

KRTK Adatbank bemutatása

Tir Melinda

.

2025.03.27.

Tevékenységünk

Nagyméretű-
panel adatbázisok
építése

Legfontosabb hazai
adatfelvételek
beszerzése,
harmonizálása

KSH-KRTK
Kutatószoba

Küldetésünk:
Használatra kész adatbázisok
létrehozása
tudományos
kutatások számára

KRTK
Adatszoba

Kísérleti labor

Big Data
csoport
(SSCUB)

Kutatásra előkészített állományok létrehozása

Miért van erre szükség? Tipikus hibák:

- Azonos név, eltérő tartalom, vagy fordítva
 - Időben változó kódértékek, ugratások
 - Diszkontálás, nettó/bruttó értékek számítása
- Harmonizálás
- Az adatbázis kiegészítése máshonnan származó adatokkal
 - Kiegészítés belülről: az információ ott van, de a kinyerése komoly munkával jár
- Kiegészítés
- Dokumentálás, a reprodukálhatóság biztosítása
 - **A Kutatók másodelemzést végeznek, az adattulajdonos továbbra is az adatgazda, az Adatbank és a kutatók használati jogot kapnak**

Adatok elérhetősége



Adatkérő lap kitöltése és elküldése

- Kizárólag tudományos cél
- Többségében KRTK-s társszerző, konzulens szükséges, vagy külső kutatói státusz



Hozzáférés többségében adatbanki szerveren keresztül

- A teljes munkafolyamat a szerveren történik
- Fájlok Statában, supportot is csak így tudunk biztosítani, de használható R és a jövőben Python is



Publikáció elküldése az Adatbanknak

KSH-KRTK Kutatószoba

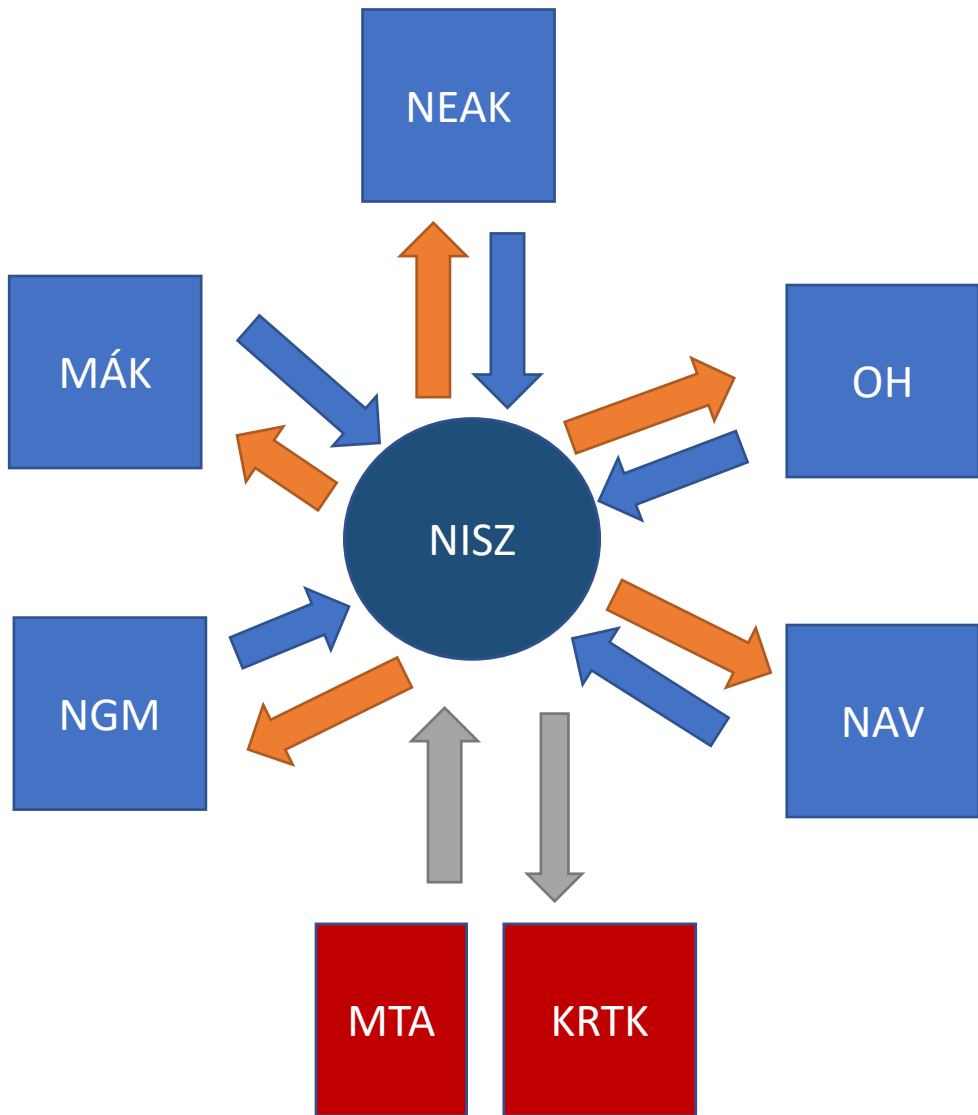
Projektet csak KRTK-s, vagy TK-s kutató vezethet.
De KRTK-s projekt esetén a résztvevők lehetnek
külsős kutatók.

