

CSERES-GERGELY ZSOMBOR–CSORBA GERGELY

Műkincs vagy működő tőke?

Gondolatok a kutatási célú adatok hozzáférhetőségéről

A cikk célja a közérdekű adatokhoz való hozzáférés lehetőségeinek tisztázása, ezen belül is a kutatási célra használható adatok hozzáférhetőségének elemzése. Bemutatjuk a Magyarországon jelenleg érvényben lévő jogi szabályozást, majd a közgazdaságtan normatív szempontjai szerint megvizsgáljuk, milyen lenne az optimális adat-szolgáltatási intézményrendszer. A cikk második felében rámutatunk a magyarországi intézmények működési hiányosságaira, illetve a helyzet javítására javaslatokat fogalmazunk meg.*

Journal of Economic Literature (JEL) kód: C80, D60, D78, K23, L43.

Az elmúlt húsz év nagy horderejű társadalmi-gazdasági folyamatainak kutatásához a különböző állami intézményekben felhalmozott hatalmas adatvagyon kivételesen jó alapot kínál. Ezek felhasználása nagyban segítené a tudományos megismerést, és a politika is könnyebben kontrollálhatóvá válhatna, ha bármely tőle független közösség is elvégezheti az elemző ellenőrzést. A pozitív példák mellett mégis számos olyan hazai esetet ismerünk, amikor kutatók komoly, sokszor leküzdhetetlen akadályokba ütköznek adatkéréseik során. Megfelelő adatok hiányában pedig sokszor le kell mondaniuk kutatási kérdéseik alapos megválaszolásáról vagy akár a teljes kutatásról. Ennek a minden fél számára rossz és rendezetlen helyzetnek az az oka, hogy nincs egyértelműen kijelentve: az adatok nemcsak az adatkezelő intézmények céljait szolgálják, hanem számos lehetőség kínálkozik további társadalmi hasznosításukra/hasznosulásukra is. Mivel azonban az adatok nyilvánosságra hozatala kényes kérdés, és a lehetséges hasznok nem minden esetben nyilvánvalók, ezért az adatgazdák sokszor inkább a kockázatkerülő bezárkózást választják.

Napjainkban az állam kiemelt szerepet tölt be a különféle adatok létrehozásában. Mivel ezt polgárai pénzén teszi, megalapozott az igény, hogy az ilyen adatokhoz való hozzáférést csak a minimálisan szükséges mértékben korlátozzák. Az elektronikus ügyintézés előretörésével a tájékoztatás gyakorlati problémáinak megoldásában jelentős előrelépés történt Magyarországon is (az *Ügyfélkapu-rendszer*, amelyben több állami szerv információk szolgáltatásai válnak egy helyen elérhetővé). Bár az elemzésre alkalmas alapadatok területén a magyar adatkezelő intézményrendszer elvileg továbbra is központi

* A cikk a Nemzeti Hírközlési Hatóság által támogatott, az Eötvös Károly Intézet keretein belül végzett Hozzáférés a közzsféra adataihoz című kutatás egyik eredménye. Köszönetet mondunk *Bálint Mónikának*, *Kézdí Gábornak*, *Köllő Jánosnak*, *Majtényi Lászlónak*, *Miklósi Zoltánnak*, *Nárai Csillának* és *Szabó Máténak*, akik számos vitában ösztönöztek minket a témáról kialakított elképzeléseink finomítására. A cikkben elhangzott állítások a szerzők saját véleményét tükrözik, nem tekinthetők sem az MTA KTI, sem a CEU Közép-európai Egyetem, sem a Gazdasági Versenyhivatal hivatalos álláspontjának.

szervezett, mégis sokszor átláthatatlanul szétaprózott. A számos törvényben lefektetett joganyag azonban szinte kizárólag az adatok védelmével foglalkozik. Ezért nem véletlen, hogy egyes adatkezelő intézmények sok esetben sajátjuknak tekintik az adatokat, többször ilyen módon védelmezik, sőt értékesítik azokat. A külső szemlélő számára úgy tűnhet, e téren nincs egységes rend, az állami intézmények adatkezelését, az adatvagyonnal végzett kereskedelmi tevékenységét, árképzését senki sem felügyeli.

Cikkünk célja a közérdekű, azon belül is a kutatási célra használható adatokhoz való hozzáférés lehetőségeinek tisztázása. Először röviden bemutatjuk a jelenleg érvényben lévő jogi szabályozást, majd részletesebben is megvilágítjuk a kutatási adatok speciális tulajdonságait. Ezután közgazdaság-elméleti nézőpontból megvizsgálunk néhány, adatszolgáltatással kapcsolatos hiedelmet, illetve bemutatjuk, hogy milyen lehetne az optimális adatszolgáltatási intézményrendszer. A cikk második felében kritikai szemmel, példákön keresztül elemezzük a magyar adatszolgáltatási intézmények működését, és – a fennálló jogi-közgazdasági környezetet figyelembe véve – ajánlásokat fogalmazunk meg annak javítására.

A közérdekű adatokra vonatkozó jogi szabályozás áttekintése

Az, hogy egy érdeklődőnek milyen lehetősége van az állami finanszírozásban létrejött adatvagyon elérésére – például tudományos kutatás számára – mindenekelőtt jogi kérdés.¹ A közérdekű adatokhoz való hozzáférés igénye már a Magyar Köztársaság Alkotmányában is megfogalmazódik.

„A Magyar Köztársaságban mindenkinek joga van a szabad véleménynyilvánításra, továbbá arra, hogy a közérdekű adatokat megismerje, illetőleg terjessze.” [Alk. 61. § (1) bekezdés.]

Maga az adat meglehetősen tág fogalom. A jog minden ismeretet, információt adatként kezel, amelyből bármiféle következtetés levonható, és adatnak minősül minden ilyen ismeretből keletkező további ismeret is. Egy statisztikai felmérés esetében már magát a kérdőívet és annak kitöltési útmutatóját, a kitöltött kérdőíveken szereplő válaszokat, az ezekből született számítógépes adatbázist is ide kell érteni. Hasonlóképpen az adatok előbbi értelmezése nyomán keletkezett elemzés is adatnak minősül, valamint a hozzáférésükről szóló információk is.

A különféle, természetes személyekre vonatkozó adatfajtákat a jog érzékenysége alapján határolja el. Ezt a megkülönböztetést és az adatok nyilvánosságra hozatalának alapelveit az *adatvédelmi törvényben* találjuk meg.²

„E törvény alkalmazása során

(1) személyes adat: a meghatározott természetes személlyel (a továbbiakban: érintett) kapcsolatba hozható adat, az adatból levonható, az érintettre vonatkozó következtetés. A személyes adat az adatkezelés során mindaddig megőrzi e minőségét, amíg kapcsolata az érintettel helyreállítható;

...

(3) közérdekű adat: az állami vagy helyi önkormányzati feladatot, valamint jogszabályban meghatározott egyéb közfeladatot ellátó szerv vagy személy kezelésében lévő, a személyes adat fogalma alá nem eső adat;” (1992. évi LXIII. törvény 2. §).

¹ A jogszabályok részletesebb bemutatására és értelmezési lehetőségeire lásd *Majtényi-Szabó* [2006].

² 1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról. Megtekinthető: http://www.mkogy.hu/adatved_biztos/AvTv.htm

A legalapvetőbb kérdés tehát az, hogy az adott intézmény kezelésében levő adat személyes-e, vagyis kapcsolatba hozható-e az érintettel. Amennyiben a válasz nem, akkor az adat közérdekű, harmadik eset nem létezik. További fontos következmény, hogy amennyiben egy személyes adatot „*állami vagy helyi önkormányzati feladatot, valamint jogszabályban meghatározott egyéb közfeladatot ellátó szerv vagy személy*” kezel, és megszüntethető az adat kapcsolatba hozhatósága az érintett személlyel, akkor az adat művelet elvégzése után közérdekűvé válik. Hasonlóan igaz, hogy az előbbi szervek által közérdekű adatok alapján létrehozott újabb adatok (adatbázisok, elemzések, jelentések) szintén közérdekűnek minősülnek.

A nyilvánosság az adatokra vonatkozó adatokra, az úgynevezett metaadatokra (például az adatok elérhetősége), az adott intézmény gazdálkodására, különösen pedig az adatok nyilvánosságra hozatala (vagy személyes, vagy üzleti titkot tartalmazó adatok *potenciális* nyilvánosságra hozatala) során felmerült költségekre, illetve bevételekre is vonatkozik:

„(1) Az állami vagy helyi önkormányzati feladatot, valamint jogszabályban meghatározott egyéb közfeladatot ellátó szerv és személy (...) a feladatkörébe tartozó ügyekben – ideértve a gazdálkodásával kapcsolatos ügyeket is – köteles elősegíteni a közvélemény pontos és gyors tájékoztatását.

