkoztatottak” megjelölés a hivatalos foglalkoztatásra utal, nem tartalmazva az informális vagy illegális foglalkoztatási formákat. Az országspecifikus, kisebb kivételek (például kis- vagy nagyvállalatok, bizonyos ágazatok vagy foglalkozások stb.) itt nincsenek feltüntetve.

^b A mintavétel általában egyszerű véletlen mintavétel az egyének szintjén. A zárójelben szereplő szám a kutatók számára elérhető adatok arányát jelzi az összes elérhető vagy mintázott populáció százalékában.

^c A legkisebb elérhető felbontás. Előfordul, hogy ez a felbontás csak az adatok egy részmin-táján áll rendelkezésre, nem az egész időszakra vonatkozóan. A folytonos olyan országokat jelöl, amelyek rendelkeznek részletes időszakszintű (lényegében napi) adatokkal.

^d Az utolsó évek az utolsó, valamilyen publikációban használt évre vagy a legutolsó olyan évre vonatkoznak, amelyet egy, az adatokhoz hozzáférő kutató megerősített.

időzítésének elemzését. Ez a részletesség különösen fontos a rövid távú változások – például jogszabályváltozások – hatásának vizsgálatánál, ahol az éves adatok már túl durva felbontást jelentenének. Míg korábban jelentős korlát volt az adatbázis rövidsége, az Admin4 már a maga közel húszéves időtartamával az elérhető adatbázisok középmezőnyébe tartozik.

3. táblázat

Munkáltató-munkavállalói paneladatok elérhetősége – az amerikai kontinens és egyéb országok

Ország	Alapsokaság ^a	Mintavétel ^b	Felbontás ^c	Időablak ^d
Argentína	magánszféra	minta (3%)	havi	1996–2015
Ausztrália	foglalkoztatottak	teljes	folytonos	2011–2020
Brazília	foglalkoztatottak	teljes	éves	1985–2018
Chile	foglalkoztatottak	teljes	havi	2002–2022
Costa Rica	foglalkoztatottak	teljes	havi	2006–2017
Ecuador	foglalkoztatottak	teljes	havi	2006–2012
Egyesült Államok	foglalkoztatottak	teljes	éves	1993–2021
Izrael	foglalkoztatottak	teljes	havi	1983–2015
Kanada	foglalkoztatottak	teljes	éves	1983–2019
Kolumbia	foglalkoztatottak	teljes	havi	2008–2014
Mexikó	magánszféra	teljes	havi	2005–2019
Tajvan	magánszféra	teljes	havi	1998–2004
Új-Zéland	foglalkoztatottak	teljes	havi	2000–2017

^a A „foglalkoztatottak” megjelölés a hivatalos foglalkoztatásra utal, nem tartalmazva az informális vagy illegális foglalkoztatási formákat. Az országspecifikus, kisebb kivételek (például kis- vagy nagyvállalatok, bizonyos ágazatok vagy foglalkozások stb.) itt nincsenek feltüntetve.

^b A mintavétel általában egyszerű véletlen mintavétel az egyének szintjén. A zárójelben szereplő szám a kutatók számára elérhető adatok arányát jelzi az összes elérhető vagy mintázott populáció százalékában.

^c A legkisebb elérhető felbontás. Előfordul, hogy ez a felbontás csak az adatok egy részmintáján áll rendelkezésre, nem az egész időszakra vonatkozóan. A folytonos olyan országokat jelöl, amelyek rendelkeznek részletes időszakszintű (lényegében napi) adatokkal.

^d Az utolsó évek az utolsó, valamilyen publikációban használt évre vagy a legutolsó olyan évre vonatkoznak, amelyet egy, az adatokhoz hozzáférő kutató megerősített.

Természetesen vannak példák, főleg a skandináv országokban, ahol még ennél is több évtizedre is vissza tudnak nyúlni a kutatók.

Tartalmi szempontból az Admin4 nemzetközi viszonylatban a kifejezetten gazdag adatbázisok közé tartozik. A rendelkezésre álló modulok lefedik az egészségügyi, oktatási, pénzügyi, szociális és demográfiai információk széles körét. Különösen ritka, hogy egy adatbázis egyszerre tartalmaz például egészségügyi ellátási eseményeket, részletes oktatási karrierutakat és szülő-gyermek kapcsolatokat. A 4. táblázatban a Boza–Pető [2025] által készített felmérésre támaszkodva részletesen is bemutatjuk az egyes témakörök kapcsán elérhető információkat.