Hozzáférés módja:

- 12 munkaállomás
- KSH egyetértésével indított projektek
- Szigorú szabályok
- Kamerás megfigyelés
- Output checking
- Stata, R, Office használat biztosítása



Kapcsolt Adminisztratív panelek



Adatbank → MTA Elnök → NISZ a 2007.CI tv alapján

NISZ → adatgazdák: hash1 (TAJ), hash2 (adóazonosító) kódoló algoritmusok

NEAK: 50%-os TAJ minta

Adatgazdák → TAJ alapján leválogatott állományok, hash1

MÁK, NAV, NGM: vállalati adóazonosítók alapján leválogatott állományok, hash2

NISZ: anonimizálás → KRTK: kapcsolt állomány

Eddig megvalósult adatkapcsolások

Időszak:2002-2009

4 602 000 egyén
(15-74 50%)
Nincs foglalkoztatói
információ
Adatgazdák száma: 3
OEP, ONYF, ÁFSZ

Időszak:2003-2011

4 601 999 egyén
(5-74 50%)
898 722 vállalat
Adatgazdák száma: 5
OEP, ONYF, NAV,
ÁFSZ, OH

Időszak:2003-2017

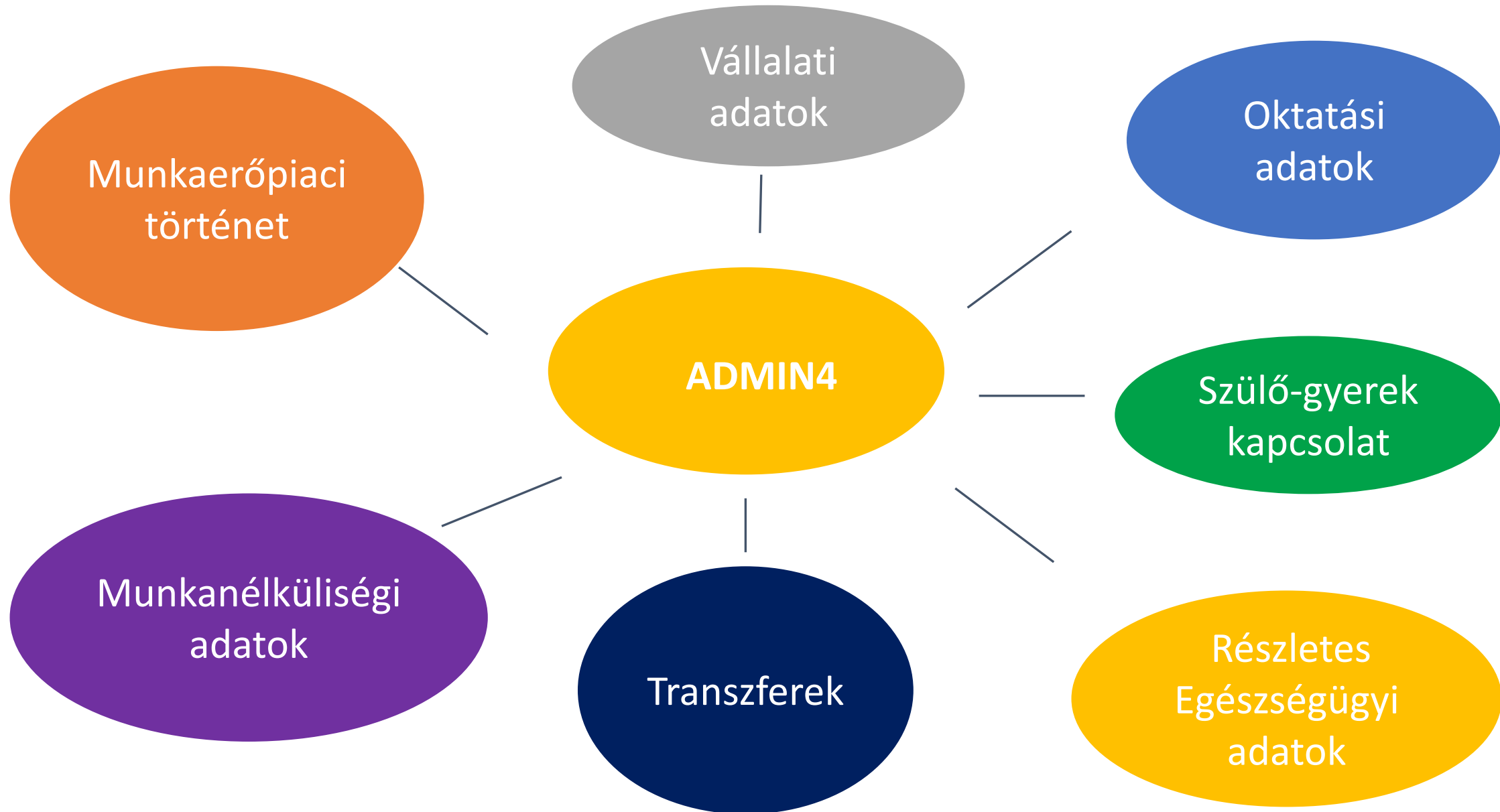
5 174 486 egyén
(0-tól 50%)
1 282 763 vállalat
Adatgazdák száma: 5
NEAK, ONYF, NAV,
NMH, OH
Egészségügyi, oktatási
adatkör bővítése