(2) Az (1) bekezdésben említett szervek rendszeresen közzé- vagy más módon hozzáférhetővé teszik a tevékenységükkel kapcsolatos legfontosabb – így különösen a hatáskörükre, illetékességeikre, szervezeti felépítésükre, a birtokukban lévő adatfajtákra és a működésükről szóló jogszabályokra vonatkozó – adatokat.” (1992. évi LXIII. törvény 19. §.)

A törvény tehát az adatkérésnek nem pusztá tűrését teszi kötelezővé, hanem azt is, hogy az egyes intézmények proaktív módon, időről időre „adatvagyon-nyilatkozatot” tegyenek.³ Ennek értelmében például az APEH külön kérés nélkül köteles informálni a közvéleményt arról, hogy birtokában vannak többek között olyan, számítógépes formában tárolt adatbázisok, amely a gazdasági társaságok mérlegadatait vagy az adózók száadatait tartalmazza. Noha például az APEH tevékenységét átgondolva, ez az ismeret triviális, egész más az adatkérő helyzete, ha nyitott menüből válogatva fogalmazza meg adatkérését, mintha úgy „merészkedik oda” az adatgazdához, hogy lehetőségeivel nincs pontosan tisztában.

„(3) Az (1) bekezdésben említetteknek lehetővé kell tenniük, hogy a kezelésükben lévő közérdekű adatot bárki megismerhesse, kivéve, ha az adott törvény alapján az arra jogosult szerv állam- vagy szolgálati titokká nyilvánította, továbbá, ha a közérdekű adatok nyilvánosságához való jogot – az adatfajták meghatározásával – törvény (...) korlátozza.” (1992. évi LXIII. törvény 19. §).

Az adatkezelő szervek birtokában lévő közérdekű adatoknak tehát alapértelmezésben nyilvánosan hozzáférhetőeknek kell lenniük. Ez alól kivételt jelentenek azok az adatok, amelyeket a megfelelő törvények felsorolnak, ilyen például az üzleti vagy szolgálati titok. Ezekben az esetekben az adatkezelő szerv megtagadhatja annak kiadását, de ennek bizonyítása minden esetben az ő feladata, és nem az adatkérőé:

„(2) A megtagadás jogszerűségét és megalapozottságát az adatot kezelő szerv köteles bizonyítani.” (1992. évi LXIII. törvény 21. §.)

Ha egy adatról viszont bebizonyosodik, hogy az személyes vagy például üzleti titoknak minősül, akkor az adatgazdát jelenleg nem lehet arra kötelezni, hogy mentesítse adatait személyes jellegűtől, illetve hagyja el az üzleti titoknak minősülő elemeket. Ez a

³ Érdemes megfigyelní, hogy ez a kötelezettség nemcsak a közérdekű adatokra vonatkozik.

helyzet független attól, hogy sok esetben könnyedén elvégezhető a szükséges művelet. A szabályozás ugyanakkor világossá teszi, hogy ha erre lehetőség lenne, a közérdekű adatok száma jelentősen megnőne.

Az OEP kezelésében levő gyógyszerfogyasztási adatbázis jó példa a személyes és közérdekű adat közötti határvonal rendkívüli vékonyságára. Ez az adatbázis igen gazdag, választ ad például arra, hogy egy biztosított egy adott időszakban hány és milyen receptre felírt gyógyszer fogyasztott. Az ilyen, főleg időben összekapcsolható adatok igen fontos kérdések megválaszolására adnak lehetőséget. Mint minden adminisztratív adatbázis, ez is tartalmaz egyéni központi azonosítót (itt a taj-azonosítót), amely az adatokat személyessé teszi, az adatgazda tehát az adat jellegének ismertetésén túl semmire sem köteles, főleg nem köteles kiadni az adatokat. Érezhető azonban, hogy ennek megtörténtehez csak egy igen apró lépésre van szükség: a taj-számok elhagyására. A művelet percek veszt igénybe, és óriási értékkel növeli a rendelkezésre álló adatvagyon.

Egy másik példa az APEH kezelésében levő, az egyes gazdasági társaságok kötelező adatszolgáltatásából származó mérlegadatokra vonatkozik. Az egyes cégek adatai mindenki számára nyilvánosak, kikérhetők a cégbíróságtól. Az APEH adatbázisa pusztán ezeknek az adatoknak az összessége, semmilyen üzleti titkot, érzékeny pluszinformációt nem tartalmaz. Kézenfekvő lenne, hogy szabadon hozzáférhető legyen, akár az interneten letölthető formában, ami óriási megtakarítást jelentene az érdeklődő kutatók számára. Az adatbázis azonban nemcsak nem hozzáférhető, de kiadását kutatók számára az APEH több ízben is megtagadta, holott erre nincs törvényi lehetőség.

A közérdekű adatok a jelenlegi szabályozás szerint ingyenesek – az ilyen adatok árazásával kapcsolatos megfontolásokról később részletesen szólunk. A vonatkozó szabályozás ma az adatok külön feldolgozása, átadása során keletkező költségek felszámítását teszi lehetővé:

„(1) A közérdekű adat megismerésére irányuló kérelemnek az adatot kezelő szerv a kérelem tudomására jutását követő legrövidebb idő alatt, legfeljebb azonban 15 napon belül, közérthető formában tesz eleget. Az adatokat tartalmazó dokumentumról vagy dokumentumrészről annak tárolási módjától függetlenül – költségterítés ellenében – a kérelmező másolatot kérhet.

...

(3) A közérdekű adat közléséért az adatkezelő szerv vezetője – legfeljebb a közléssel kapcsolatban felmerült költség mértékéig – költségterítést állapíthat meg. A kérelmező kérésére a költség összegét előre közölni kell.” (1992. évi LXIII. törvény 20. §.)

Mivel a legtöbb szóban forgó adatbázist elektronikus formában, számítógépen tárolják, az átadás technikai költsége minimális, a folyamat jóformán csak a dokumentáció elkészítésében merül ki.⁴ Ha tehát valaki az APEH mérlegadat-állományát szeretné megkapni, az ezért felszámított költség reálisan nem lépheti át a (néhány) tízezer forintot. A szabályozás fonákja ugyanakkor, hogy az adat átadásának költségébe nem számítható be az adat gyűjtésének költsége, legyen az akár több milliárd forint. Más a helyzet akkor, ha az adatokat a hozzáférés lehetővé tétele miatt kell anonimizálni. Mivel lehetséges, hogy ez hosszadalmas munka, jelentősen megemeli az adatkérés költségeit – kivételt jelentenek természetesen azok a helyzetek, amikor az adatigénylő olyan anonimizált adatot igényelt, amelyet egyszer már előállítottak.

Látnunk kell, hogy az adatkérő nemcsak az átadás díját, hanem a díjszámítás módját, az alkalmazott technológiát is megismerheti, hiszen az is közérdekű adat. Ez automatikus vagy gyorsított jogorvoslatra nem ad lehetőséget, de egy jogi eljárásnál segítség

⁴ Egy adatbázis-kezelő programból egy adatbázist kinyerni negyedórát, az adathordozóra másolás pedig percek kérdése.

lehet, amennyiben az adatkérő irreálisan magasnak találja az adatátadásért felszámított összeget.

Kérdés azonban, hogy tételesen mely intézmények tartoznak az adatvédelmi törvény által megjelölt intézményi körbe. Sajnos nem létezik ma olyan jogszabályban rögzített definíció vagy lista, amely egyértelműen kijelölné a közfeladatokat vagy az azokat ellátó szerveket, pedig az adatvédelmi törvény ezzel a megjelöléssel él. Ez az értelmezés csak néhány törvény esetében történik meg az éppen szabályozni kívánt területen. Ilyen értelmezés található például a számos, az adatkezelést ágazati és intézményi szinten szabályozó jogszabály közül kiemelkedő *statisztikai törvényben*.⁵ Már a törvény első paragrafusa világossá teszi, hogy a statisztikai tevékenység során keletkezett adatok esetében azok közzététele felvételükkel és kezelésükkel egyenrangú fontosságú esemény:

„... e törvény – összhangban a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról szóló 1992. évi LXIII. törvénnyel – az adatok statisztikai módszerekkel történő felvételét, feldolgozását, tárolását, átadását, átvételét, elemzését, szolgáltatását, közlését, valamint közzétételét (a továbbiakban: statisztikai tevékenység) szabályozza.” (1993. évi XLVI. törvény 1. §.)

A statisztikai adatgyűjtés eredményeinek nyilvánosságra hozhatóságáról a törvény egyértelműen rendelkezik, ha tudjuk, mi az adatgyűjtés eredménye. A probléma abban rejlik, hogy nincs pontosan definiálva, mi ez az eredmény. Az egyik szélsőséges álláspont szerint már a kitöltött kérdőívek összességét is annak tekinthetjük, de a másik szélsőség szerint az eredmény az, amit az intézmény annak minősít, ami akár egy egyoldalas összefoglaló is lehet.

