

ELTE – Közgazdaság- és Regionális Tudományi Kutatóközpont Adatbankja

A KRTK ideje korán (2005-ben) felismerte a manapság a társadalomtudományokat is egyre inkább jellemző adatrobbanás jelentőségét és az eladdig decentralizált kutatói, vagy kutatócsoportok szintjén futó adatbeszerzési, adatfeldolgozási eljárásokat egy szervezeti egységbe foglalta és felvért az azt a szükséges adattudományi, adatgazdász, jogi, közgazdaságtani és IT ismeretekkel, továbbá a szükséges infrastruktúrával.

Így lett a KRTK Adatbankja a Kutatóközpont professzionálisan működő adatüzeme, amely 1) a kutatási adatpiacon elsőként végzett és végez nagy volumenű regiszter alapú adatkapcsolásokat (a 2007/Ci Tv. alapján), 2) kiépítette és üzemelteti Magyarország első, nem a KSH telephelyén működő KSH-Kutatószobáját, ahol 12 ellenőrzött terminálon keresztül a KSH teljes adatvagyonára elérhető a ELTE társadalomtudományi kutatásokkal foglalkozó kutatói és szerzőtársai számára, 3) szenzitív adatok kutatását lehetővé téve KRTK Adatszobát épített ki és üzemeltet, amely megfelel a TUKEB szigorú előírásainak, 4) 40 végpontos Kísérleti Labort épített ki és üzemeltet a kísérleti közgazdaságtan művelői számára, 5) adatbanki szerveres infrastruktúrát fejleszt és üzemeltet, 6) kurrens adatokat kumulál, tisztít, harmonizál és tesz elérhetővé a széles tudományos közösség számára, 7) Big data alapú adatfeldolgozási eljárásokat és kapacitást fejleszt, 8) módszertani tanácsadást nyújt az adathasználat, adatfeldolgozás és adatvédelem terén, továbbá 9) kurrens adathátteret nyújtva támogatja a hazai, társadalomtudományi profilú egyetemi képzéseket.

2025. I. félévében 449 kutató rendelkezik adatbanki szerveres adathozzáféréssel, 2012 óta 2024 adatkérést szolgáltatunk ki. Jelenleg 221 projekttag dolgozik 86 nemzetközi és hazai projekten adatbanki adatokon. A Kutatószobában 2018 óta 19 837 időpontfoglalást regisztráltunk. Az adatainkra építve 896 publikáció, 136 BA, MA és Ph.D. disszertáció, számos szakpolitikai hatásvizsgálat született itthon és külföldön, többek között olyan programokban, mint az ERC, H2020, Lendület, Kiválósági Együttműködési Program és OTKA. Fejlesztéseink és elért eredményeink révén 2021-ben kiérdemeltük a TOP 50 hazai tudományos infrastruktúra megtisztelő címet.

Az Adatbankban 7-8 fő dolgozik, akik intenzív, napi szintű kapcsolatot ápolnak a kutatókkal, és maguk is rendszeresen részt vesznek különféle kutatásokban. Ez elengedhetetlen a megfelelő színvonalú működéshez, mert az adatállományokban meglévő hibák és inkonzisztenciák egy részére csak az elemzés és az adathasználat során derülhet fény. A munkatársak közgazdász, vagy szociológus végzettséggel rendelkeznek, ami a magas szintű informatikai ismeretek mellett nélkülözhetetlen a tudományos igényeket kiszolgáló előkészítő munka elvégzéséhez.

