

Dessewffy Tibor – Láng László

## Big Data és a társadalomtudományok véletlen találkozása a mütőasztalon<sup>1</sup>

### A Big Data dialektikája

Napjaink egyik legeredetibb közgazdásza, Dan Ariely szerint: „a Big Data olyan, mint kamaszkorban a szex: mindenki beszél róla, senki nem tudja, hogy kell csinálni, mindenki azt gondolja, hogy mindenki csinálja már, ezért mindenki azt állítja, hogy ő is csinálja”. Valóban, mint a felkapott, divatos fogalmak esetében általában, a Big Data is zavarba ejtően sokszínű értelmezési lehetőségeket kínál. Nem ambíciónk, hogy a Big Data átfogó, minden szférára kiterjedő és érvényes definícióját adjuk – kérdés, hogy ez egyáltalán lehetséges-e. A magunk részéről az alábbiakban a társadalomtudományi felhasználás szempontjából kívánjuk áttekinteni ezt a jelenségekört. Ezért számunkra a Big Data a digitális adatok robbanásszerű mennyiségi növekedése miatt előálló minőségileg is új társadalmi gyakorlatok és megismerési módok együttese. Azért is beszélhetünk a Big Data dialektikájáról, mert a digitális működés irányába elmozduló, átalakuló társadalmi valóság és az e valóság megértését szolgáló, a digitális lábnyomokra fókuszáló eljárások egymást erősítik. Az alábbiakban elsősorban ennek a dialektikus viszonynak a második elemére és a társadalomtudományok számára jelentkező kihívásokra koncentrálunk. De nem árt észben tartani: a társadalmi folyamatok és gyakorlatok fokozódó digitalizációja nyomán a megértés új útjai nem egyszerű-

<sup>1</sup> E tanulmány az elmúlt másfél évben számos előadáson és kollegiális vitában formálódott, ahol fontos impulzusokat és kritikákat kaptunk mondanánk továbbgondolásához. Külön is szeretnénk köszönetet mondani Benczúr András, Bozsonyi Károly, Győri Gábor, Lányi Gusztáv, Kertész János, Nagyfi Richard, Németh Renáta és Z. Karvalics László észrevételeiért, kritikái megjegyzéseiért. Hálásak vagyunk Csepeli Györgynek azért, hogy e tanulmány kapcsán is, de ezen túlmenően is a témában folyamatos intellektuális eszmecsereben és dialógusban lehetünk partnerei. Továbbá szeretnénk megköszönni Berger Vikornak írásunkhoz fűzött megjegyzéseit, valamint a szerkesztői szerepen túlmutató szöveggondozási segítségét. Természetesen, a tanulmányban található állításokért, valamint az esetleges hibákért, ahogy mondani szokták: a felelősség kizárólag minket terhel.

en választható alternatívák, hanem egyre inkább a megismerés adekvát módozatai lesznek. Lehetséges mondjuk a fiatalok kulturális fogyasztásairól vagy az idősek betegségeikkel kapcsolatos szorongásairól kutatásokat készíteni, de aligha kockáztatunk sokat, ha azt állítjuk, Facebook-csoportok vagy egészségügyi honlapok látogatottságának elemzésével ezekben a kérdésekben is mélyebb és pontosabb válaszokat kaphatunk.

A Big Data jelentőségének megértéséhez tehát két folyamatból kell kiindulni. Az egyik a világban elérhető adatmennyiség robbanásszerű növekedése. E folyamat érzékeltetésére jó néhány szakértői számítás forog közkézen. Az egyik visszafogott becslés szerint, míg 2002-ben a világon összesen 27 exabyte-nyi adat keletkezett, addig ma 7 nap alatt jön létre ennyi információ. Ha a pontos számok kérdésesek is lehetnek, az exponenciális növekedés ténye vitathatatlan.