(1) A hivatalos statisztikai szolgálathoz tartozó szervek által végrehajtott adatgyűjtések eredményei – a (2) bekezdésben foglalt kivétellel – nyilvánosak. A nyilvánosságra hozásról e szervek saját hatáskörükben gondoskodnak.

(2) Nem lehet nyilvánosságra hozni az államtitoknak vagy szolgálati titoknak minősített adatokat, valamint – a 18. §-ban foglaltak kivételével – a statisztikai célt szolgáló, a természetes és a jogi személy, valamint a jogi személyiséggel nem rendelkező adatszolgáltatóval kapcsolatba hozható adatot (a továbbiakban: egyedi adat).” (1993. évi XLVI. törvény 17. §.)

A kétértelműség ellenére a statisztikai törvény nem írhatja felül az adatvédelmi törvény rendelkezéseit. Éppen ezért, ha azt szolgálati titoknak nem minősítik, az egyes adatgyűjtési munkálatok részleteinek, költségének, a statisztikai szolgálat tagjai bevételeinek és kiadásainak is nyilvánosaknak, megismerhetőeknek kell lenniük.

Jól látszik továbbá, hogy a statisztikai törvény „egyedi adat” megjelöléssel egyszerre védi a természetes személyeket és másokat. Megfogalmazása ugyanakkor bizonyos esetekben értelmetlenül restriktív, mert nincs tekintettel arra, hogy a szolgálat tagjainak birtokában levő adatok mennyire nyilvánosak. Az egyes intézmények akkor sem adhatnak ki egyedi, mondjuk vállalkozásokkal kapcsolatos adatokat, ha azok más, esetleg kényelmetlenebb módon elérhetők.

Az 1993. évi XLVI. törvény 3. § (2) pont felsorolása szerint a hivatalos statisztikai szolgálathoz tartozó szervek körébe tartozik a Központi Statisztikai Hivatal, a minisztériumok és a Miniszterelnöki Hivatal, a Magyar Nemzeti Bank, a Gazdasági Versenyhivatal és a Legfelsőbb Bíróság. A törvény előírja, hogy a statisztikai szervek között a közérdekű adatok áramlásnak akadálytalanul kell lezajlania.

⁵ A statisztikáról szóló 1993. évi XLVI. törvény. Megtekinthető: http://portal.ksh.hu/portal/page?_pageid=37,126432&_dad=portal&_schema=PORTAL

„(1) A hivatalos statisztikai szolgálathoz tartozó szerv a program végrehajtásából rendelkezésre álló adatokat a hivatalos statisztikai szolgálathoz tartozó másik szervnek – annak feladatai ellátásához – kívánságára köteles továbbítani (adatátadás). Az adatigénylés és adatátadás – a (4) bekezdésben foglalt kivétellel – egyedi adatra nem vonatkozhat.

(2) Az adatátadó a rendelkezésére álló, kérdőíven, táblán tárolt adatok kigyűjtését, átmásolását az adatátvevőnek térítésmentesen köteles lehetővé tenni. A gépi adathordozón tárolt adatok adatátadó által végzett gépi átmásolásának, újbóli feldolgozásának költségeit az adatátvevő az adatátadónak – ha igényli – megtéríti.” (1993. évi XLVI. törvény 21. §.)

Bár furcsa, hogy a kérdőívek adott esetben rendkívül költséges feldolgozását a szolgálat tagjainak egy másik tag kérésére térítésmentesen kell elvégeznie, az adatátadás értelmezhetetlenül csekély költsége pedig felszámolható, ez az árazás közgazdasági érvekkel még megmagyarázható. Ugyanakkor sokkal nagyobb figyelmet igényel az a védelem, ami tükrözi az adatvédelmi törvény szabályozását, a korábban már említett korlátozással: egyedi adat nem adható át.⁶ Noha az adat itt is megfosztható egyediségétől, ezt a lehetőséget a törvény nem említi, erre az adatgazda ismét nem kényszeríthető. Bár az adatátadás díjmentes, a másolás költségeit az adatátadó felszámolhatja. Itt is érvényes azonban az elv, amit az adatvédelmi törvény esetében már megismertünk: ha az adatok léteznek, és már feldolgozták, anonimizálták őket, az adatkérés költsége csak az átadás költsége lehet.

Az adatok felhasználás szerinti osztályozása és a kutatási célú adatok sajátosságai

A jogi szabályozásból látható, hogy nemcsak a kutatók érdeklődési körébe tartozó adatok tartoznak a közérdekű adatok körébe. Maga a közérdekű elnevezés is azt sugallja, hogy ezek azok az adatok, amelyek potenciálisan minden állampolgár érdekét szolgálják. Az adatvagyon hatékony felhasználását firtatva azonban ésszerűnek látszik egy alternatív, a felhasználás módját leképező kategorizálás bevezetése.

Korábbi kutatásunkban az adatok két formáját különböztettük meg (*Cseres-Gergely-Csorba* [2006]). Egyrészt beszélhetünk *klasszikus adatokról*, amelyeket minden állampolgár közvetlenül felhasználhat, jellemzően egyéni döntések hatékonyságának növelésére.⁷ Másfelől definiáltuk a *kutatási célú adatokat*, amelyek jellemzően csak elemző feldolgozás után hasznosulhatnak a társadalom tagjai számára. Jelen tanulmányban mi csak a második csoportba tartozó adatokra vonatkozó következtetéseket tárgyaljuk.

A kutatási célú adatok jellegzetessége, hogy sok esetben magánemberek személyére vonatkoznak, és egy harmadik személy, a kutató szerez hozzájuk hozzáférést (ilyen például az összes lakossági statisztikai adatfelvétel). Az ilyen adatok lehetnek jogi értelemben személyesek (amennyiben az ismeret a konkrét személlyel összekapcsolható), vagy közérdekűek (ha a kapcsolat nem teremthető meg).⁸ Az utóbbi esetben semmilyen adat-

⁶ Meg kell jegyeznünk, hogy ez a korlátozás igen szélsőséges. Nemcsak azt tiltja, hogy az egyedi adatok a nagyközönség kezébe kerüljenek, de azt is, hogy az egyébként egy cél érdekében létrehozott statisztikai szolgálat tagjai kezeljék azokat. Mivel az egyes adatbázisok könnyen anonimizálhatók, összekapcsolásukat pedig az egységes adminisztratív azonosító hiánya eleve akadályozza, a tiltás nagyfokú, megfelelő nyomon követhetőség biztosítása esetén indokolatlan bizalmatlanságot tükröz. Mindez nyilvánvalóan óriási hatékonyságvesztést jelent.

⁷ Ilyenek például a jogszabályok, egy intézmény ügyfélszolgálati rendje, a telekkönyvi kivonat vagy a másnapi időjárásra vonatkozó meteorológiai adatok (utóbbi már kutatási célú alapadatok feldolgozásának eredménye).

⁸ Léteznek természetesen olyan adatbázisok, amelyek személyekre vonatkozó, nem személyes adatokat kezelnek, de nem közérdekűek. Ilyen minden, a vállalkozások kezelésében levő egyéni szintű adatbázis. Lévé, hogy ezek nem közpénzen jöttek létre, ezért figyelmen kívül hagyjuk őket.

védelmi probléma nincsen, az adatok ugyanúgy kezelhetők, mint a semmilyen egyéni jelleggel nem bíró, például nemzetgazdasági szintű nem titkos adatok. Az első esetben azonban az adatvédelmi intézkedések különféle változataira lehet szükség az adat és az elemzés jellege függvényében.

A konfliktus egyszerű: a kutató olyan adatokat akar elemzéséhez használni, amelyek „láthatását” az adatvédelmi törvény tiltja. Hiába alkalmaz jellemzően olyan eszközöket, amelyek csak az adatok összességén képesek működni, és hiába érdektelenek számára önmagukban az egyes emberek személyes adatai, sajnos, ma még nem oldható meg az, hogy a kutatóknak kiadott adatokat másként, az egyediségre koncentráló módon se lehessen felhasználni. Ahhoz tehát, hogy az ilyen adatok elemzésre átadhatók legyenek, meg kell akadályozni, hogy az egyes személyes adatokhoz az adatkérő hozzáférhessen, vagy visszaélhessen velük.

