

BŐGEL GYÖRGY

**Az adatrobbanás mint közgazdasági jelenség**

Az adatrobbanás, vagyis a korábbiakat nagyságrendekkel meghaladó méretű adattömegek, adatbázisok megjelenése a tudományos, a gazdasági és a társadalmi élet számos területén korunk gazdaságtudományi szempontból is figyelemre méltó jelensége. Cikkünkben a dimenziók érzékeltetése után bemutatjuk a jelenség technológiai hátterét, elemezzük fontos gazdasági vonatkozásait, felvázoljuk a körülötte kialakuló, sokféle szereplőből álló adat-ökoszisztémát. Felhívjuk a figyelmet arra, hogy a nagy adatbázisok kezelése és hasznosítása fontos versenyképességi tényezővé vált vállalati és nemzetgazdasági szinten egyaránt, a „nemzeti adatvagyon” fogalmát ezért tágan kell értelmezni, és ki kell dolgozni az azzal kapcsolatos stratégiákat és politikákat. A fejlődés nem mentes a kockázatoktól és a dilemmáktól – a cikk végén ezekről is képet adunk.\*

Journal of Economic Literature (JEL) kód: A12, B40, C02, D83, L26, L52, L86, M15, O31

A világ egyik vezető természettudományos folyóirata, a Nature 2008. szeptember 4-i számát a *big data* témának szentelte. A szerkesztők a bevezetőben felhívták a figyelmet arra, hogy a kutatóknak és intézményeiknek világszerte hatalmas adattömegek kezelésére és hasznosítására kell felkészülniük. A természettudományos kutatás nyilván mindig adatigényes volt, most azonban nagyságrendi változásról van szó, soha nem látott adattömegekről – ezeket jelöli a *big data* elnevezés.

A nem kevésbé tekintélyes Science 2011. február 11-én jelentkezett hasonló témájú tematikus számmal. A kérdést különböző tudományterületek szempontjából tárgyaló cikkek bevezetőjében különleges tudományos kihívásnak nevezik a hatalmas adattömegek megjelenését, amiben különleges lehetőségek rejlenek a kutatók számára. Ezek kihasználása érdekében, a technikai problémák megoldása mellett, a kutatás kultúrájának is meg kell változnia – állapítják meg a szerkesztők.

Mindkét folyóirat szerkesztői és szerzői egyetértenek abban, hogy a jelenség radikális változásokat hoz a tudományos kutatásban, annak szinte minden ágában, legyen szó akár természet-, akár társadalomtudományokról.

E két tematikus szám jó néhány példát hoz az elmondottakra olyan területekről, mint például a biológia, a genetika, az orvostudomány vagy a magfizika. Egyes írások a kutatás mellett az eredmények hasznosítására is kitérnek, és azt állítják, hogy ott is szembe kell nézni az adatrobbanás jelenségével.

Ebben a cikkben azt szeretnénk bemutatni, hogy e jelenség közgazdasági szempontból is nagyon fontos. A hatalmas adattömegek megjelenése a közgazdaságtudomány számára

\* A szerző köszönetet mond az IQSYS cég üzleti intelligenciával foglalkozó csapatának: e cikk egyes gondolatait írásaik és előadásaik inspirálták.

ugyanolyan kihívást és lehetőséget jelent, mint a természettudományok területén, mind a kutatást, mind a hasznosítást tekintve. A cikk elején magát a jelenséget igyekszünk leírni, megvilágítva annak technikai hátterét is. Ezután példákon keresztül a fontosabb közgazdasági jellemzőit mutatjuk be. Hasonlóképpen példák segítségével tárgyaljuk a gazdasági hasznosítás módjait és az adatrobbanás jelenségre épülő tevékenységi köröket és piaci szerepeket. Végül külön szakaszt szánunk a kockázatok elemzésének.

Magyarországon 2010-ben törvény született<sup>1</sup> a nemzeti adatvagyon védelméről. Ennek első paragrafusában a következő definíciót használja a nemzeti adatvagyonra: „a közfeladatot ellátó szervek által kezelt közérdekű adatok, személyes adatok és közérdekből nyilvános adatok összessége”. Ebben a cikkben azt igyekszünk bizonyítani, hogy a nemzeti adatvagyon ennél jóval tágabban is értelmezhető, legalábbis közgazdasági szempontból.

### Az adatrobbanás jelensége és technológiai háttere

Mit értünk az adatrobbanás, a hatalmas adattömeg (*big data*) megjelenésének jelenségén? Mint látni fogjuk, sokfelé találkozhatunk vele (akár a hétköznapi életben is), rendszeresen tapasztaljuk a létezését, még sincs pontos definíciója, ilyet a két említett lap sem ad.

A *big* tulajdonképpen azt jelenti, hogy nagyon sok adatból álló adatbázis vagy adatkészlet. Olyan sokból, hogy azt már nehéz kezelni a megszokott, általánosan rendelkezésre álló adatbázis-kezelő alkalmazásokkal. Óriási adattömeg, amelynek kezelése (létrehozása, tárolása, feldolgozása, továbbítása, lekérdezése stb.) a technikai lehetőségek határát feszegeti. Mivel a technika állandóan fejlődik, ez a határ folyamatosan előre tolódik, tehát értelmezhető, de folyamatosan mozgó célponttal van dolgunk.

Hol húzódik ez a bizonyos határvonal manapság? Erre a kérdésre többféle módon válaszolhatunk. A Science említett számának szerkesztői gyorsfelmérést készítettek a cikkek értékelésében részt vevő kutatók körében. Megállapították, hogy a válaszadók 20 százaléka rendszeresen használ, illetve elemez 100 gigabájt<sup>2</sup> (1 gigabájt =  $10^9$  bájt), 7 százaléka pedig egy terabájt<sup>2</sup> ( $10^{12}$  bájt) nagyobb méretű adatbázisokat (King [2011] 692–693. o.). Ilyen méretű adatbázisok kezelése már komoly problémákat okozhat, a tényleges technikai határvonal azonban ennél magasabban húzható meg, az exa-, illetve zettabájtok ( $10^{18}$ , illetve  $10^{21}$  bájt) nagyságrendjénél. Kétségtelen, hogy sok olyan tudományos program van (például genetikai, meteorológiai, részecskefizikai, hálózatkutatói területeken), amelyek évente több terabájt adatot produkálnak.

A méretezés kérdését megközelíthetjük más irányból is. Martin Hilbert és Priscila López közelmúltban végzett számításai szerint az emberiség 2007-ben  $2,9 \times 10^{20}$  bájtnyi adatot volt képes tárolni optimálisan tömörített formában,  $2 \times 10^{21}$  bájtnyi tudott közvetíteni, és másodpercenként  $6,4 \times 10^{18}$  utasítást tudott végrehajtani általános célú számítógépeken. E gépek kapacitása 1986 és 2007 között évente átlagosan 58 százalékkal növekedett; ugyanez a mutató a kétirányú távközlés kapacitására vonatkozóan évi 28 százalék, a tárolt információ mennyiségére pedig 23 százalék. Az új évszázad eleje óta a digitálisan tárolt információ mennyisége meghaladja az analóg módon (például nyomtatott könyvekben) rögzítettekét, 2007-ben a tárolt információ 94 százaléka digitális volt (Hilbert–López [2011] 60. o.).