ADATBÁZISOK HARMONIZÁLÁSA

A KSH és más szervek (Oktatási Hivatal, Állami Foglalkoztatási Szolgálat) adatfelvételei megbízhatóak és jó minőségűek, de tudományos célú felhasználásra közvetlenül nem minden esetben alkalmasak. A mai szttenderdeknek megfelelő kutatás csak hosszú időszakot átfogó longitudinális adatokon folytatható. Ehhez időben konzisztens panel vagy kvázipanel adatállományokra van szükség, ami megköveteli a változók és a kódolás időbeni harmonizálását, az ágazati, foglalkozási és területi kódváltozások áthidalását, kompromisszumos kódkészletek kialakítását. Az alapos elemzéshez a megfigyelt egyének és vállalatok rekordjait célszerű kiegészíteni a felvételeken kívüli információkkal, például földrajzi adatokkal és a diszkontáláshoz felhasználható ár-, bér- és adózási idősorokkal. A felvételekben meglévő, de az elemi rekordokban meg nem jelenő információkat ugyancsak érdemes kiaknázni, családi és háztartási, vagy ágazati és területi indikátorokat rendelve az egyénekhez és vállalatokhoz. Az Adatbank ezeket a munkákat folyamatosan elvégzi, sok hónapnyi munkát takarítva meg az egyes kutatóknak. Ennek köszönhetően például a Munkaerő-felvétel 1992 első negyedétől, a Bértarifa-felvétel 1986-tól napjainkig egységes adatállományként áll a tudományos igényű elemzés rendelkezésére. Elérhető adatbázisainkról bővebben: <https://adatbank.krtk.mta.hu/adatbazisok/elerheto-adatbazisok/>

KAPCSOLT ÁLLAMIGAZGATÁSI ADATGYŰJTEMÉNYEK

A nagyméretű kapcsolt államigazgatási panel adatbázisok a modern társadalomkutatásban nagy teret nyertek az elmúlt 10 évben (skandináv országok, Németország, Olaszország, Portugália), és jelentős részben kiváltották a sokkal drágább és sokszor kevésbé megbízható survey adatokat. Az Adatbank eddig három nagyméretű, államigazgatási regiszterek összekapcsolásával létrehozott egyéni szintű, munkáltatói, oktatási és egészségügyi adatokon alapuló panel adatbázist épített és jelenleg egy negyedik a megépítésén dolgozik, melyben több mint hatmillió ember követhető havi, vagy akár napi szinten 2003 és 2021 között. A panel iránt széleskörű kutatói érdeklődés mutatkozik, mivel változótartalma a legkülönbözőbb témákat öleli fel (oktatási részvétel, kompetencia eredmények, munkaerőpiaci státusz, foglalkozás, jövedelem, transferek, mortalitás és morbiditás, vállalati jellemzők, részletes egészségügyi adatok).

KSH-KRTK KUTATÓSZOBA

2018. február 1-jén a KRTK székhelyén megnyílt Magyarország első, a KSH telephelyén kívül működő Kutatószobája, ahol a kutatók teljes körű, egyéni és vállalati adatokat tartalmazó, egymással is összekapcsolható, 1970-ig visszanyúló adatállományokat használhatnak távoli ellenőrzés mellett. Az adatokat a KSH biztosítja, a programok a KRTK használtában lévő, a KSH főépületében elhelyezett szervereken futnak. A Kutatószoba nagy előrelépést jelent azon az úton, melyen a modern társadalomtudományi kutatás halad. A KSH-val közösen kialakított rendszer nem csak a régióban egyedülálló, hanem világviszonylatban is az élenjárók közé tartozik a technikai felszereltség, a kutatható adatvagyon és a hozzáférési feltételek szempontjából. A Kutatószobát a ELTE társadalomtudományi intézeteinek munkatársai, szerzőtársaik és asszisztenseik használhatják. A munka 12 terminállal felszerelt helyiségben végezhető szigorú biztonsági rendszabályok betartásával, amit – egyebek mellett – egy tizennégy kamerával felszerelt megfigyelőrendszer biztosít. A Kutatószoba felügyeletét a KSH és az Adatbank közösen látják el. Az eredmények a KSH által elvégzett, térítésköteles adatvédelmi ellenőrzés után hozhatók ki.