Mivel a személyes adatokat a statisztikailag is hasznos információnak egy adott személyhez kötése teszi jogilag személyessé, vagyis adatvédelmi szempontból aggályossá, kézenfekvő lehetőség a kapcsolat megszakítása. Erre sok esetben lehetőség van gyakorlatilag nulla információvesztéssel, esetleg némi pluszköltség vállalásával. A technikai megvalósíthatóságra jó példa, ha egy lakossági adatfelvétel személyekre vonatkozó adatai között a földrajzi azonosítók értékesek egy kutatás számára – mondjuk azért, mert a kutató a helyi mikrokörnyezetnek a munkavállalásra gyakorolt hatását szeretné számszerűsíteni. Megoldást jelenthet, ha a kutató nem kapja meg a valódi azonosítókat, csak véletlen számokat, amelyeknek a valódi azonosítóhoz rendelését az adatgazda megőrzi. Így az adatgazda a kívánt pluszinformációt a felhasználó kérésére hozzákapcsolhatja a törzsadatállományhoz anélkül, hogy adatvédelmi aggályok merülnének fel (ilyen gyakorlat alakult ki például a KSH munkaerő-felvétele esetén). Ilyenkor az adathozzáférés és kutatási célra átadás költsége mindössze a fenti műveletek elvégzésének elhanyagolható munkaköltségével nő meg.

Meg kell jegyeznünk, hogy a jogszabályokat a lehető legszigorúbban értelmezve, a fent leírt anonimizálási eljárás önmagában nem elégséges. Elképzelhető ugyanis, hogy egy személyről csak annyit lehet tudni, hogy melyik településen lakik, de jellemzői annyira sajátosak, hogy az illető helyen – mondjuk egy kis faluban –, ha mindenkit végigkérdezzünk, van esélyünk megtalálni az adott személyt. Ez az eljárás sértheti az adatvédelmi törvény előírásait. A nemzetközi gyakorlat azonban azt mutatja, hogy ha már csak ilyen, igen komoly befektetéssel megteremthető kapcsolat maradt, azt nyugodtan tekinthetjük nem létezőnek. Mindez legmarkánsabban az 831/2002/EK bizottsági rendeletben⁹ jelenik meg, amely figyelembe veszi a személyek azonosításának költségét. Ha az rendkívül nagy, az adatot anonimizáltnak tekinthetjük.

Vannak azonban olyan esetek, ahol a kapcsolat vagy nem szakítható meg, vagy a megszakítás megengedhetetlen mértékben rontaná az adat használhatóságát. Ilyenkor két megoldás alkalmazható. Az első az, hogy a kritikus azonosító megmarad, de az adattartalmat beható vizsgálatnak vetik alá, megállapítandó, hogy van-e valóban azonosítható személy. Amennyiben ilyet találnak, az adott személyre vonatkozó információ részben vagy egészében eltávolítható az adatbázisból (lásd később a Népszámlálás egyéni mintájával kapcsolatos eljárást). Ez a módszer igen költséges, hiszen ha valaki komolyan veszi, az egész adatbázist át kell fésülni, kritikus esetek után kutatva. A munka heroikus méretéből adódóan így sem biztos a százszázalékos találati arány.

⁹ A bizalmas adatokhoz való tudományos célú hozzáférés tekintetében a közösségi statisztikákról szóló 322/97/EK tanácsi rendelet végrehajtásáról szóló, 2002. május 17-i 831/2002/EK bizottsági rendelet (<http://eur-lex.europa.eu/Notice.do?mode=dbl&lang=hu&ihmlang=hu&lng1=hu,hu&lng2=cs,et,hu,lt,lv,mt,pl,sk,sl,&val=386940:cs&page=>). A rendelet e kérdés mellett több, témánkba vágó kérdést is tárgyal; részletesen kitér például a kutatási célú adatfelhasználás módjára és támogatására.

A második megoldás az adatokat érintetlenül hagyva olyan jogi és gyakorlati konstrukciót találni, amelyben a kutató ellenőrzött körülmények között fér hozzá az adatokhoz, illetve kötelezettséget vállal azok titkos kezelésére (ilyen gyakorlat ma Magyarországon nem létezik). Egy-két felhasználás esetében a költség viszonylag nagy: megfelelő számítástechnikai és jogi háttérrel kell kiépíteni. Mivel azonban a költség egyszeri, az ilyen jellegű adathozzáférés rutinná válásával egyre gazdaságosabb.

Az adathozzáférés közgazdasági szempontjai

Tekintsük most át a közhasznú adatok felhasználásáról szóló vita néhány fontosabb állítását, amelyek szerintünk közgazdasági szempontok alapján finomításra vagy mélyebb megalapozásra szorulnak.

A nemzeti adatvagyon közjósága. Egy szolgáltatást akkor nevezünk közjóságnak, ha egy gazdasági szereplő számára való juttatása esetén minden további gazdasági szereplő is költségmentesen hozzájuthat. A közjóság egyik tulajdonsága, hogy az azt használók között nincs rivalizálás, vagyis egy szereplő „fogyasztása” következtében a többi szereplő által igénybe vehető mennyiség nem csökken. Ez a tulajdonság adatok esetben nyilvánvalóan teljesül. A közjavak másik tulajdonsága, hogy fogyasztásukból nem zárhatók ki egyes szereplők. Ennek értelmében a nyilvánosan közzétett adatok valóban közjavak, ilyenek például a szabadon hozzáférhető internetes honlapokról letölthető adatok.

Az adatszolgáltatás azonban önmagában még nem tekinthető közjóságnak, ugyanis az adat kezelőjének lehetősége nyílhat árazási vagy diszkriminatív eszközökkel egyes szereplőket kizárni a hozzáférésből, vagyis az ő tudatos döntése, hogy az adat mennyire ölt közjóságjellegűt. Az adatvédelmi törvény és az elektronikus információszabadságról szóló törvény¹⁰ szelleme szerint a közhasznú adatokat valóban közjósággá kell tenni, de ez egyrészt nem minden esetben valósul meg, másrészt – mint amellet érvelni fogunk – ez nem is minden esetben optimális.

A közhasznú adatokhoz való hozzáférés gyakorlatilag költségmentes. Mivel a legtöbb adat ma már információs jószág,¹¹ ezért amennyiben egyszer már előállították, akkor az már tetszőleges mennyiségben reprodukálható (másolható), határköltsége tehát konstans és elhanyagolható. Az első példány létrehozásának fix költsége ugyan igen magas, de mivel az állami adatszolgáltató és -kezelő szervezeteknek vagy törvényi kötelessége az adatok gyűjtése, vagy tevékenységük melléktermékeként állnak ezek elő, ezért úgy tűnhet, hogy az aktuális hozzáférés optimális mértékének meghatározásánál csak a minimális változó költségeket kellene figyelembe vennünk.

Ez az érvelés ugyanakkor nem teljes. Először is, nem veszi figyelembe azokat a fix költségeket, amelyek az adatszolgáltatás egymást követő periódusaiban jelentkeznek. Ez egyrészt a meghatározott adat-hozzáférési szint biztosításának a költsége (például a szerverek kapacitása vagy sebessége), ami annál magasabb, minél alacsonyabb az ár, vagyis minél nagyobb a kereslet. Másrészt, a hozzáférés színvonalát folyamatosan fenn kell tartani, ami újabb adatok felvételéből, az adatbázisok aktualizálásából áll. Ezekre a jövőben jelentkező költségekre meg kell teremteni a forrást, ellenkező esetben hosszú távon romlik a szolgáltatás minősége.

¹⁰ 2005. évi XC. törvény az elektronikus információszabadságról. Megtekinthető: www.nda.hu/resource.aspx?ResourceID=2005_evi_XC_tv_VI

¹¹ Néhány kivétel azért létezik, mint például az Országos Földtani Intézet kezelésében lévő talajminták.

Másodszor, bár egy adott típusú adat termelési folyamata során valóban megfigyelhető egyfajta méretgazdaságosság (másképpen növekvő mérethozadék), aminek alapján az adott adat szolgáltatása tekinthető természetes monopóliumnak, érdemes észrevennünk a különböző típusú adatok előállításában és szolgáltatásában felmerülő kölcsönhatásokat. Ha egy szervezet végzi a munkavállalókra vonatkozó statisztikák összesítését, és ebből különböző mélységű (minőségű) adatbázisokat állít elő, akkor azt jó eséllyel költséghatékonyabban teheti, mint ha külön szervezet foglalkozna minden foglalkozási ággal. Ez az úgynevezett választékgazdaságosság, valamint hogy a különböző típusú adatok kezelésében egy idő után fellép a csökkenő mérethozadék adja az egyes adatokra specializált szervezetek létjogosultságát.

Az önköltségen alapuló ár elve. Az adatvédelmi törvény előírása szerint közérdekű adatok esetében a költségtérítés legfeljebb a közléssel kapcsolatban felmerült költség mértékéig terjedhet, a tényleges költségek megtérítésén túl profitelemet nem tartalmazhat. Ez lényegében a klasszikus határköltségen alapuló árazás, amely a fentebb elemzett költségstruktúra miatt gyakorlatilag az állam kezelésében lévő adatvagyon teljes nyilvánosságához vezetne.¹² Az eredmény társadalmilag optimális (úgynevezett első legjobb), mivel a piacon minden értékes tranzakció megvalósul abban az értelemben, hogy azok haszna meghaladja azok költségét.