<sup>1</sup> 2010. évi CLVII. törvény a nemzeti adatvagyon körébe tartozó állami nyilvántartások fokozottabb védelméről.

<sup>2</sup> Egy mai dvd-lemezre négy-öt gigabájtnyi adat fér rá. Az amerikai kongresszusi könyvtár (*Library of Congress*) honlapján található adat szerint az intézményben 2011-ben 235 terabájtnyi adatot tároltak.

Érdemes egy pillantást vetni az adatok jellegét mutató információkra is. Az infokommunikációs piac elemzésével foglalkozó, tekintélyes IDC cég *Digital Universe Study* elnevezésű rendszeres felmérése szerint 2009-ben a digitálisan tárolt információ 56,7 százaléka jelent meg filmekben, a televízióban és reklámokban, 31,5 százaléka kamerák és videofelvételek születte volt, 2,8 százaléka zene, 1,5 százaléka a kommunikációhoz kapcsolódott, 7,5 százaléka pedig az egyéb kategóriába volt sorolható (idetartozik például a világháló is). Ebből a felsorolásból (idézi *Waters* [2011] 7. o.) az is látható, hogy a keletkező és tárolt információ nagy része – az IDC szerint több mint 90 százaléka – strukturálatlan.

Becslések szerint a tárolt adattömeg körülbelül 18 hónaponként megduplázódik.

Ezek a számok mindenképpen óriásiak, a nem szakemberek számára gyakorlatilag felfoghatatlanok, ugyanakkor nagy bennük a bizonytalanság, amit az is fokoz, hogy a források hol „adatot”, hol „információt” emlegetnek, miközben a kettő nem ugyanaz: az általános felfogás szerint az információ kontextusba helyezett adat. *Gleick* [2011] alapos történeti áttekintést ad az információval kapcsolatos nézeteink, ismereteink és fogalmaink fejlődéséről, így egyebek között arról is, milyen javaslatok születtek az információ mérésére. A fenti adatok az információ mennyiségét modern felfogás szerint bitekben (1 bájt = 8 bit) mutatják. Megfelelő statisztikák birtokában elég jó becslést lehet készíteni például arról, hogy a rendelkezésünkre álló digitális tárolóeszközöknek hány bit lehet a teljes kapacitása. Arról viszont nagyon nehéz képet alkotni, hogy ezeken az eszközökön mennyi „új” vagy „eredeti” információ lehet, hiszen ugyanarról a filmről, képről, szövegről, adatbázisról stb. számtalan másolat létezhet különböző helyeken.

A big data jelenségén nem csak az adatok mennyiségének növekedése, a mennyiségi értelemben vett adatrobbanás értendő. Az infokommunikációs piac elemzésével foglalkozó, kiemelkedő szakmai tekintélynek örvendő Gartner Group a nagy adatbázisok három fontos tulajdonságát emeli ki. Az első ezek közül az adatok mennyisége (*volume*) – erre vonatkozóan már közöltünk adatokat. A második az adatok változatossága (*variety*), az adatfajták és források sokfélesége. A sejt kutatás területén például nagyon sokféle és egymással bonyolult kölcsönhatásban lévő adatot kell összegyűjteni és feldolgozni. A harmadik tulajdonság a sebesség (*velocity*), ami egyrészt az adatok keletkezésének gyorsaságára vonatkozik, másrészt a hasznosításhoz rendelkezésre álló időre.

Mindezek alapján a jelenséget a következőképpen írhatjuk le: hatalmas, nagyfokú változatossággal és komplexitással jellemezhető, gyorsan keletkező és szaporodó adattömegek megjelenése, amelynek hasznosítására kevés idő áll rendelkezésre.

Láthattuk, hogy az emberiség a modern digitális technológiáknak köszönhetően óriási kapacitásokkal rendelkezik információk feldolgozásához és tárolásához. Ezekre szükség is van, hiszen a világban ugyanilyen óriási tömegű adat keletkezik szinte folyamatosan. Egyre több olyan forrás van, amely a hatalmat adattömeg jelenségét produkálja. Lássunk néhány példát!

Korunk egyik leglátványosabban fejlődő tudományága a genetika. Az emberi genom bájtokban kifejezett méretéről különböző forrásokban eltérő számokat lehet találni, de ha abból indulunk ki, hogy benne minden bázispár (*base pair*) két bittel kódolható, akkor körülbelül 725 megabájt adatról van szó. A CERN nagy hadronütköztetőjének használatánál évente 10–15 petabájt ( $10^{15}$  bájt) adat biztonságos tárolására van szükség. Óriási adattömeget ontanak folyamatosan a csillagászati teleszkópok. Szeizmográfiai, meteorológiai és környezetvédelmi mérések során is hatalmas tömegű adat keletkezik.

A technológiai fejlődés egyre fejlettebb „adattermelő” eszközöket hoz magával. Számos területen használnak például apró szenzorokat, amelyek a környezetükben képesek egy vagy több dolgot (például hőmérsékletet, rázkódást) érzékelni és az adatokat rádiójelek formájában valamilyen központba továbbítani. A közelmúltban a *The Economist* egy teljes

mellékletet szentelt az okos és intelligens rendszereknek (*smart systems*). E rendszerek működése nagyon hasonló logikára épül. Ki kell választani valamit, ami ésszerűsítésre, racionalizálásra szorul: ilyen lehet például az egészségügy valamelyik területe, valamilyen termelési folyamat, egy város közlekedése, valamilyen közműrendszer működése, egy termelési folyamat stb. Az említett szenzorok segítségével nagy tömegű adatot kell gyűjteni róla, ezeket fel kell dolgozni, a feldolgozásból új ismereteket nyerni, majd mindezek alapján beavatkozási módokat kidolgozni, és mindezt, ha lehetséges, folyamatosan és automatizáltan tenni. Az említett külföldi „okos városok” (*smart city*) példáját is említi,<sup>3</sup> ahol az így felépített rendszereket egyebek mellett hatékony közösségi szolgáltatások biztosítására használják (*Economist* [2010]).

Hasonló projektek sok régi városban is indulnak, és minden jel szerint nagy szükség van rájuk. Csak egyetlen példa: London 16 000 kilométer hosszúságú föld alatti vízhálózással rendelkezik, a csövek több mint fele száz évnél öregebb; a rendszerből becslések szerint mindennap közel 900 millió liter víz szivárog el, és naponta átlagosan 240 repedést, csőtörést kell rendbe hozni. Amikor a rendszert gondozó Thames Water közműcég új csöveket fektet le, rádiós szenzorokat kapcsol hozzájuk, hogy folyamatosan figyelni tudja az állapotukat, időben, lehetőleg preventív módon intézkedjen, minimalizálva a veszteségeket (uo. 6–7. o.).

Hatalmas adattömeggel dolgozó, intelligens, racionális és olcsó működést célzó rendszerek más közműveknél is megjelennek. A brit kormány például 2009 májusában bejelentette, hogy 2020-ig minden háztartást fel kíván szerelni gázzal, illetve elektromossággal működő készülékekhez kapcsolt intelligens mérőeszközökkel (*smart meters*). Ezek egyrészt a lakókat, másrészt a közműcéget tájékoztatják valós idejű adatokkal a fogyasztásról, információs házist biztosítva az energiával való racionális és takarékos gazdálkodáshoz, intelligens hálózatok (*smart grid*) működtetéséhez.