A másik meghatározó folyamat az adattermelés és az adatok digitalizációja. 2000-ben még a világ adatvagyonának hetvenöt százalékát analóg formában, papíron, filmekben, kazettákon tároltuk. 2013-ban az információtömegnek a 98 százaléka digitális formában érhető el. *Tehát nem pusztán a mennyiségből, hanem az adatok digitális voltából fakadnak a megismerés új lehetőségei.* A digitális jel megmutat ott is, ahol korábban nem értük utol a reprezentációt: egy nyomtatott újság esetében maximum az eladott példányszámot tudhatjuk, a Big Data esetében viszont akár azt is, hogy az olvasók mely cikkeket olvasták, mennyi időt töltöttek az egyes írásokkal, és ez milyen mintázatokat rajzol ki, például érdeklődik a külpolitikai cikkek olvasói a gasztronómia? A digitális jel korlátlanul *időtálló* – évekkel ezelőtti olvasási szokásaink éppúgy kereshetőek, mint a jelenlegiek. A digitális adatbázisok *összekapcsolhatóak és kereshetőek*, az újságolvasási szokásokat kombinálhatjuk vásárlási adatokkal. Végül érdemes észben tartanunk, hogy digitális jel *nemcsak az emberi tevékenység* nyomán jön létre, Facebookos feltöltéseink és like-olásaink mellett okostelefonunk különböző instrukciók nélkül is napi 24 órában rögzíti, hogy éppen hol vagyunk. Ez a napról napra növekvő digitális adatvagyon hívta életre a Big Data alapú adatbányászatot, amely az új problémafelvetések, módszerek és megközelítések, amint arra Ariely is utalt, nehezen megragadható gyűjtőfogalma.

## I. Big Data – szociológiai elméleti előképek

Ez az *újszerűség* azonban *nem szociológiai előképek nélkül való*. E helyütt nem aspirálhatunk arra, hogy a társadalomtudományok történetét áttekintsük a Big Data szempontjából. Ezzel szemben *a múlt század két végpontjáról mutatunk be egy-egy példát*, melyek a szociológián belül jelezték azokat az igényeket és lehetőségeket, amelyeket ma a Big Data révén lehetséges realizálnunk.

1903-ban egy a gondolatmenetünk szempontjából különösen érdekes, a szociológia természetére vonatkozó vita zajlott Gabriel Tarde és Émile Durkheim között. A szociológia kanonizált története szerint ezt a vitát Durkheim, a tudomány alapítóatyáinak egyike „nyerte”. Az ő holisztikus társadalomszemléletében a szociológia feladata a „társadalmi tények” megismerése.

Durkheim szerint a társadalmi tények az egyéntől, a személyes tudástól és gyakorlattól függetlenek, és kizárólag a kollektív tudat, illetve annak változásai határozzák meg. E kollektív társadalmi tudat változásai kizárólag társadalmi okokkal magyarázhatóak (Durkheim

1895). Vagyis az egyéni, individuális, biológiai jellegzetességeket nem vizsgálni kell, hanem éppenséggel megszabadítani tőlük a kutatást. Módszertanilag ugyan Durkheim elkötelezett az empirikus kutatások mellett, de szerinte a társadalmi tények kizárólag aggregát, összesítő statisztikákon keresztül ragadhatóak meg – mert úgy vélte, ezek az egyéni nézetek és vélemények sokféleségét egy magasabb szinten összegzik.

Ezzel szemben Tarde, a vita másik résztvevője a saját individualista, vagy ahogy ő nevezte, monadikus megközelítése mellett érvelt. Ebben az elméletben a sokaságot alkotó egyének sokszínűsége, különbözősége nem pusztán figyelembe veendő, megőrizendő erény, hanem éppenséggel az eltérések, egyéni sajátosságok megértése és magyarázata a szociológia feladata.