Az elv megvalósításának további előnye, hogy információigénye minimális, mindössze tudnunk kell, hogy az adott adat iránti kereslet negatív meredekségű, és hogy azoknak az adatoknak a szolgáltatását írjuk elő, amelyekből származó bruttó fogyasztói többlet meghaladja az adat előállításának fix költségét. Ugyanakkor arra a kérdésre a válasz, hogy ki viselje az adatok előállításának költségét, nem jelenik meg ebben az elvben.

Az érvényes szabályozás implicit módon tartalmazza: az állam határozza meg a biztosítandó adatok körét, és teremtsen meg erre az anyagi feltételeket. A gyakorlatban azonban a megfelelő törvények csak általánosságban jelölik ki ezeket az adatokat, a specifikálás már az adatszolgáltató szervezetek feladata. Mivel a szervezet a pénzbeli költségeket nem, vagy csak részlegesen, a munkához kapcsolódó erőfeszítés költségeit viszont teljes egészében viseli, ezért nem áll érdekében az optimális mértékű és minőségű adatszolgáltatás, általában érdemes nagyobb mennyiséget az optimálisnál alacsonyabb minőségen előállítania.¹³

Az adatszolgáltató szervezetek részben vagy egészében finanszírozzák saját tevékenységüket. Az intézmények túlköltekezésének megakadályozására az állam sokszor érvényesíti ezt az elvet (sokszor csak implicit követelményként vagy támogatásmegvonással). Ha veszteségeiket a szolgáltatásukért megszabott árból kell fedezniük, akkor annak szükségképpen meg kell haladnia annak határköltségét, vagyis hatékonyságvesztést okoz.

Az állam ugyanakkor szabályozhatja az egyes adatszolgáltatások árait, és azokat olyan szinten határozhatja meg, amelyek mellett a szervezet éppen nullszaldós. A mai magyar gyakorlat a részleges szabályozás: az adatszolgáltató szervezetek alapszolgáltatására (a közhasznú adatokra) ingyenességet állapít meg, ugyanakkor lehetséges biztosít arra, hogy a szervezetek „értéknövelt szolgáltatásokat” nyújtsanak, és azokat szabadon meghatáro-

¹² Az optimális árak kérdése módszertanilag ekvivalens az optimális hozzáférés kérdésének a vizsgálatával. Ha az adatszolgáltatás ára nulla, akkor ez szabad hozzáférést jelent az adatkerők számára, az ár növekedésével pedig a biztosított hozzáférés folyamatosan csökken.

¹³ Például egy szervezet a birtokában lévő összes adatot elérhetővé teszi a honlapján, de csak aggregált formában, az alaposabb elemzésekre alkalmas mikroadatokat pedig nem adja ki.

zott áron érvényesítsék. Az alapszolgáltatás minősége ugyanakkor a legtöbb esetben nincs pontosan definiálva, így a szervezeteknek érdekükben áll az alapszolgáltatásként is megkövetelhető minőséget értéknöveltként feltüntetni és beárázni, valamint egy értékcsökkentett változatot ingyenesen nyilvánosságra hozni.

Vizsgáljuk meg most azt az esetet, amikor az adatszolgáltató fix összeget szabhat meg (úgynevezett licenc- vagy előfizetési díjat), és csak azoknak a személyeknek szolgáltat egy meghatározott egységáron, akik előfizettek a szolgáltatásra! Az egységárat meghatározhatjuk úgy, hogy az pontosan fedezze a változó költségeket, a belépési díjat pedig úgy, hogy azokból összességében megtérüljenek a fix költségek, vagyis ez a tarifarendszer lényegében a költségek szerkezetét másolja le.¹⁴ Ennek az optimálisan választott úgynevezett kétrészes tarifának a két különböző eszköze alkalmas lehet arra, hogy külön-külön teljesítsék a két célt: az egységárat egyenlővé tehetjük a határköltséggel, így kialakulhat a társadalmilag optimális szolgáltatási szint, míg a belépési díjakat úgy kell meghatározni, hogy a szolgáltatás éppen ne legyen veszteséges.

Az adatszolgáltatás ára nem lehet diszkriminatív. Az adatvédelmi törvény egyértelműen kimondja, hogy az adatszolgáltatásért kért költségtérítés összege nem tehető függővé a közérdekű adat felhasználásának céljától. Ha a korábban áttekintett első legjobb, határköltségen alapuló árazás egységesen teret nyerne, akkor az árdiszkrimináció valóban jólétszökkentő hatású lenne. Ha viszont el kell térnünk a társadalmilag optimális árazástól (például finanszírozási nehézségek miatt), akkor az egységes ár alkalmazása többé már nem hatékony, a diszkrimináció engedélyezésével növelhető a jólét.¹⁵

Ha minden adatszolgáltatás ára szabályozott, akkor mindenképpen optimális eltérni az átlagköltségen alapuló árazástól, és a számos iparágban (például gáz- és energiaszolgáltatás) alkalmazott Ramsey-árazás alapgondolatát alkalmazni: ahhoz, hogy az ártorzítás okozta hatékonyságvesztést minimalizáljuk, a magasabb árugalmasságú piacon kisebb mértékben kell eltérni a határköltségtől, mint az alacsonyabb árugalmasságú piacon. Már pusztán ennek az elvnek az alapján megalapozottnak tűnik, hogy a kutatási célokra kért adatokat optimális alacsonyabb áron értékesíteni (vagyis szélesebb hozzáférést biztosítani hozzájuk), mint az üzleti célokra kért adatokat, mivel a kutatási feladatokat ellátó felhasználók fizetőképessége jellegzetesen kisebb, ami magasabb árugalmassággal párosul.

Ehhez hasonló árazási stratégiával ma is találkozunk a közérdekű adatok piacán, de ez mégsem jelenti azt, hogy máris megvalósult a hatékony árazás. Az árdiszkrimináció már bemutatott iránya ugyanis egybevág azzal, ami akkor következne be, ha az adatszolgáltatónak megengednénk, hogy profitját maximalizálja: ő is a kevésbé árérzékeny felhasználói csoport számára szabna nagyobb árat, de mindkét részpiacon tovább emelné az árat.¹⁶ Pusztán abból a tényből tehát, hogy diszkriminatív árazást figyelünk meg, és az árak a határköltségnél magasabbak, még nem tudjuk elkülöníteni, hogy az a második legjobb társadalmi optimumhoz vagy egy profitmaximalizáló szolgáltatóhoz – jelen esetben az adatvagyonnal a szabályozás szándékával ellentétesen monopolistaként gazdálkodó adatgazdához – tartozik. A kérdés eldöntéséhez meg kell figyelnünk a szolgáltató bevételeit,

¹⁴ Az érintett intézmények jelenlegi működését ismerve, meg kell jegyeznünk, hogy a költségek becslésekor rendkívül körültekintően kell eljárni, hiszen az intézmények elemi érdeke azok eltűzése. Éppen ezért a becslést független szakértőknek kell végezniük, az nem lehet alkudozás tárgya. Ezzel párhuzamosan meg kell vizsgálni, miként pótolható az adatközléssel kapcsolatban felmerülő, az intézmények bevételeiből valóban hiányzó összeg.

¹⁵ Ez az állítás „a második legjobb tételéből” is következik: ha már érvényben van egy hatékonyságtorzító elem, akkor még egy önmagában hatékonyságtorzító elem bevezetése növelheti a jólétet.

¹⁶ A szabályozási szakirodalom ezt úgy nevezi, hogy az árak struktúrája optimális, csak a szintjük nem.

amelyeket össze kell vetni a kiadásaival. Ha úgy látjuk, hogy az adott szolgáltató „túlzóan” profitorientált diszkriminatív árpolitikát folytat, arra következtethetünk, hogy az indokoltan tekintett profitszint eléréséig arányosan kell csökkenteni a különböző fogyasztói csoportoknak szabott árakat.

Az árdiszkrimináció megengedésével azonban még akkor is növekedhet a jólét, ha az adatszolgáltató teljesen szabad kezet kap árai meghatározásában, bár ehhez további feltételeknek kell teljesülniük.¹⁷ Hasonló az eredmény abban az esetben, amikor ugyan a harmadfokú árdiszkrimináció nem lehetséges, de a szolgáltatók különböző minőségű szolgáltatások révén másodfokú árdiszkriminációt folytathatnak, amely szintén jelenthet optimális elmozdulást az egységes árázashoz képest.