Az emberiség egyik legnagyobb problémája az egészségügyi szolgáltatások ésszerűsítése, hozzáférhetővé és fenntarthatóvá tétele. Kétség sem fér hozzá, hogy ez nem fog menni masszív informatikai támogatás nélkül. Egészségügyi digitalizálási programokra (például az orvosi diagnózisok eredményeinek digitalizálására) számtalan példa van. Rendkívül intenzív innovációs tevékenység tapasztalható például a távdiagnosztika területén; az új eszközök tulajdonképpen a leírt logika szerint működnek: szenzorok segítségével történő mérés, nagy tömegű valós idejű adat továbbítása az erre kijelölt központokba, riasztási és beavatkozási rendszer működtetése lehetőleg megelőzési céllal.

A hatalmas adattömeg megjelenésére kifejezetten hétköznapi példákat is említhetünk: folyamatosan keletkezik például a távközlési hálózatok forgalmáról (telefonhívások, e-mail stb.), de a források közé sorolhatjuk természetesen a földrajzi hely-meghatározó (GPS) rendszereket is, amelyek a közeli jövőben várhatóan az autók alapfelszereltségéhez tartoznak majd. Egy modern nagyvárosban kamerák milliói figyelik az emberek mozgását,<sup>4</sup> a legújabb rendszerek már az arcfelismerés területén is jó eredményeket tudnak felmutatni. Kimeríthetetlen, hólabdaszerűen növekvő adatforrások az internetes közösségi hálók: 2011 közepén a Facebooknak már 750 millió regisztrált felhasználója volt, a regisztrált és nem regisztrált felhasználók egyedül júniusban ezermilliárd (!) oldalt töltöttek le.<sup>5</sup> Az internet és a világháló minden felhasználója digitális nyomok tömegét hagyja maga után strukturált és/vagy strukturálatlan adatok (kattintások, letöltések, szövegek, képek, filmek stb.) formájában.

<sup>3</sup> Ilyen okos város például *Masdar* Abu-Dzabiban, *Songdo City* Dél-Koreában és *PlanIT Valley* Portugáliában.

<sup>4</sup> Az Egyesült Államokban 2005-ben már körülbelül 25 millió biztonsági kamerát üzemeltettek (tehát minden utcát polgárra egy kamera jutott), a számuk azóta minden bizonnyal tovább nőtt.

<sup>5</sup> A letöltési adatokat például a DoubleClick cég közli rendszeresen.

A hatalmas adatbázisok keletkeztetésének, tárolásának, továbbításának, rendszerezésének, kombinálásnak, feldolgozásának, lekérdezésének, megosztásának számtalan technikai feltétele sorolható fel. E jelenség technológiai hátterének részletes bemutatása és elemzése messze meghaladná e cikk kereteit. A felsorolt példák közül látható, hogy az adatoknak, illetve adatbázisoknak több szakaszból álló, a keletkezéstől az ismételt felhasználásig terjedő életciklusa van, amelyek mindegyikéhez megfelelő technikai hátteret és támogatást kell biztosítani. Erre a hagyományos asztali gépek, tömegesen használt szerverek technikai és gazdasági szempontból egyre kevésbé alkalmasak. Az adatok feldolgozása növekvő részben megosztott rendszereken, különleges, több futballpálya méretű, számtalan egymáshoz kapcsolt gépből (*computer grid*) álló adatközpontokban, az úgynevezett „számítástechnikai felhőkben” történik (lásd erről például *Doctorow* [2008], *Dömölki* (szerk.) [2008], *Bögel* [2009a]). A felhasználással arányosan fizetett, a felhasználóktól nagy indító beruházást nem igénylő internetes szolgáltatások, valamint a nyitott forráskódú szoftverek megjelenése és elterjedése lefelé szorítja a belépési korlátokat, olcsóbbá teszi az adatkezelési és feldolgozási technológiák használatát (*Cusumano* [2008]). Internetes szolgáltatások nyújtásához, valamint természetesen a hatalmas adatbázisok továbbításához megbízható szélessávú távközlési kapcsolatok kellenek.

A hardver természetesen nem elegendő: szoftverre is szükség van, amin nemcsak a számítógépes programokat, hanem az alkalmazott algoritmusokat, adatkezelési és -feldolgozási eljárásokat is értjük. A hatalmas adattömegek lehetővé teszik, hogy a tudományágak művelői az általuk feltett kérdésekre matematikai-statisztikai, adatbányászati eszközökkel keressenek választ, az adattömegekben mintázatok, sokszor nagyon bonyolult, sokváltozós összefüggések után kutassanak tanulásra is képes alkalmazásokkal. Fejlődnek a strukturálatlan adatok feldolgozására szolgáló technológiák, így például a szövegbányászat is (*Fajszí–Cser–Fehér* [2010] 12. fejezet).

A sokféle forrásból érkező, sokféle módon keletkeztetett adat feldolgozásához, kombinálásához és lekérdezéséhez természetesen szabványokra és szabványos metaadatokra is szükség van.

Sajátos eljárást és technológiát igényel az adatfeldolgozási műveletek eredményeinek tálalása, kutatók és döntéshozók számára felfoghatóvá és érthetővé tétele. Valós idejű rendszerekben a döntésekhez, a beavatkozáshoz sokszor nagyon kevés idő áll rendelkezésre. Nem véletlenül indult lendületes fejlődésnek például a nagy adattömegek kezelésére is alkalmas vizualizálási technológia és azon belül a háromdimenziós megjelenítés módszertana és eszközkészlete (lásd például *McCandless* [2010]). A megjelenítés fontossága és különleges szaktudásigénye miatt a vizualizálással foglalkozó szakembereket – akár képzőművészeket is – gyakran már a kezdet kezdetén, a tervezési fázisban bevonják a tudományos kutatási programokba (*Frankel–Reid* [2008]). A közérthetőség érdekében a vizualizálás területén is szükség van szabványokra, típusmegoldásokra. Mindezekből az is látható, hogy a jelenség kezelése mennyire multidiszciplináris csapatokat (például szakterületi kutatók + matematikusok + statisztikusok + informatikusok + grafikusok) követel.

A felsorolt és más, a jelenséghez kapcsolódó technológiák fejlesztésére európai uniós projektek is indultak Ezekben magyarországi kutatóközpontoknak is van szerepük. Példaként említhetjük például a NESSI-Hungary projektet<sup>6</sup> (az alapról lásd például *NESSI* [2009], *Reding* [2009]), vagy az MTA Rényi Alfréd Matematikai Kutatóintézetében működő, európai uniós támogatású bioinformatikai kutatócsoportot (*Palugyai* [2011]).

<sup>6</sup> A NESSI elnevezés a Networked European Software and Services Initiative rövidítése. A NESSI-Hungary (Szoftver és Szolgáltatások Nemzeti Technológiai Platform) honlapja: [www.nessi-hungary.com](http://www.nessi-hungary.com).



## A hatalmas adattömeg jelensége a gazdaságban és a gazdaságtudományban

A Financial Times a közelmúltban „digitális aranybányának” nevezte a hatalmas adattömegek előállításában, feldolgozásában és felhasználásában rejlő lehetőséget (Waters [2011]). A cikk számos példát mutat be üzleti célú hasznosításra. Ez már elég ahhoz, hogy a jelenség gazdasági, illetve üzleti elemzés tárgya is legyen, a gazdasági vonatkozások tárgyalását azonban kezdjük a gazdaságtudományi vonatkozásokkal!