KRTK ADATSZOBA

2021 júniusában megkezdte működését az Adatbank által létrehozott KRTK Adatszoba. A kutatók itt biztonságos környezetben, a rangos nemzetközi publikációk esetében elvárt kutatásetikai és adathasználati elveknek eleget téve tudnak olyan adatokat elemezni, melyekhez külső környezetben nem férhetnek hozzá. Jelenleg elsősorban a Kapcsolt Államigazgatási Adatbázis (Admin3) részletes egészségügyi adatainak, illetve a teljes input állományainak a használatára van lehetőség, de kérés esetén más adatforrás is használható. Az Adatszoba kizárólag tudományos célból biztosít hozzáférést az adatállományokhoz, az egyedi statisztikai adatok kiemelt szintű védelme, illetve az adatvédelemmel kapcsolatos jogszabályok szigorú betartása mellett. Az Adatszobában kizárólag tömegjelenségek vizsgálhatók, nem indíthatók olyan kutatási projektek, melyekhez az egyes természetes egységek felfedése szükséges. Az Adatszoba kártyás beléptető rendszerrel és kamerás megfigyeléssel egyaránt rendelkező biztonságos informatikai környezetben használható. Kutatási eredmények az Adatbank munkatársai által elvégzett adatvédelmi ellenőrzést követően hozhatók ki.

BIG DATA CSOPORT

2021-ben az empirikus kutatási lehetőségek bővítésének, valamint a modern kutatási ökoszisztémák elterjedésének támogatására megalakult a Társadalomtudományi Számítástechnika Csoport. A csoport célja a magas színvonalú, reprodukálható és kiterjeszthető tömegjelenségeken alapuló kutatás támogatása, új módszerek alkalmazása. Munkájuk része nagy adathalmazok gyűjtése, feldolgozása és elemzése Python segítségével. Az általuk gyűjtött és feldolgozott adatokhoz többségében nyílt, korlátozásmentes hozzáférést tudunk biztosítani ld: <https://sscu-budapest.github.io/explorer/>

Jelentősebb adatgyűjtéseik: COVID-19 halálozások, geolokalizált Twitter adatok, magyarországi EU által finanszírozott projektek- adatai, kórházi várólisták, választási eredmények.

KÍSÉRLETI LABOR

A kísérleti közgazdaságtan a tudományág gyors ütemben fejlődő területe, melynek több kiemelkedő képviselőjét Nobel-díjjal ismerték el. 2018-ban felállítottunk egy 40 végpontos, társadalomtudományi kísérletek lefolytatására alkalmas Kísérleti Labort, mely segítséget nyújt kutatóinknak lépést tartani ezzel a tudományággal. Az infrastruktúra használatára nemcsak a HTK épületében van lehetőség, kidolgozott eljárásunkkal külső helyszínen is folytathatók kísérletek. Az elmúlt években már 13 kísérlet zajlott, melynek jelentős része külső helyszínen, középiskolákban valósult meg. A Kísérleti Labor kutatási célú használata külsős kutatók számára térítésköteles.

ADATBANKI GYAKORNOKI PROGRAM

Fontosnak tartjuk az adatközpontú szemlélet átadását a fiatal kutató generáció részére, ezért gyakornoki programunk keretében biztosítunk számukra képzést és egyben lehetőséget a központban folyó kutatásokba történő bekapcsolódásba. A résztvevők megvalósíthatják kutatási terveiket az általunk gondozott állományokon és munkájukhoz segítséget kapnak munkatársainktól.

ELÉRHETŐSÉGEINK:

Adatkérések, adathasználattal kapcsolatos ügyintézés: adatkeres@krtk.hu

Kutatószobai ügyintézés: kutatoszoba@krtk.elte.hu

Adatszobai ügyintézés: adatszoba@krtk.hu

Big data csoport: sscub@krtk.elte.hu

Központi kapcsolatfelvétel: adatbank@krtk.elte.hu