Statistikai adatkezelő szervezetek közötti adatszolgáltatásért cserébe nem számolható fel díjazás. Ezt az elvet a statisztikai törvény írja elő, és jogosultsága több közgazdasági elméletből is levezethető, de számos példát ismerünk, amikor mégis sérül.

Az elvet elsősorban alátámasztja az, hogy ha mindkét szervezetet az állam finanszírozza, akkor a közöttük létrejövő jóvedelemáramlás semmiképpen sem lehet hatékony. Másodsorban az is előfordulhat, hogy az adatkérő szervezet a kapott adatot valamilyen termelési folyamat inputjaként használja fel, és egy újabb terméket szolgáltat. Az első adatszolgáltató által meghatározott magasabb inputár ilyen módon szükségszerűen beépül a második szolgáltató által meghatározott árba. A végtermék magasabb ára csökkenti az annak piacán végbemenő tranzakciók nagyságát, és így a társadalmi jólétet is.

Hasonlóan fontos annak belátása, hogy ezeknek az adatkérő szervezeteknek a tevékenysége jelentős pozitív externáliát gyakorolhat a társadalom többi tagjára is, és minél nagyobb ez a pozitív externália, annál inkább indokolt az adott tevékenység folytatását támogatni, vagyis az árát csökkenteni. Meglátásunk szerint ezt az elvet érdemes érvényesíteni a kutatási célú adatfelhasználás esetén is, bár az nem, vagy csak részben állami finanszírozású.

A magyarországi gyakorlat példák tükrében

Noha számos nehézség és költség felmerül a kutatási adatok kezelése során, azok láthatóan megoldhatók; felhasználásuk révén pedig számottevő társadalmi haszon keletkezhet. Egy döntéshozó számára azonban elképzelhető, hogy mindez nem elégséges érv az adatok hozzáférhetővé tétele mellett, mivel ő nem feltétlenül van tudatában ezeknek a sokszor externáliaként megjelenő hasznoknak. Mivel gyakran csak az adatok felhasználásával járó kockázat a nyilvánvaló számára, ez arra ösztönözheti, hogy a hozzáférést túlzott szigorral korlátozza. Azért, hogy a helyzetet világosabbá tegyük, és segítsünk az ilyen helyzetet elkerülni, a következőkben példákat hozunk a kutatási adatok, azon belül pedig a számunkra ismerős empirikus közgazdasági kutatások alapjául szolgáló mikroadatok felhasználásával kapcsolatos Magyarországon tapasztalt gyakorlatra.

Hosszan tartó siker: az ÁFSZ bértarifa-felvétele. Az Állami Foglalkoztatási Szolgálat (és jogelődei) 1986 óta (1993 óta évente) bonyolítja le a magyar munkavállalók bérvizonyairól átfogó képet rajozó bértarifa-felvételt. Az adatfelvétel a versenyszféra több mint 100 ezer, valamint a közszféra sok százezer munkavállalójának egyéni béradatait, demográfiai adatait, illetve munkahelyének jellemző adatait tartalmazza. Létrehozásának célja

¹⁷ Ennek legfontosabb szükséges feltétele, hogy az árdiszkrimináció megengedése után a kielégített kereslet ne csökkenjen (megítélésünk szerint ez a csökkenés az adatszolgáltatás piacán fennáll).

alapvetően államigazgatási-statisztikai: ebből a felvételtől származnak a keresetek alakulására vonatkozó hivatalos adatok.

A felvételnek az a tulajdonsága azonban, hogy adatait az egyének szintjén gyűjti, az egyszerű leíró táblázason túl összetett statisztikai-ökonometriai módszerek alkalmazását is lehetővé teszi. Ez a lehetőség azonban önmagában csak technikai. Mivel a felvételt az Állami Foglalkoztatási Szolgálat semmilyen oknál fogva nem köteles kiszolgáltatni a nagyközönségnek, annak az eredeti célon túlmutató hasznosulását semmi sem garantálja. Az Állami Foglalkoztatási Szolgálat felelős munkatársai azonban úgy döntöttek, hogy kihasználják a számukra adott mozgásteret, és a tudományos kutatók számára hozzáférhetővé teszik az elemi adatokat. Ennek a nyitottságnak eredményeképpen mára a világ több országában született olyan közgazdasági elemzés, amely a bértarifa-felvétel adatait használja, amely közül több rangos szakfolyóiratban (például *Campos-Jolliffe* [2005]) és nagyra értékelt kötetben jelent meg (*Kertesi-Köllő* [2002], illetve *Köllő* [2000]).

Ezek az adatok nem „pusztán” tudományos célokat szolgálnak. Sok más területtől eltérően a kutatók olyan gyakorlati kérdésekhez is megalapozottan tudtak hozzájárulni, mint a diplomás túlképzés vélelme, a minimálbér emelésének hatása, az iskolázottság jelentőségének átalakulása vagy a munkaerő keresletének rugalmassága.

Kudarc: az APEH mérlegadatai. Szorosan a bértarifa-felvételhez kapcsolódik az APEH mérlegadatainak állománya, amelyről már említést tettünk. A bértarifa-felvétellel ellentétben a mérlegadatokat nem statisztikai felvétellel gyűjtik – azok úgynevezett adminisztratív adatbázis részei. Noha használatával igen fontos kutatásokat lehetne elvégezni, a mérlegadatokat az APEH nem adja ki kutatási célra, csak a statisztikai szolgálaton belül. Indoklásként rendszerint két érv hangzik el: 1. az APEH-nek nem feladata a kezelésében levő adatok kiadása külső érdeklő számára, 2. erre az adott esetben nincs is módja, mert az adatok kiadása üzleti titkot sértene. Mint azt korábban kifejtettük, egyik érv sem elfogadható az adatkérés visszautasítására.

Hosszan tartó siker: a KSH munkaerő-felvétele. A KSH munkaerő-felvételének története igen hasonló az ÁFSZ bértarifa-felvételéhez. A munkaerő-felvétel a hivatalos foglalkoztatási és munkanélküliségi statisztikák forrása. Információját lakossági adatgyűjtés útján szerzi, aminek eredménye minden év minden negyedévében mintegy 70 ezer egyedi megfigyelés, amely a teljes magyar népességet reprezentálja (noha az alapvető demográfiai adatokon túl csak a 15–74 évesekre vonatkozó kérdéseket tesznek fel). A felvétel 1992 első negyedévéától napjainkig érhető el.

A munkaerő-felvétel hasonló „karriert” futott be, mint a bértarifa-felvétel. Kezelői felismerték a benne rejlő tudományos lehetőségeket, és mind a fontosabb hazai kutatóműhelyek, mind a külföldi felhasználók rendelkezésére bocsátották. A két felvétel között van azonban két jelentős különbség. Egyrészt a munkaerő-felvétel adatlapjainak eredeti egyéni azonosítója a lakóhely viszonylag finom, de nem pontos azonosítását teszi lehetővé. Van, aki úgy érvel, hogy ezen azonosítók és az ismert egyéni jellemzők alapján az egyes emberek megkereshetők. Ennek elkerülésére az adatokat anonimizálni kell, ez esetben arra, hogy az egyes adatlapokról az aggodalomra okot adó területi információt el kell távolítani. Másrészt a munkaerő-felvétel lehetőséget ad arra, hogy az egyes emberekre vonatkozó információkat hat negyedéven át figyeljük, következtetéseket vonjunk le nemcsak a résztvevők aktuális helyzetéből, hanem annak változásából is. Ehhez azonban egyedi azonosítókra van szükség, amelyek az előbbi oknál fogva nem állnak rendelkezésre.

Szerencsére van egy egyszerű mód arra, hogy időben folytonos egyéni azonosítók álljanak rendelkezésre. A KSH munkatársai az adatok átadása előtt a valódi azonosítók

kritikus elemét lecserélik egy véletlenszerű azonosítóra, amely azonban egy adott személy esetében minden évben azonos. Ha a kutatónak mégis szüksége volna a számlálókörzeti információkra, a valódi és a generált azonosító közötti fordítókulcs birtokában a KSH-ban ez a kapcsolat pillanatok alatt megteremthető. A művelet elvégzése után az összefésült állományokat a kutató a korábban generált mesterséges, információit nem hordozó azonosítóval kapja meg, így adatvédelmi aggályok nem merülnek fel.

Kudarca: a népszámlálás mikroadatai. A népszámlálás Magyarország legdrágább adatfelvétele, a költségeit ezermillió forintokban kell mérni. A népszámlálás azonban nemcsak drága, de igen érzékeny adatfelvétel is. Mivel minden magyar állampolgárt megkeres, és az adatoknak meglehetősen széles körét érinti, elemi adatainak hozzáférése legalábbis aggálllyal töltheti el a döntéshozókat.