A közgazdaságtudomány a társadalomtudományok körébe tartozik. Számos jele van annak, hogy a hatalmas méretű digitalizált adatbázisok előállításának, rendezésének és feldolgozásának technikai lehetősége a társadalomtudományi kutatásokra is jelentős hatást gyakorol. Ahogy Gary King, a Harvard Egyetem professzora írja: a társadalomtudományok művelői felismerik, hogy olyan kérdések vizsgálatára nyílik lehetőség, amelyeket eddig nem, vagy csak alacsony hatékonysággal lehetett kezelni, miközben azok rendkívül fontosak lehetnek a társadalom szempontjából. Ha például valaki a tömegek véleményére, attitűdjeire, viselkedésére kíváncsi bizonyos kérdésekkel vagy jelenségekkel kapcsolatban, az internetes közösségi hálókon bejegyzések millióit analizálhatja automatizált szövegelemzési módszerrel (King [2011] 719. o.).

Az emberi viselkedés a közgazdaságtan viselkedési irányzatának (*behavioural economics*) is tárgya. Akerlof–Shiller [2011] például a jelenlegi válságot viselkedési, illetve „jelki” tényezőkkel magyarázza. Az előző szakaszban említett digitális nyomoknak köszönhetően az emberi viselkedés ma hatalmas tömegű valós idejű adat segítségével elemezhető. Példaként lehet említeni a Barabási [2010] által leírt hálózatkutatói projekteket. Ezek legfontosabb eszköze az emberek által önkéntesen magukkal hordott „szenzor”: a mobiltelefon. 2010-ben több mint 4 milliárd ember, a föld lakosságának 60 százaléka használt mobiltelefont, minden ötödik készülék sokféle tevékenységre használható okostelefon volt, amelyek piaci forgalma évi 20 százalékos növekedést mutat (McKinsey Global Institute [2011] 2. o.).

A technológia fejlődése új lehetőségeket nyit gazdaságtudományi vizsgálatokhoz, a tudományos célú kísérletezéshez, régi és új törvények igazolásához vagy éppenséggel megcáfolásához. Tudjuk például, hogy a közelmúltban – a gazdasági válság ellenére – gyors fejlődésnek indult az elektronikus (internetes) kereskedelem. A hazai internetes boltok forgalma 2010-ben meghaladta a 133 milliárd forintot, ami a teljes kiskereskedelmi forgalom 1,8 százalékát jelentette.<sup>7</sup> Vannak olyan országok, ahol ez az arány lényegesen nagyobb. Magyarországon is létezik olyan nyitott, többféle árképzési rendszerrel dolgozó elektronikus piactér, ahol egy átlagos napon 15 ezer tranzakció (adásvételi esemény) zajlik, amelyek rengeteg digitális adatot „termelnek”. Az elektronikus kereskedelem kiváló terep adatigényes közgazdasági vizsgálatok számára. Hogy mást ne mondjunk: az elektronikus piacok egyes szegmensei közel állnak ahhoz, amit a közgazdaságtudomány „tökéletes piacnak” (*perfect market*) nevez – nyilván nem véletlen az irántuk mutatkozó kutatói érdeklődés (lásd például Brynjolfsson–Dick–Smith [2009]).

A vállalatok és a gazdaság más intézményei a kilencvenes évektől kezdődően rengeteg pénzt fektettek informatikai fejlesztésekbe. Számptalan vállalat használ fejlett számítógépes rendszereket folyamatai irányítására, ügyfélkapcsolatai kezelésére, ellátási láncának működtetésére, logisztikai rendszerének automatizálására stb. Ezek a rendszerek folyamatosan öntik magukból a valós idejű állapot- és műveleti adatokat. Sokat fejlődött az adattisztítás, -tárolás és -lekérdezés üzleti szervezetekben is használt technológiája (adattárházak, dokumentum-menedzsment rendszerek stb.). Digitalizált adatok állnak rendelkezésre a bakkártyás műveletekről, az ingatlanpiaci forgalomról, a tőzsdei tranzakciókról, a nyitott társaságok pénzügyi és egyéb mutatóiról, különböző készletekről, a vonalkódokat felváltó

<sup>7</sup> A hazai internetes kereskedelemről a GKIeNET közöl rendszeresen adatokat (lásd például Sz. B. [2011]).

rádiófrekvenciás azonosítók (RFID) terjedésének köszönhetően sokféle árucikk és jármű fizikai mozgásáról. A *McKinsey Global Institute* becslése szerint 2010-ben már több mint 30 millió hálózatba kapcsolt szenzor (*sensor node*) működött a közlekedésben, az iparban, a közműveknél és a kiskereskedelemben (*McKinsey Global Institute* [2011] 2. o.).

Mivel egyre több a számítógépes munkahely, az alkalmazottak munkahelyi viselkedése is elemezhető digitalizált adatbázisok segítségével. A szervezeteken belüli elektronikus levél- és telefonos forgalom például hasznos képet adhat az informális hálózatokról, a hatalmi viszonyokról, a részlegek és szakemberek közötti együttműködésről.

Mindezekből az következik, hogy a természettudományokhoz és más társadalomtudományokhoz hasonlóan a közgazdaságtudománynak is „adatgazdag” jövőre kell felkészülnie, a adatok hatalmas tömege a közgazdaságtudományt is mindent bizonytalansággal megváltoztatja. A gazdaság szereplői viselkedésének megértéséhez is részben saját adatbázisaikhoz való viszonyuk, adatkezelési és -felhasználási módjaik feltárásán keresztül vezet az út.

Térjünk vissza témánk szempontjából nyilvánvalóan fontos vállalati informatikai beruházásokhoz! A velük kapcsolatos számos trend közül kettőt mindenképpen ki kell emelnünk. Az első: a vállalati informatikai beruházások éves nagysága manapság több ezer milliárd dollárra rúg<sup>8</sup> és az elmúlt években jelentős emelkedést mutatott. Eközben az információk keletkeztetésének, kezelésének és tárolásának fajlagos költsége az elmúlt öt évben a hatodára zsugorodott.<sup>9</sup> Az egyik oldalon tehát hatalmas befektetések állnak – a kérdés az, hogy mi van a másik oldalon, ezek a befektetések miként térülnek meg, milyen üzleti eredményt hoznak.

A *Financial Times* említett cikke (*Waters* [2011]) a jelenséget üzleti aranybányának nevezi. Ha e hasonlatnál maradunk, kijelenthetjük, hogy a bánya működtetéséhez beruházások kellene. Mint láttuk, az üzleti célú informatikai beruházások óriási összegekre rúgnak. Mérési bizonytalanságok ellenére ezekről elég pontos képünk van. Nehezebb feladat azt megválaszolni, hogy ezek a beruházások (köztük a kifejezetten adatkezelési jellegűek, például adattárház-projektek) hogyan, milyen áttételeken keresztül térülnek meg. Üzleti kifejezéssel élve úgy is feltehetjük a kérdést, hol és milyen értékteremtő szerepe lehet a nagy digitalizált adatbázisok keletkeztetésének és felhasználásának.