A 4. táblázat alapján jól látszik, hogy bár egyre több ország rendelkezik adminisztratív adatokból épülő, kapcsolt munkáltató-munkavállalói adatbázissal, a törzsadatokhoz kapcsolódó kiegészítő információk lefedettsége korántsem

4. táblázat

Nemzetközi munkáltató-munkavállalói paneladatokban elérhető információk

	Admin3	Admin4	Van adat	Nincs adat
Munkakörülmények (munkaórák, túlóra, bónuszok)	×	×	10	5
Cég pénzügyi/tulajdoni viszonyai, adatai	✓	✓	11	4
Szociális juttatások/transzferek	✓	✓	10	5
A munkahely elhelyezkedése	×	×	12	3
Lakóhely	✓	✓	12	3
Származási hely	×	✓	12	3
Egészségügyi adatok	✓	✓	7	8
Oktatási intézmények	✓	✓	8	7
Kognitív készségek	✓	✓	6	9
Gyermekek száma	×	✓	9	6
Családi kapcsolatok	×	✓	7	8
Összekapcsolható más adatokkal	×	×	11	4

Megjegyzés: a cellák azt mutatják, hogy hány országban érhető el legalább valamilyen részletességben adat az adott témakörre vonatkozóan. Részletes információkkal jelenleg 15 ország esetén rendelkezünk. Az egyes kategóriákban az országonkénti elérhetőségre vonatkozó információk megtalálhatók *Boza–Pető* [2025]-ben.

általános. Míg a magyarországi adatbázisban a törzsadatok kiegészíthetők a cégek pénzügyi adataival, valamint egyéni szinten például szociális transzferekkel és földrajzi információkkal, addig ezek az információk a vizsgált országok mintegy egynegyedében nem állnak rendelkezésre. Különösen ritkák az olyan érzékenyebb adatok, mint az egyéni szintű egészségügyi ellátások igénybevétele, az oktatási, a kognitív készségekre vagy a családi kapcsolatokra vonatkozó információk. Míg ezek Magyarországon elérhetők az adatbázisban, a lekérdezett országok körülbelül felében hiányoznak.

Az Admin4 egyik legnagyobb nemzetközi előnye, hogy a kutatható témák köre rendkívül széles a részletes kiegészítő moduloknak köszönhetően. Fontos azonban megjegyezni, hogy néhány lényeges területen továbbra is vannak hiányosságok. Az Admin4 egyik újdonsága a korábbi hullámokhoz képest, hogy már tartalmaz információt a családi kapcsolatokról is. Ezek azonban jelenleg korlátozottak: elsősorban az anya-gyermek kapcsolatok azonosíthatók, és a mintavételi eljárás (50 százalékos véletlen minta) miatt az összes lehetséges kapcsolatnak csak egy töredéke jelenik meg. Ahhoz ugyanis, hogy egy gyermeket az édesanyjához lehessen kapcsolni, mindkettjüknek szerepelnie kell a mintában.

Az Admin4 egy másik hiányossága, hogy az iskolai végzettségre vonatkozó információk közvetlenül csak a fiatalabb korosztály esetében állnak rendelkezésre, azon belül is azoknál, akik a megfigyelési időszak alatt aktív tanulói jogviszonnyal

rendelkeztek. Bizonyos esetekben azonban ez a hiányosság megkerülhető: az iskolai részvételre vonatkozó adatok alapján a végzettség becsülhető (*Köllő* [2022]). Hátrány továbbá, hogy a munkavállalókat nem telephely, hanem csak a foglalkoztató szintjén látjuk. Ily módon sem a munkavégzés pontos földrajzi helye nem ismert, sem az, hogy kik dolgoznak egymással közvetlenül egy munkahelyen.

Összességében elmondható, hogy a hazai Admin4 a kapcsolt munkáltató-munkavállaló adatbázisok nemzetközi mezőnyének élvonalába tartozik. Bár történetileg fiatalabb, mint például az osztrák vagy dán adatbázisok, lefedettsége, tartalmi gazdagsága és finom időbeli felbontása alapján számos szempontból kimagasló. A magyar adatbázis így nemcsak hazai elemzésekre nyújt kivételes lehetőséget, hanem a nemzetközi összehasonlító kutatások új hullámának is szerves részévé válhat.

A legfrissebb kutatási irányok és jövőkép

A tanulmány zárófejezete röviden bemutatja a legújabb innovációkat, valamint a több országot átfogó adatbázisokra épülő kutatások fejlődési irányait. Részletesebb elemzésért lásd *Boza–Pető* [2025].