A jelenlegi helyzet azt mutatja, hogy az aggály feloldása nem volt sikeres. Az adatforrás gazdái úgy döntöttek, hogy a népszámlálás elemi adatait nem hozzák nyilvánosságra. Emiatt – szemben a bértarifa- vagy a munkaerő-felvétellel – a népszámlálás elemi adataira alapozott önálló tanulmány a KSH-n kívül dolgozó kutató tollából tudomásunk szerint még nem látott napvilágot. Az ezzel járó veszteséget megbecsülni lehetetlen.

Felmerülhet, hogy talán a jelentős adathozzáférési költségek indokolják a nyilvánosságra hozás megtagadását. A hozzáférés technikai jellegű költségeiről nem érdemes szót ejteni, hiszen azok nyilván minimálisak egy strukturált formában tárolt adatbázis esetében. Komoly költségeket jelenthetnek azonban az adatvédelmi intézkedések. Ha „csak” észszerű adatvédelmi igényeknek kellene megfelelni, akkor joggal feltételezhetjük, hogy a szükséges munka elvégezhető, hiszen számos ország – első helyen az Egyesült Államok felelős hivatala – az interneten teszi elérhetővé és letölthetővé a helyi népszámlálások egy- és ötszázalékos állományait (köztük a legfrissebbekét is).¹⁸ Több országban – a skandinávokkal az élen – nemcsak mintákkal, de magával a teljes népszámlálási adatbázissal is lehet dolgozni. Ezekben az esetekben a szóban forgó adatbázis nem hozzáférhető, elemzése csak az adatokat kezelő hivatal megfelelően felkészített és ellenőrzött számítógépes munkahelyén (kutatószobában) lehetséges. A kutatók beiktatását alaposan kidolgozott, nevesített büntetőjogi háttér és adatkezelési protokoll támogatja.

Kérdésünkre, hogy a magyar népszámlálás elemi adatai (tudományos kutatás céljára) hozzáférhetők-e, nemleges volt a válasz. Az 1990-es népszámlálásból készült ugyan egy, az adatait a településeknél finomabb bontásban bemutató CD-kiadvány, az mára már nem kapható. A 2001-es népszámlálásból már nem is készült ilyen anyag, noha az adatok igen speciális, táblázatos formában megtalálhatók a KSH honlapján. Sajnálatos módon ezek szerkezete fejlett statisztikai eszközökkel végzett kutatásra teljesen alkalmatlan.

Noha a KSH-ban korszerű számítástechnikai apparátus működik, az elemi adatokhoz kontrollált körülmények között hozzáférést biztosító kutatószoba kialakítására mind ez idáig nem került sor. Személyes beszélgetések során azonban felmerült, hogy a KSH munkatársai külföldi tapasztalatcsere formájában éppen most vizsgálják ennek elméleti lehetőségét, és ismét csak informálisan olyan tájékoztatást kaptunk, hogy egy kutatószoba üzembe állítása akár már 2006 telén elképzelhető. Ha ez csakugyan megtörténik – és ami hasonlóan fontos: a külföldön megszokott széles körű jogokkal és kötelezettségekkel –, akkor a KSH igen fontos lépést tesz meg adatainak hasznosítása irányába.

További érdeklődésünkre, amelyben egy kisebb egyéni minta használata felől érdeklődtünk, meglepő választ kaptunk: a KSH munkatársainak sikerült a népszámlálásból származó minta elemzésével járó személyiségi jogi problémát megoldaniuk, és elkészült

¹⁸ Az Egyesült Államok esetében az adatok a <http://www.census.gov/main/www/pums.html> oldalról tölthetők le.

a népszámlálásnak egy ötszázalékos mintája, amelyben megoldották az adatvédelmi problémákat. A KSH népszámlálási osztálya 2000 után szerződést kötött a Minnesotai Egyetem mellett működő *Minnesota Population Centerrel* (minnesotai népesedéstudományi központtal). Ebben vállalja, hogy annak igen előremutató *IPUMS International* elnevezésű programjához, amely a világ számos országából gyűjt elemi népszámlálási adatokat, rendelkezésre bocsátja a magyarországi állományt.¹⁹ A KSH munkatársai elmondták, hogy a Minnesotába szállított állományt alaposan áttekintették, megtisztították, adatvédelmi szempontból megfelelő állapotba hozták.

A fentiek akár forradalmi lépésként is értelmezhetők. A KSH egy hagyományosan féltve őrzött, eredetileg egyedi adatot fosztott meg egyediségétől, így mindazokat a tartalmi akadályokat feloldotta, amelyek akadályozták a további felhasználást. Bár a Minnesotai Egyetem számára sikerült előállítani megfelelően anonimizált állományt, amely regisztráció után az amerikai intézetből elérhető lesz a kutatók számára, az jelenleg nemcsak az interneten keresztül, de tudomásunk szerint a magyarországi kutató közösség számára más módon sem érhető el.

Mint már hangsúlyoztuk, a népszámlálás esete is csupán példa, így a tapasztalatokból ismét nem az egyedi, hanem inkább az általános tanulság levonása gyümölcsöző. Joggal merül fel azonban a költői kérdés: megfelelő-e az a szabályozás, amely lehetővé teszi, hogy az adófizetők pénzének felhasználásával készült legköltségesebb adatfelvétel ne hasznosuljon teljes egészében akkor, amikor ennek láthatóan adott minden technikai feltétele?

Áttörés és a siker ígérete: adminisztratív adatkérés az OEP-től és az ONYF-tól. Végül beszámolunk egy még folyamatban levő, de igen előremutató, a hazai adathozzáférés történetében úttörő kezdeményezésről. 2005-ben a Pénzügyminisztérium közgazdasági kutató osztálya felvette a kapcsolatot az Országos Egészségbiztosítási Pénztárral, az Országos Nyugdíjbiztosítási Főigazgatósággal, az APEH-hel, valamint a Magyar Államkincstárral egy tervezett kutatás megvalósítása érdekében. A kutatásnak szüksége van olyan egyéni szintű adatbázis összeállítására, amelyben nemcsak a két intézmény által tárolt egyéni adatok találhatók meg, de azok között kapcsolat is van. Ennek segítségével egy rendkívül fontos kérdésre lehet választ kapni, amelyről ma még csak igen durva vonalakban vannak ismereteink: miként alakul a társadalom egyes – akár igen finoman meghatározott – csoportjai és az állam közötti pénzáramlás egyenlege. Más szóval: ha a kutatás sikerrel zárul, megállapítható lesz, hogy mely társadalmi csoportok részesülnek és milyen mértékű nettó pozitív vagy negatív transzferben.

A jogi szabályozás tükrében az adatkérőnek három problémát is le kell küzdenie. Az első, hogy az adatok eredeti formájukban személyesek: a három hivatal a birtokában levő azonosítók segítségével az adatokat egyértelműen összefüggésbe tudja hozni az érintett személyekkel. Ugyanakkor az anonimizálás ebben az esetben igen egyszerű: el kell hagyni róluk az azonosítót, valamint a földrajzi azonosításra alkalmas irányítószámot egy jóval durvább, területi szintű azonosítóval kell helyettesíteni. E műveleteket elvégezve, adatvédelmi aggály nem merülhet fel.

A második, az adatok összekapcsolásával felmerülő nehézség talán a legkényesebb, megoldása pedig példaértékű lenne, ha az abban kulcsszerepet játszó BM Központi Adatfeldolgozó, Nyilvántartó és Választási Hivatal együtt tudott volna működni a kutatókkal. Megfelelő, a kért lépésre kimondottan felhatalmazó jogszabály hiányában erre sajnos nem kerülhetett sor. Ez az eset megmutatja, miként lehetne a jelenlegi jogszabályi keretet (az említett kivételével) maradéktalanul figyelembe véve haladó lépéseket tenni. A prob-

¹⁹ A projekt honlapja a <http://www.ipums.umn.edu/international/index.html> címen érhető el.

léma alapja az, hogy az OEP és az ONYF adatait taj-szám, az APEH adatait pedig az adószám kapcsolja egyes személyekhez. Az adatok összekapcsolásának legegyszerűbb módja az lenne, hogy a BM adatnyilvántartója, ahol megvan az azonosítók összekapcsolásához szükséges kulcs, átadja azt az érintett intézményeknek. A személyi számot megszüntető törvény célja azonban éppen az, hogy adminisztratív adatokat az azokat kezelő intézmények saját hatáskörben ne kapcsolhassanak össze. Mindez azt sugallja, hogy a kitűzött cél elérhetetlen.