A vállalatok és más intézmények most kísérleteznek ezekkel a szerepekkel, most kezdik „kiaknázni a bányát”. Néhány értékteremtő módozat már elég jól kirajzolódik. Soroljuk fel a fontosabbakat!

A vezetés és a döntéshozatal szempontjából a legfontosabb változás az intuitív, múltbeli tapasztalatokra, ösztönre, megérzésekre, kétes információk alapján kialakított véleményekre alapozott döntések visszaszorulása. Számos kutatási program bizonyítja, hogy a felsoroltaknak fontos szerepük van egy olyan világban, ahol a döntésekhez szükséges információk hiányosan és megkétszerezve állnak rendelkezésre (lásd például *Finkelstein* [2003]). A fontos döntési tényezőkről rendelkezésre álló nagy tömegű, valós idejű adat csökkenti a döntések kockázatát, ráadásul a valós idejűség több preventív döntést tesz lehetővé, az utólagos, megkésett reakció helyett a megelőzésre téve át a hangsúlyt. (Aligha kell részletezni, hogy ennek mekkora jelentősége van például az egészségügy területén.)

Az adatokra alapozott döntések könnyebben algoritmizálhatók és automatizálhatók. Az automatizálásnak többféle pozitív hatása lehet: kisebb költségek, rövidebb döntési ciklusidők, kisebb esély emberi hibákra, több lehetőség a döntések decentralizálására.

Az adatbányászat (*data mining*) elnevezés mintegy két évtizede jelent meg a tudományos és az üzleti világban. Mai művelői kifinomult matematikai statisztikai eszközökkel

<sup>8</sup> Az IDC adatai szerint 4 ezer milliárd dollárról van szó, ennyit költenek a vállalatok hardverre, szoftverre, informatikai szolgáltatásokra, saját informatikai személyzetükre, rendszereik működtetésére és irányításukra ([http://www.emc.com/digital\\_universe](http://www.emc.com/digital_universe), 2. o.).

<sup>9</sup> Uo.

elemeznek hatalmas adatbázisokat. Az üzleti célú adatbányászat fontos célja olyan összefüggések, mintázatok feltárása, amelyek közvetlen hatással vannak a gazdasági eredményre (lásd például *Hand–Mannila–Smyth* [2001], *Fajsi–Cser–Fehér* [2010]). A vállalatokat – érthetően – elsősorban azok a felismerések érdeklik, amelyek birtokában képesek azokat az ügyfélkapcsolati és értékesítési tevékenységüket javítani, amelyek vevőik viselkedéséről, a kereslet változásáról szolgáltatnak információt. Az eredmények hasznosításának többféle módjával találkozhatunk: megjelennek például az ügyfélcsoportok azonosításában (a demográfiai alapú piacszegmentálást tömeges adatelemzéssel alátámasztott, viselkedésalapú tagolás váltja fel), a bankok kockázatértékelési (*credit scoring*) eljárásaiban, az internetes boltok ajánló algoritmusaiiban (*recommendation engines*), a távközlési cégek lemorzsolódás-előrejelző rendszereiben.

A vállalatok példáját a gazdasági élet más intézményei is követik: az adóhivatalok például adatbányászati vizsgálatok segítségével dönthetik el, kiket érdemes alaposabb (ezért szükségképpen drágább) ellenőrzésnek alávetni, vagyis növelni tudják ellenőrzési és behajtási munkájuk találati valószínűségét.

Gazdasági célú statisztikai elemzések persze régen is készültek, az adattömeg-jelenség azonban megváltoztatta a nagyságrendeket. Hogy mennyire, azt érzékeltessük egyetlen példával: *Baker* [2008] megállapította, hogy egyedül a Yahoo cég havonta 110 milliárd (!) adatot gyűjt össze az ügyfeleiről. A hatalmas adattömegek statisztikai elemzése a természettudományos kutatás mellett nemcsak a gazdaságban kap egyre nagyobb szerepet, de más területeken is, így például a bűnüldözésben, a politikában vagy éppenséggel a szexuális viselkedés feltárásában.

Az adatok előállításához, tárolásához, feldolgozásához szükséges informatikai beruházások természetesen nem hoznak automatikusan üzleti eredményeket. Kutatási eredmények bizonyítják, hogy hasznosításuknak fontos feltételei vannak, és az sem kizárt, hogy több évet kell várni a pozitív következményekre (*Bögel* [2009b] 12. fejezet). A vállalati teljesítmény és a fenti értelemben vett (hatalmas adatbázisok, kifinomult statisztikai módszerek) üzleti analitika közötti összefüggés feltárását célzó kutatások egyelőre szórványosak, de az eredményeik biztatók. Az MIT-n dolgozó Eric Brynjolfsson és munkatársai például ökonometriai eljárással határozott pozitív kapcsolatot mutattak ki az adatalapú döntéshozatal (*data-driven decisionmaking*) és a vállalati termelékenység között, ügyelve a fordított oksági kapcsolat kiszűrésére is (*Brynjolfsson–Hitt–Kim* [2011]).

Mindezekből arra következtethetünk, hogy az adatbányászat a gazdaságban új szintet hoz a piaci versenybe. Versengés adatelemzéssel (*competing on analytics*) – így nevezi ezt *Davenport–Harris* [2007]. Az adatbázisok a vállalatok vagyónának fontos részét képezik, amit kezelni, hasznosítani és értékelni kell. Ma már sok olyan vállalat van, amelyek petabájtos nagyságrendben tárolnak adatokat, a száz terabájtos nagyságrend pedig már szinte általánosnak tekinthető a nagyobb cégek körében. Az adatbázisokkal kapcsolatos tudás, az adatbázisok tartalma, feldolgozásuk technikája növekvő fontosságú versenyképességi tényező. Minden jel arra vall, hogy ez nemcsak vállalatokra vonatkozik, hanem régiókra, sőt országokra is. A lehetőségek szempontjából különös figyelmet igényelnek az olyan adatintenzív területek, mint például az egészségügy vagy a pénzügyi szolgáltatások; az előbbin a McKinsey cég kutatóközpontja az Egyesült Államokban évi 300 milliárd dolláros „értékteremtési potenciált” jelez (*McKinsey Global Institute* [2011] 8. o.). Jól látható, hogy informatikai tekintetben lemaradó, elhanyagolt szektorokban és régiókban nehéz lesz a lehetőségek kihasználása.

A verseny munkaerő-piaci változásokat is hoz. A nagy digitális adatbázisok kezelése és elemzése különleges szakértelmet kíván, ami nem áll mindenütt korlátlan mennyiségben rendelkezésre. A keresleti előrejelzések egyes országokban tíz-, sőt százezres nagyságrendű munkaerőhiányt jósolnak. Ne feledjük, nemcsak informatikusokról és matematikusokról van szó, hanem olyan üzleti, államigazgatási, egészségügyi és más



vezetőkről, akik megértik az adatokban rejlő lehetőségeket, és ki is tudják azokat használni. Az utóbbiakból jóval többre van szükség, mint az előbbiekből. Kérdés, hogyan reagálnak ezekre az igényre az oktatási rendszerek.

### Új piaci szerepek és tevékenység típusok

Az adatrobbanás leírt jelensége élénk mozgást hoz magával az üzleti életben: a meglévő szervezetek alkalmazkodni próbálnak, miközben új piaci szerepek, tevékenység típusok és ezeknek keretet adó új üzleti modellek jönnek létre. Eddig az „adatpiacnak” főleg a keresleti oldalával foglalkoztunk, most vessünk egy pillantást a kínálatra is!