Időszak:2003-2021

6 393 158 egyén
(0-tól 50%)
1 717 598 vállalat
Adatgazdák száma: 6
NEAK, ONYF, NAV, NMH,
OH, NSZFH
Új belépők
egészségügyi, oktatási
adatkör további bővítése,
családi pótlék adatok
(anya-gyerek kapcsolat),



Admin4: 50 % Taj minta 2003-2021



Adatok elérhetősége

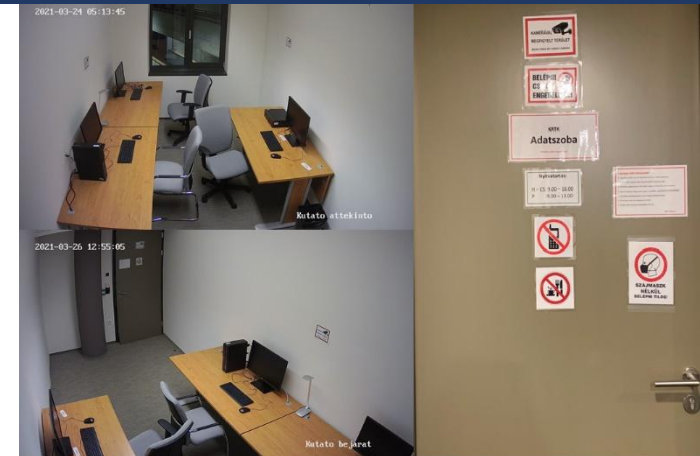
- Kizárólag szerveren keresztül
- Jelenleg elérhető: Admin3
- Admin4: Várhatóan tavasz végén KRTK-s kutatók és azok szerzőtársaik számára, Külsősök végső verzió elkészülése után
- **Hozzáférési szintek:**
 - **H1:** alapfájl és egyszerű kiegészítő fájlok. Szerveren keresztül, kutatási céllal
 - **H2:** részletes egészségügyi és oktatási adatok. KRTK-s kutatók, adott téma szakértői, előzetes engedélyezés után, szerveren
 - **H3:** eredeti részletességű adatok, teljes egészségügy. Adatszobai környezetben, részletes projekt tervvel, bizottsági engedélyeztetés után

KRTK Adatszoba

Cél: Adatvédelmi szempontból biztonságos környezet biztosítása érzékeny adatok felhasználása számára

Adatok: Teljes admin adatok, részletes egészségügyi adatok, kérés esetén más adatforrás is behelyezhető.

Működési mód: KSH szabályok átvétele, 3 munkaállomás, 4 kamera, külön beléptető rendszer, zárt serveres infrastruktúra, OC szabályok



Kapcsolat

- Adatbanki honlap: <https://adatbank.krtk.mta.hu/>
- Adatbanki adatok adatkérése: adatkeres@krtk.hun-ren.hu
- Kutatószobai ügyintézés: [kutatoszoba@krtk.hun-ren.hu](mailto:kutatooszoba@krtk.hun-ren.hu)
- Big-data csoport: sscub@krtk.hun-ren.hu
- Letölthető adatszettek: <https://sscu-budapest.github.io/explorer/>
- Adatbank általános kapcsolatfelvétel: adatbank@krtk.hun-ren.hu