Nemzetközi összefogások. Míg a legtöbb kapcsolt munkáltató-munkavállaló adatot használó elemzés egyetlen ország adataira épül, az egyre elterjedtebb nemzetközi összehasonlító vizsgálatok átfogóbb bizonyítékokat szolgáltatnak a modern gazdaságok közös viselkedési mintáiról és piaci folyamatairól. Az országokon átívelő elemzések megvalósítását ugyanakkor jelentősen megnehezítik az adathozzáférési korlátozások, valamint az adatok felhasználásához szükséges ország- és adatspecifikus szakértelem szükségessége. Egyéni kutatók számára emiatt ritkán reális cél az ilyen jellegű vizsgálatok létrehozása. A megoldást az országok közötti, sokszervező együttműködések jelentik, amelyek közös, harmonizált módszertanon alapuló, összehasonlító elemzéseket végeznek. Három jelentős nemzetközi kezdeményezést szeretnénk kiemelni ezen a téren.

A Comparative Organizational Inequality Network (COIN) egy 2015-ben indult szerveződés, amely mára 16 ország kutatóit fogja össze azzal a céllal, hogy a cégek szerepét vizsgálják a jövedelmi egyenlőtlenségek kialakulásában (*Tomaskovic-Devey és szerzőtársai* [2020], *Penner és szerzőtársai* [2022]). Az OECD LinkEED kutatócsoport 20 ország harmonizált LEE-adatain alapulva elemzi a vállalati bérpolitikák, a munkahelyváltás és a bérkülönbségek szerepét a munkaerőpiaci egyenlőtlenségek nemzetközi alakulásában (*OECD* [2021]). A Global Repository of Income Dynamics (GRID) projekt nyílt hozzáférésű, nemzetközi adatbázist hoz létre az egyéni szintű jövedelmi egyenlőtlenségek és dinamikák összehasonlítható mérésére. Jelenleg 13, a közeljövőben pedig több mint 25 ország adminisztratív adatainak felhasználásával (*Güvenen és szerzőtársai* [2022]). Ezek az együttműködések jelentősen hozzájárulnak az általános gazdasági és társadalmi folyamatok mélyebb megértéséhez.

Adatkapcsolások. Az adminisztratív adatbázisok egyik fő korlátja, hogy kizárólag az állami intézmények által gyűjtött információkat tartalmazzák. Ezt a hiányosságot

azonban ki lehet küszöbölni más adatforrások – például bérfelmérések, háztartási költségvetési adatok vagy időmérleg-vizsgálatok – integrálásával. Ilyen összekapcsolások révén a kapcsolt munkáltatói-munkavállalói adatbázisok tartalma jelentősen gazdagítható, amennyiben az adattulajdonosok engedélye és az adatvédelem (különösen az anonimizálás) megfelelően biztosított. Számos országban már folynak ilyen munkák, és egyre több helyen integrálnak új típusú adatokat, például oktatási, egészségügyi vagy vállalati hálózati információkat.

A német IAB (*Institute für Arbeitsmarkt- und Berufsforschung*) például élen jár ezen a téren: olyan adatbázisokat fejleszt, amelyekben az adminisztratív nyilvántartások mellett megjelennek a cégek munkaerő-felvételi gyakorlatairól, önbevalláson alapuló bérstruktúrájáról vagy akár laboratóriumi kísérletekből származó adatok is. Ezek az innovációk a gyakorlatban új kutatási lehetőségeket teremtenek: összehasonlíthatóvá válik például a cégek által becsült és a tényleges bérpozíció, elemezhető a technológiai változások munkaerőpiaci hatása, vagy feltárhatók a munkaerő-felvétel során jelentkező nehézségek.

Egyes kezdeményezések – szigorú adatvédelmi szabályok mellett – például a munkavállalók nevei alapján próbálják megbecsülni származásukat, ezzel lehetővé téve az etnikai alapú munkaerőpiaci egyenlőtlenségek vizsgálatát is (például Németország esetében lásd *Hermansen és szerzőtársai* [2025]). A jövőben a nyelvi és osztályozási algoritmusok további lehetőségeket nyithatnak az adminisztratív adatok fejlesztésében, amennyiben ehhez az adatgazdák is hozzájárulnak.

A magyarországi Admin-adatbázisok ugyan nem bővíthetők szabadon új, egyéni szintű kérdőíves modulokkal, azonban aggregált szintű információk kapcsolhatók hozzájuk. Erre jó példa *Boza–Reizer* [2024] kutatása, amely során a szerzőpáros valószínűségi kapcsolás segítségével sikeresen összekötötte a Bértarifa-felvételből származó cégszintű aggregált bérjellemzőket az Admin3 adatbázisban megfigyelhető bérelőzlásokkal, így részletesebb képet kapva a kereseti szerkezetekről.