Az adatok keletkeztetése, tárolása, feldolgozása, továbbítása sajátos értéklánccá áll össze, amelynek egyes láncszemeihez sokféle technikai eszköz és szolgáltatás tartozik. A lánc minden elemét (hardver, szoftver, szolgáltatások együtt) csak kevés nagyvállalat képes hasznosítani. A kereslet bővülése, a növekedési lehetőségek sok vállalkozó és befektető érdeklődését keltik fel. A lánc különböző szemeinek vagy szemkombinációinak hasznosítása egy sajátos sokelemű, nyüzsgő adat-ökoszisztémát alakít ki.

A teljesség igénye nélkül soroljunk fel néhány figyelemre méltó piaci szerepet és vállalkozási formát!

a) *Vertikális szolgáltatók.* Az informatikai óriáscégek körében jól megfigyelhető, hogy egyre inkább igyekeznek kiaknázni az adatrobbanás kínálta előnyöket. A nagy részvénytársaságoknak állandóan növekedési lehetőségek után kell kutatniuk, és kétségtelen, hogy ma az adatüzlet az egyik legforróbb terület az informatikában. Az elmúlt évek piaci mozgásai (innovációk, felvásárlások és összeolvadások) következtében néhány nagy cég a hardver-, szoftver-, szolgáltatói és tanácsadási piacokon egyaránt jelen van, vagyis integrált megoldások nyújtására képes. Más kérdés, hogyan ítélik meg vevői oldalon az egy kiszolgáltatótól való függés kockázatát.

b) *Közművek, infrastruktúra-szolgáltatók.* Korábban már jeleztük, hogy egyfelől technikai innovációknak, másfelől gazdasági kényszereknek köszönhetően a nagy tömegű adatok tárolása, feldolgozása és archiválása egyre nagyobb részben megosztott adatközpontokban történik. Ezeknek több fajtája létezik, egyes központok például csak kapacitásokat biztosítanak, mások feldolgozást és más szolgáltatásokat is vállalnak. Több olyan vállalat van, amelyek eredetileg saját tevékenységük kiszolgálásához építettek adatközpontokat, most viszont a piacon értékesítik azok tartósan vagy ideiglenesen kihasználatlan kapacitásait.

c) *Eszközspecialisták.* Az adatokkal kapcsolatos műveletekhez helyenként az általános informatikai háttér mellett nagyon speciális eszközök is kellenek. Eszközspecialista lehet például egy intelligens kamerákat fejlesztő vállalkozás vagy valamilyen különleges szenzorfajta előállítója.

d) *Távközlési szolgáltatók.* A sávszélesség növekedése jól megfigyelhető tendencia a távközlési cégek körében. Az összefüggés a nagy adatközpontok modelljével egyértelmű: minél olcsóbbak és megbízhatóbbak a nagy sávszélességet biztosító távközlési szolgáltatások, annál több tevékenységet lehet és érdemes nagy adatközpontokba vinni.

e) *Kereső- és szűrőszolgáltatások.* Üzleti modelljüket a hatalmas adatbázisokban való eligazodás nehézsége táplálja. A keresésnek és a szűrésnek többféle módja van: a verseny tulajdonképpen algoritmusok rivalizálása. Mivel egyes algoritmusok öntanuló voltak miatt eleve adatigényesek, a hálózati hatás miatt a piacon könnyen monopóliumok vagy kvázimonopóliumok alakulhatnak ki.

f) *Adatkereskedők, brókerek és aggregátorok.* Az adat, illetve az adatbázis maga is árucikknek tekinthető, aminek élénk piaca van. Az adatbázisok adásvétel tárgyai: hozzáférést lehet hozzájuk biztosítani, meg lehet őket osztani, bérbe lehet őket adni, közvetítői szerepet lehet játszani eladók és vevők között.

g) *Algoritmusfejlesztők.* A tudományos, gazdasági és társadalmi élet legkülönbözőbb területein gyakran egymáshoz hasonló adatfeldolgozási (adatbányászati) eljárásokat alkalmaznak. Egy fejlett statisztikai alkalmazáscsomagot sokféle célra lehet felhasználni. Az általános célú, széles körben használható algoritmusok mellett speciális eljárások is vannak, amelyek fejlesztői egy-egy szűkebb ügyfélkört vesznek célba.

h) *Szakterületi hibridek.* Egy-egy speciális szakterületet szolgálnak ki algoritmus-készítési, adatfeldolgozási és más kapcsolódó szolgáltatásokkal. Ilyeneknek tekinthetők például a bioinformatikai vállalkozások: hibrid voltakra a megnevezésük is utal.

i) *Integrátorok és felhasználási tanácsadók.* Az ilyen vállalkozások a lehetőségek felismeréséhez, a megoldások kiválasztásához, az eredmények hasznosításához nyújtanak segítséget, sokszor úgy, hogy saját hardver- és szoftvereszközökkel egyáltalán nem rendelkeznek. Fontos szerepük van az ismeretek és megoldások megosztásában, újabb és újabb területekre való adaptálásában.

j) *Adatmegjelenítők.* Ők is az adat-ökoszisztéma sajátos, de adott esetekben nagyon fontos szereplői: vizualizálási szakemberek, művészek, kommunikátorok, vagyis mindazok, akik az adatfeldolgozási műveletek eredményeit a szűkebb vagy tágabb közönség számára fogyaszthatóvá teszik.

k) *Biztonsági szolgáltatásokat nyújtó vállalkozások.* Mivel az adatbázisok egyrészt nagyon értékesek lehetnek, másrészt a kezelésüknél eleget kell tenni az adatvédelmi előírásoknak, szakszerű védelemre is szükség van.

Alaposabb vizsgálat nélkül is belátható, hogy az egyes tevékenységek nagyon eltérő közgazdasági sajátosságokat mutatnak, és az ezekből eredő lehetőségek és kényszerek különböző stratégiáknak és üzleti modelleknek kedvezhetnek, miközben a technikai fejlődés sebessége miatt az esélyek és a helyezések folyamatosan változnak. Az is belátható, hogy az információs piacra jellemző nagy trendek hatása alól (lásd például Carr [2004]) ez a terület sem vonhatja ki magát.

## Kockázatok

Az elmondottak alapján az a benyomás alakulhat ki az olvasóban, hogy az adatrobbanás leírt jelensége tulajdonképpen mindenkinek jó: a sok valós idejű adatnak köszönhetően jobb döntések születnek, csökkennek a piac sűrűlódásai, kevesebb lesz a veszteség, sikerebbek lesznek az innovációk. Ebben sok igazság van, de nem jelenti azt, hogy a fejlődés kockázatmentes, és nincsenek benne súlyos, nehezen feloldható dilemmák.