A kutatók igényét egy adatátadási protokoll kialakításával lehetett mégis kiszolgálni. E szerint elsőként az egyik adatgazda (például az OEP) mintát vesz a taj-azonosítókból, amelyekhez véletlen számokból más célra nem használható azonosítót generál. Saját adatait a taj-szám nélkül, de a generált azonosítóval elküldi a kutatóknak. A mintát azonosító taj-szám és generált azonosító párokat elküldi a BM adatnyilvántartó hivatalába, ahol azt összekapcsolják az egyéni adószámokkal, majd csak az adószámokat és a generált azonosítót továbbítják az APEH-hez. Figyeljük meg, hogy az azonosítók összekapcsolása csak a nyilvántartónál valósul meg, akinek erre törvény adta joga van. Az APEH az adószámok alapján leválogatja a kért adatokat, majd azokat csak a generált azonosítókkal elküldi a kutatóknak. Végül három adatminta áll a kutatók rendelkezésére ugyanazzal a véletlen, de a minták azonosítására alkalmas azonosítóval. Az egyes adatminták anonimizáltak, és ebben az esetben (ez nem feltétlenül teljesül mindig) az összekapcsolt adatbázis is az lesz. Így valósulhat meg az, hogy az eredetileg személyes adatok összekapcsolásra alkalmasan eljutnak a kutatókhoz anélkül, hogy a részt vevő hivataloknál egyszer is összekapcsolódnának.

Bár a BM adatnyilvántartó hivatalának szűkre szabott hatásköre miatt az eredeti elképzelés nem valósulhatott meg, az OEP és az ONYF, illetve az APEH és a Magyar Államkincstár egymás között a fentihez hasonló protokollt alkalmazva, létre tudja hozni a megfelelő adatbázisokat. Ezek azonban egymással nem kapcsolhatók össze, így a kutatás eredeti célját nem érheti el. Noha ez is a jelenleginél jóval több információt fog szolgáltatni, a veszteség jelentős.

A végső nehézség az, hogy a fenti terv működőképességéről meg kell győzni az adatok kezelőit. A jogi részleteket taglaló fejezetben láttuk, hogy hiába egyszerű az anonimizálás, hiába tartja tiszteletben a protokoll a jogszabályokat, az adatgazdának a jog által kezébe adott rugalmasságot aktívan használnia kell ahhoz, hogy az adatokat kiadhassa. A terv sikeres végrehajtása azt mutatja, hogy mind az említett négy szervezet illetékesei felmérték a kutatás jelentőségét, és együttműködik a szokatlan, de ígéretes munkát kezdeményező kutatókkal.

Összefoglaló következtetések

Tanulmányunk utolsó részében az állami finanszírozású adatvagyon kezelésére sikeres és kevésbé sikeres példákat is bemutattunk. Láttunk olyan eseteket, amelyekben az adatokról döntést hozó személy maga kezdeményezte az együttműködést, olyat, ahol ez a kutatók megkeresésére jött létre, illetve olyat is, ahol nem sikerült megállapodást kötni. Minden esetre érvényes tanulság azonban az, hogy mivel a kutatási adatok hozzáférése nem szabályozott, a döntéseket helyben meghozó személyek szerepe nagyon jelentős. Ezért a kedvező gyakorlatok is elromolhatnak, a kedvezőtlenek is megjavulhatnak.

A törvények szövegét olvasva és a közgazdasági érvelést végiggondolva, sokakban megfogalmazódhat a gondolat: ha vitathatatlanul hasznos a kutatási adatokhoz való hozzáférés, és a törvény látszólag erős eszközöket ad az adatot igénylők kezébe, miért mennek ennyire nehezen a dolgok? Amint azt több ízben megjegyeztük, a kutatási adatok használatának költségei kézzel foghatók, azonnal és helyben jelentkeznek. A keletkező hasznok

ezzel szemben közvetettek, és rendszerint nem a költséget viselő intézményeknél jelentkeznek. Ez egyedül azonban nem vezetne a mára kialakult helyzethez. A szervezeti felaprózottság, a hiányzó felügyeleti és koordináló intézményrendszer, a felhasználás támogatásának közös célját nélkülöző és ilyen szempontból hiányos jogszabályok azonban együtt olyan helyzetet teremtenek, amelyben az adatkezelő intézmények az adatnyilvánosság törvényalkotási szellemének ellenében, saját érdekeiket követve cselekedhetnek.

Az előzőekben kifejtett megfontolások alapján azt gondoljuk, hogy az adatok hozzáféréseinek körülményeit javítani lehet és szükséges. Ehhez először is fontos lenne, hogy azt a gondolatot, amely szerint az állami költségen létrehozott adatvagyonat érdemes a kutatók rendelkezésére bocsátani, a kormányzat érezze magáénak, és ezt aktívan hirdesse is. Hangsúlyozni kell, hogy az adatok gyűjtése, rendszerezése és elemző feldolgozása mind fontos részei ennek a folyamatnak, ugyanakkor szakítani kellene azzal a gondolattal, hogy minden munkafázist egy intézménynek kell elvégeznie.

A jelenlegi törvényi szabályozás alapgondolati tiszták, gyakorlati megvalósításukhoz azonban további kiegészítések szükségesek, amelyek egyértelművé teszik a külső felhasználók adathozzáféréseinek módját és eszközeit. Elsősorban azt az állapotot kell elérni, hogy minden adminisztratív szervezeteknek kötelessége legyen a birtokában levő adatok kezelésével aktívan foglalkozni. Bár ez törvényi kötelesség, e kötelezettség kiszabását külön nem tartalmazza az egyes szervezetek tevékenységét közvetlenül szabályozó joganyag. Talán éppen ezért találkozhatunk számos esetben azzal a kibúvóval, hogy az adatok kezelése nem tartozik a hivatalos tevékenységi körbe, és nincs rá erőforrás. Továbbá a jogrendszer minden szintjén meg kell születnie azoknak a konzisztens törvénymódosításoknak, amelyek lehetővé teszik az adatok hozzáférését.²⁰ Optimális esetben az adatkezelő szervezeteket kötelezni lehetne arra, hogy ha a birtokukban levő személyes adatok anonimizálhatók, vagyis közérdekűvé tehetők, akkor ezt meg is tegyék. Az adatvédelem szempontjából ugyanakkor indokoltnak érezzük, hogy jöjjön létre olyan jogszabály, amely lehetővé teszi az érzékeny adatokkal dolgozó kutatók személyes felelősségvállalását.

Sok intézmény a jelenlegi helyzetet alapul véve bevételi forrásokat épített adatvagyonára, nekik a fentiekben javasolt nyitás bevételekiesést fog okozni. Ennek ellensúlyozására, de elsősorban az optimális helyzet elérése érdekében javasoljuk, hogy az intézményeknek legyen lehetőségük a felhasználás célja alapján meghatározni az adatok kiadásának módját,²¹ illetve ennek alapján diszkriminálni az adatok kiadásának árazásában.

Az intézmények működésének, árazásának szabályozását és felügyeletét egy független és állandó intézmény hatáskörébe célszerű utalni. Noha időről időre létrejönnek olyan *ad hoc* szervezetek, mint például az Informatikai Tárcaközi Bizottság Adatgazdálkodási Szakmai Bizottsága, ezek mindeddig nem kaptak megfelelő jogkört és átfogó feladatot ahhoz, hogy kellő hatásfokkal működhessenek.

Hivatkozások

- CAMPOS, N.–JOLLIFFE, D. [2005], Does Market Liberalisation Reduce Gender Discrimination? Econometric Evidence from Hungary 1986 to 1998. *Labour Economics*, Vol.12. No. 1. 1–22. o.
- CSERES–GERGELY ZSOMBOR–CSORBA GERGELY [2006]: Közgazdasági gondolatok az állami intézmények által kezelt adatokhoz való hozzáféréseiről. Megjelent: Hozzáférés a közszféra adataihoz. Eötvös Károly Intézet, Budapest.

²⁰ Egy ilyen átfogó módosításcsomag tervezete a közelmúltban el is készült az akkori Ifjúsági Családügyi Szociális és Egészségügyi Minisztérium megbízásából.

²¹ Lásd a már említett kutatószoba számos országban működő intézményét, amikor a kutató csak ellenőrzött körülmények között fér hozzá az adatokhoz.

- KERTESI GÁBOR-KÖLLŐ JÁNOS [2002]: Economic Transformation and the Returns to Human Capital. Megjelent: *De Grip, A. – Van Loo, J. – Mayhew, K.* (szerk.): The economics of skills obsolescence. Research in Labor Economics, Vol. 23, Elsevier.
- KÖLLŐ JÁNOS [2000]: Transformation before the transition, Megjelent: *Maskin, E. – Simonovits, A.* (szerk.): Planning, Shortage and Transformation. Essays in honour of János Kornai, MIT Press, Cambridge Mass.
- MAJTÉNYI LÁSZLÓ-SZABÓ MÁTÉ [2006]: A közszférában kezelt adatok típusai a magyar jogrend-szerben. Jogi-fogalmi elemzés. Megjelent: Hozzáférs a közszféra adataihoz. Eötvös Károly Intézet, Budapest.