Az akadémián túl. Az adminisztratív forrásokra épülő paneladatok nemcsak az akadémiai kutatások számára fontosak, hanem a szakpolitikai döntéshozatal és a közpolitikai programok tervezése szempontjából is hasznos eszközt jelentenek. Jelenleg ezen adatbázisokat leginkább gazdasági szabályszerűségek vagy nagyobb szabályozási változások utólagos értékelésére használják. Ha azonban az adatok naprakészen, minimális késleltetéssel elérhető lennének, közvetlenebb szerepet játszhatnának a döntéshozatalban is. Bár sok ország évente frissíti munkáltató-munkavállaló adatbázisait, ez nem mindenhol bevett gyakorlat, pedig az adminisztratív nyilvántartások jellemzően havi szinten frissülnek. A folyamatos, gyorsan elérhető adatszolgáltatás segíthetne a gazdasági sokkok – például a koronavírus-járvány vagy a migrációs hullámok – azonnali munkaerőpiaci hatásainak megértésében, és ezáltal hatékonyabbá tehetné a közpolitikai reakciókat. Ráadásul a tudományos kutatásokat is élénkítené.

A tervek szerint a KRTK Adatbank Admin5 adatbázisa már évente fog frissülni, így biztosítva a fenti előnyöket. Erről azonban majd csak néhány év múlva írhatunk részletesebben. Addig is jó kutatást kívánunk!

Hivatkozások

- ABOWD, J. M.–KRAMARZ, F. [1999]: The analysis of labor markets using matched employer-employee data. Megjelent: *Ashenfelter, O.–Card, D.* (szerk.): Handbook of Labor Economics. Chapter 40. Vol. 3. Part B, 2629–2710. o. [https://doi.org/10.1016/S1573-4463\(99\)30026-2](https://doi.org/10.1016/S1573-4463(99)30026-2).
- BENEDEK DÓRA–ELEK PÉTER–KÖLLŐ JÁNOS [2012]: Adóelkerülés, adócsalás, fekete- és szürkefoglalkoztatás. Megjelent: *Fazekas Károly–Benczúr Péter–Telegdy Álmos* (szerk.): Munkaerőpiaci Tükör, 2012. MTA KRTK KTI–OFA, Budapest, 165–190. o. https://kti.krtk.hu/file/download/mt_2012_hun/egyben.pdf.
- BOZA ISTVÁN–PETŐ RITA [2025]: The Rise of Linked Employer-Employee Panel Data: Where Are We Now? KRTK KTI Working Papers, 2. sz. <https://kti.krtk.hu/wp-content/uploads/2025/02/KRTKKTIWP202502.pdf>.
- BOZA ISTVÁN–REIZER BALÁZS [2024]: The Role of Flexible Wage Components in Gender Wage Difference. IZA Discussion Paper, No. 17125. <https://doi.org/10.2139/SSRN.4889211>.
- GUVENEN, F.–PISTAFERRI, L.–VIOLANTE, G. L. [2022]: Global trends in income inequality and income dynamics: New insights from GRID. Quantitative Economics, 13. köt. 4. sz. 1321–1360. o. <https://doi.org/10.3982/qe2260>.
- HERMANSEN, A. S.–PENNER, A.–BOZA ISTVÁN ÉS SZERZŐTÁRSAI [2025]: Immigrant-native pay gap driven by lack of access to high-paying jobs. Nature, 644. köt. 969–975. o. <https://doi.org/10.1038/s41586-025-09259-6>.
- KÖLLŐ JÁNOS [2022]: A kapcsolt adminisztratív paneladatok hasznáról a KRTK Adatbank Admin adatbázisainak példáján. Megjelent: *Szabó-Morvai Ágnes* (szerk.): Munkaerőpiaci Tükör, 2021. KRTK–KTI, Budapest, 67–74. o. <https://real.mtak.hu/156600/1/mpt-2021.pdf>.
- OECD [2021]: The Role of Firms in Wage Inequality: Policy Lessons from a Large Scale Cross-Country Study. OECD Publishing, Párizs, <https://doi.org/10.1787/7d9b2208-en>.
- PENNER, A. M.–PETERSEN, T.–HERMANSEN, A. S. ÉS SZERZŐTÁRSAI [2022]: Within-job gender pay inequality in 15 countries. Nature Human Behaviour, 7. köt. 2. sz. 184–189. o. <https://doi.org/10.1038/s41562-022-01470-z>.
- SEBŐK ANNA [2022]: A KRTK kapcsolt államigazgatási paneladatbázisa – Admin4. Megjelent: *Szabó-Morvai Ágnes* (szerk.): Munkaerőpiaci Tükör, 2021. KRTK–KTI, Budapest, 76–80. o. <https://real.mtak.hu/156600/1/mpt-2021.pdf>.
- TOMASKOVIC-DEVEY, D. ÉS SZERZŐTÁRSAI [2020]: Rising between-workplace inequalities in high-income countries. Proceedings of the National Academy of Sciences of the United States of America, 117. köt. 17. sz. 9277–9283. o. <https://doi.org/10.1073/pnas.2009782117>.