Az egyik legvitatottabb kérdés a személyes adatok és a magánélet védelme. Mint láttuk, az egyik legfontosabb adatforrás az ember. A szakirodalomban *digitális árnyéknak* (*digital shadow*) szokták nevezni azoknak az adatoknak a tömegét, amelyek egy-egy emberrel kapcsolatban folyamatosan keletkeznek. E digitális nyomok egy részét maga az „árnyék” gazdája állítja elő szándékos cselekedeteivel, tudatosan; azzal például, hogy vásárol, elektronikus csatornákon lép kapcsolatba másokkal, bankkártyákat használ, honlapokon regisztrál, orvosi vizsgálatoknak veti alá magát stb. Az „árnyék” egyre növekvő része viszont ismeretlen számára, nem is tud a létezéséről. A személyes adatok és a magánélet védelmének megvannak az intézményei és az eszközei, de kérdéses, hogy ezek az internet korában milyen hatékonysággal működnek.<sup>10</sup>

<sup>10</sup> A névtelenség például kétes védelmet biztosít egy olyan világban, ahol nagyon kis számú semlegesnek tűnő adat (például születési dátum, nem, irányítószám) alapján elég nagy pontossággal be lehet egy személyt azonosítani.

Nem elegendő azonban csak a védelem kérdését felvetni. Sok olyan esettel találkozhatunk, amikor a személyes adatok, magánéleti vagy annak tartott információk megosztása jelentős haszonnal járhat a megosztó és mások számára egyaránt. Aki például aktív vevőként kapcsolódik be az internetes kereskedelembe, tapasztalhatja, hogy minél többet tudnak róla és a hozzá hasonló emberek viselkedéséről a kereskedők, annál pontosabbak lesznek az ajánlataik, annál jobban tudnak reagálni az igények változásaira. De említhetjük az egészségügy példáját is: statisztikai módszerekkel közelíteni bizonyos betegségekhez vagy hajlamokhoz akkor lehetséges, ha kellő mennyiségű adat áll rendelkezésre az érintettekről, vagyis ha digitalizált diagnózisokat és egyéb személyes adatokat széles körben módszeresen gyűjtik, kombinálják és megosztják.

Az érintetteknek tehát érdekében áll az adatok megosztása, és ugyanezt mondhatjuk a közösségről is. A technológiákban rejlő gazdasági lehetőségek (innovációk, vállalkozási lehetőségek, megtakarítások stb.) kihasználásához az adatok megosztására van szükség, ez az igény viszont óhatatlanul ellentmondásba kerül a személyes adatok és a magánélet védelmével. Ellentmondásokkal, dilemmákkal, érdekütközésekkel terhelt helyzet alakul ki. A személyes adatok és a magánélet védelme nyilván fontos érték, de egy bizonyos határon túl szembe kerül az alany és a közösség egyéb érkeivel. Ez a határvonal homályos és folyamatosan mozog, pillanatnyi helyzete – ha van egyáltalán ilyen – sajátos kockázatokkal (például az elmaradt haszon kockázatával) jár az érintettek nézve.

Szeretnénk felhívni a figyelmet egy másik ellentmondásra is. Bemutattuk, hogy egyes vállalatoknál hatalmas adatvagyon halmozódik fel. Az üzleti logika azt diktálja, hogy ezt a vagyont célszerű megóvni másoktól, vagyis monopolizálni kell. Az adat azonban olyan termék, amelynek a másolása, többszörözése nagyon csekély közvetlen költséggel jár, miközben egy többféleképpen hasznosítható adatbázis újbóli előállítására nagy befektetést igényelhet. A magánérdek tehát ebből a szempontból is szembe kerülhet a közösségi érdekekkel. Mielőtt azonban valaki egyes adatbázisok államosítására gondolna, nem árt, ha összehasonlítja a magánszektor kreativitását és gyorsaságát az államéval.

Ebben az összefüggérendszerben kell megemlítenünk az adatvesztés kockázatát is. Egyes számítások azt mutatják, hogy az évente előállított digitális adattömeg nagysága meghaladja a rendelkezésre álló tárolókapacitásokét, és a rés ráadásul növekszik; az adatok egy része tehát szükségképpen elvész (*McKinsey Global Institute* [2011] 16–17. o.) – de vajon melyik része? És egyáltalán: kié az adat, ki vigyáz rá? A *Science* folyóirat cikkünk elején említett felmérése szerint a megkérdezett kutatók (minden bizonnyal a legjobb tudományos intézmények vezető szakembereiről van szó) nagyjából fele a vizsgálati adatokat csak a saját laboratóriumában tárolja, ami több szempontból is vitatható megoldás: a kutatók, legyenek bármilyen kiválóak a saját szakmájukban, nem feltétlenül jó adatkezelők. A New Orleans-t romba döntő Katrina hurrikán több laboratóriumot és számítóközpontot is elpusztított. Nem véletlen, hogy a kockázat csökkentése és a közösségi érdek védelme érdekében számos területen nagyon komoly elképzelések születtek különleges státusú, közpénzből finanszírozott adatbankok és archívumok létrehozására,<sup>11</sup> illetve az adatok kombinálását és megosztását megkönnyítő metaadat-szabványrendszerek bevezetésére, miközben természetesen meg kell oldani a szellemi tulajdon védelmének jogi kérdéseit is.

Mivel az adat, illetve az adatbázis (esetenként igen nagyon) értékes árucikk, bűnözők érdeklődését is felkelti, így fennáll az adatlopás kockázata is. Az újságokban rendszeresen olvashatunk magán- és állami adatbázisok feltöréséről, nagy tömegű adat eltulajdonításáról. Az adatbiztonsági üzlet nem véletlenül virágzik.

<sup>11</sup> Példaként említhetjük például a genetikai információk tárolására szolgáló *GenBank*ot vagy az *UniProt* protein-adatbázisát.

A kockázatok között meg kell említenünk a nagy informatikai projektek jellegzetes problémáit is. Minél nagyobb és összetettebb akcióról van szó, a kockázat annál nagyobb. Ha valaki példát keres, könnyen rábukkanhat a brit állam éveken ezelőtt indított, az egészségügyi adatok digitalizálását és internetes megosztását megelőző óriásprojektjére, és az erről szóló jelentésekben megtalálhatja a határidő- és költségvetés-túllépésekre vonatkozó meghökkenítő adatokat is. Hatalmas lehetőségek kihasználásához hatalmas befektetések szükségesek, amelyek – sajnos – egyáltalán nem mentesek a kockázatoktól.

Számolni kell környezeti kockázatokkal is. Az Egyesült Államokban a számítógépek által felhasznált energia 2000 óta több mint a duplájára nőtt (*Doctorow* [2008] 17. o.). Cikkünkben emlegetett nagy adatközpontokat az energiaellátás és a hűtés problémájának megoldása miatt – fogyasztásuk közepes méretű városokéval vetekszik – lehetőleg erőművek és hideg vízű folyók közelébe telepítik. Becslések szerint a világban évente körülbelül 50 millió tonna szemét keletkezik kiszolgált elektronikai termékekből. Az Európában keletkező elektronikai hulladéknak csak 25 százalékát hasznosítják újra. Az elhasznált számítástechnikai eszközök nagy része elmaradott országok illegálisan működő, a környezetre rendkívül veszélyes szeméttelpein<sup>12</sup> végzi. Az adatrobbanás jelenségéről mindazonáltal bonyolult feladat környezeti mérleget készíteni, a felsorolt negatívumok mellett, mint láthattuk, nagyon fontos pozitív hatásokkal is számolni kell.

A kockázatok és dilemmák – mint jeleztük, egyáltalán nem teljes – felsorolását fejezzük be egy olyan gondolattal, amely egyszerre filozofikus és hétköznapi. Az információs túlterhelés (*information overload*) évtizedek óta jelen lévő problémájáról van szó. A túl sok információ ugyanúgy problémát jelent, mint a túl kevés. Az idevágó kutatások figyelemre méltó eredményeket produkálnak. Tudjuk (és tapasztaljuk), hogy az információs túlterhelés a modern élet stresszforrása, de arra vonatkozóan is vannak már bizonyítékok, hogy bizonyos szinten negatívan hat a kreativitásra és a termelékenységre.

## Összefoglalás

Cikkünkben az adatrobbanás jelenségét elemeztük. A technika fejlődése hatalmas adatbázisok létrehozását, tárolását, feldolgozását és megosztását teszi lehetővé, számtalan fontos lehetőséget teremtve a legkülönbözőbb területeken. Bemutattuk, hogy a hatalmas adattömeg (*big data*) jelenségének érdekes közgazdasági következményei vannak: a természettudományokéhoz hasonlóan átalakíthatja a közgazdaságtudományi kutatások módszertanát és vizsgálati eszközeit, vállalkozási lehetőségeket teremt, növekvő méretű, gyorsan fejlődő gazdasági ökoszisztémát alakít ki. Lényeges megállapításunk volt, hogy az adattömegek kezelésének képessége vállalati és nemzetgazdasági szinten egyre fontosabb versenyképességi tényező. A fejlődés ugyanakkor nem mentes a kockázatoktól és dilemmáktól.

Az óriási adatbázisok világa, a cikkben leírt adat-ökoszisztéma minden bizonnyal a jövőben is közgazdaságtudományi vizsgálatok tárgya lesz, a témával – így például az adat mint áru sajátosságaival, az adatpiacok működésével, az adatalapú üzleti döntéshozattal, az adat-ökoszisztémán belül folyó versennyel, a munkaerő-piaci kereslet mozgásával, a régiók és országok versenyképességével – kapcsolatban sokféle kutatási feladat adódik.

Ha már a feladatokról van szó, arra is fel kell hívnunk a figyelmet, hogy a vállalatok és más intézmények által előállított adatbázisok a nemzeti információk nagyon részei, aminek hasznosítása átgondolt stratégiát kíván minden döntéshozói szinten. Az államnak el kell döntenie, mennyit költsön adatgyűjtésre és -kezelésre, milyen hatást gyakorolhat a bir-

<sup>12</sup> Az egyik afrikai „számítógép-temetőről” nemrég megrázó felvételeket közölt a Newsweek: <http://www.thedailybeast.com/newsweek/galleries/2011/07/24/ghana-digital-waste-pieter-hugo-photos.html>.

tokában lévő adatok megosztásával, publikálásával a gazdaságra, a környezetvédelemre és más fontos dolgokra, hogyan oldja meg a mindenféle gombamód szaporodó adatbázisokkal kapcsolatos szabályozási feladatokat.

### Hivatkozások

- AKERLOF, G.–SHILLER, R. [2011]: *Animal spirits, avagy a lelki tényezők szerepe a gazdaságban és a globális kapitalizmusban*. Corvina Kiadó, Budapest.
- BAKER, S. [2008]: *The Numerati*. Jonathan Cape, Random House, London.
- BARABÁSI ALBERT-LÁSZLÓ [2010]: *Villanások*. Nyitott Könyvműhely, Budapest
- BÖGEL GYÖRGY [2009a]: Az informatikai felhők gazdaságtana. *Közgazdasági Szemle*, 7–8. 673–688. o.
- BÖGEL GYÖRGY [2009b]: *Üzleti elvárások – informatikai megoldások*. HVG Könyvkiadó, Budapest.
- BRYNJOLFSSON, E.–DICK, A.–SMITH, M. [2009]: A Nearly Perfect Market? Differentiation vs. Price in Consumer Choice. *Quantitative Marketing and Economics*, november, [www.springerlink.com/content/0278286157686419/fulltext.pdf](http://www.springerlink.com/content/0278286157686419/fulltext.pdf).
- BRYNJOLFSSON, E.–HITT, L.–KIM, H. [2011]: Strength in Numbers: How Does Data-Driven Decision-making Affect Firm Performance? MIT research paper, április 22. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1819486](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486).
- McKINSEY GLOBAL INSTITUTE [2011]: *Big Data: The next frontier for innovation, competition, and productivity*. (Szerzők: *Bughin, J.–Corb, L.–Manyika, J.–Nottebohm, O.–Chui, M.–de Muller Barbat, B.–Said, R.*) McKinsey Global Institute, 2011. június. [http://www.mckinsey.com/mgi/publications/big\\_data/pdfs/MGI\\_big\\_data\\_full\\_report.pdf](http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf).
- CARR, N. [2004]: *Does IT Matter?* Harvard Business School Press, Boston.
- CUSUMANO, M. [2008]: *The Changing Software Business: From Products to Services and Other New Business Models*. The MIT Center for Digital Business, Research Paper, január, No. 236.
- DAVENPORT, T.–HARRIS, J. [2007]: *Competing on Analytics*. Harvard Business School Press, Boston.
- DOCTOROW, C. [2008]: Welcome to the Petacentre. *Nature*, Vol. 455. No. 4. 16–21. o.
- DÖMÖLKI BÁLINT (szerk.) [2008]: *Égen-földön informatika*. Typotex Kiadó, Budapest.
- ECONOMIST [2010]: It's a Smart World. A Special Report on Smart Systems. *The Economist*, november 6. (melléklet).
- FAJSZI BULCSÚ–CSER LÁSZLÓ–FEHÉR TAMÁS [2010]: *Üzleti haszon az adatok mélyén*. Alinea Kiadó–IQSYS, Budapest.
- FINKELSTEIN, S. [2003]: *Why Smart Executives Fail*. Portfolio, Penguin Group, New York.
- FRANKEL, F.–REID, R. [2008]: Distilling meaning from data. *Nature*, Vol. 455. No. 30.
- GLEICK, J. [2011]: *The Information*. Pantheon Books, New York.
- HAND, D.–MANNILA, H.–SMYTH, P. [2001]: *Principles of Data Mining*. MIT Press, Cambridge.
- HILBERT, M.–LÓPEZ, P. [2011]: The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, Vol. 332. április 1., 60–65. o.
- KING, G. [2011]: Ensuring the Data-Rich Future of the Social Sciences. *Science*, Vol. 331. február 11. 719–721. o.
- McCANDLESS, D. [2010]: *Az információ gyönyörű*. Infografika. Typotex kiadó, Budapest.
- NESSI [2009]: *Strategic Research Agenda. NESSI Research Priorities for FP7*. május 10, Vol. 3.2. Revision 2.0. A NESSI Roadmap series document.
- PALUGYAI ISTVÁN [2010]: *Matematikusok készítenek a kulcsot a géntérképezőknek*. *Népszabadság*, július 21. 16. o.
- REDING, V. [2009]: *Internet of the Future: Europe Must Be a Key Player*. Future of the Internet initiative of the Lisbon Council. Brüsszel, február 2.
- Sz. B. [2011]: *Az online áruházak a válság nyertesei*. *Népszabadság*, augusztus 31., 10. o.
- WATERS, R. [2011]: *A Binary Goldmine*. *Financial Times*, május 6. 7. o.