Azt hiszem, hogy a fent említett különbségek egyike sem magyarázható (...) még a pontosan megegyező atomi részecskékből kiinduló elméletekkel sem – amely hipotézis oly kedves vegyészeknek, akik e tekintetben az igazi metafizikusok (...) Azt hiszem, hogy az egyetlen módja megmagyarázni az egyéni különbségek kirobbanó növekedésének jelenségét, ha azt feltételezzük, hogy a jelenségek tarka rendjéből ered, mivel mindegyik saját egyedi jellemzőkkel rendelkezik (Tarde 1898: 15).

Tarde tehát hangsúlyozta egyes egyének vagy csomópontok sokszínűségét, a köztük levő kapcsolatok dinamikus jellegét és komplexitását. A társadalmi valóság ilyen megközelítése nyilvánvalóan gyümölcsözőbb a Big Data alapú elemzések számára. Érdekes, hogy ezt a módszertani, szemléletbeli változást maga Tarde is megjósolta.

Ha a statisztika továbbra is úgy fejlődik, mint az elmúlt jó néhány évben, ha tovább növekszik az általa nyert információk pontossága, gyorsasága, mennyisége és rendezettsége, eljőhet az idő, amikor megvalósulhat, hogy minden társadalmi eseményhez azonnal és automatikusan egy szám rendelhető, ezáltal, hogy úgy mondjam, a statisztikai nyilvántartásokban elfoglalja az őt megillető helyét, hogy aztán folyamatosan kommunikálhassuk a nyilvánosság számára és képeken terjeszthessük külföldön a napi sajtóban (Tarde, idézi Latour 2010: 159).

Bruno Latour úgy véli, hogy ez a megközelítés nemcsak sokkal hasznosabb a társadalmi jelenségek megértésében, mint Durkheimé, de napjainkban Tarde, és különösen a Big Data szempontjából, egy igazán inspiráló elméleti megközelítést kínál.

Ez az a pont, amikor rájövünk, hogy Tarde miért tűnik olyan frissnek (...) Tarde elméletének a legérdekesebb része az a világos elvárás, hogy milyen típusú információkat kell gyűjteni a társadalom tudományához. Valóban meglepő, hogy a jelen leggyorsabban növekvő témái, mint az „adatvizualizáció”, „számítástechnikai társadalomtudomány” vagy „biológiai hálózatok”, melyek itt alakulnak ki a szemünk előtt, épp azokat a fajta adatokat kínálják, amelyeket Tarde várt volna el. (...) A digitális navigáció által nyújtott, innen-oda útvonalakat bemutató „datascapé” lehet, egy évszázaddal később, igazolta Tarde meglátásait (...) [M]a, hála a digitális médiumnak, tanúi vagyunk a nyomon követhetőség elve lenyűgöző kiterjesztésének. Ez ma nem csak a tudományos eredményekben érhető tetten, hanem a véleményekben, pletykákban, politikai vitákban vagy az olyan cselekményekben is, mint a vásárlás és licitálás, a közösségekhez való csatlakozások, térbeli mozgások, telefonhívások és így tovább. Ami korábban csak a tudományos tevékenység számára volt lehetséges – hogy megmaradjon az aggregátumok káposztája, és az individuális kontribúciók kecskéje is jól lakjon –, ma a digitális nyomok, archivált digitális adatbankok korában már a legtöbb eseményre vonatkoztatva lehetséges. Ezt egyébként a Google-nak és társainak köszönhetjük (Latour 2010: 159–160).

Ez az új helyzet, amit Manuel Castells, az információs kor katalán prófétája és legátfogóbb elemzéseket kínáló szociológusa megsejtett már akkor, amikor a Big Data fogalma még nem volt közismert. Castells szerint a civilizációk szempontjából két alapidimenzió alapvető: a tér és az idő kezelése. Azért is jogosult az új információs korszakról beszélni, mert ebben ez a két dimenzió megváltozik, létrejön az *áramlások tere* és az *időtlen idő*.

Az „áramlások tere” azt a jelenséget igyekszik megragadni, hogy az új információs korszakban a társadalom a *hálózatban folyó áramlások köré épül*: tőke, információ, technológia, kódok és szimbólumok ezekben a hálózatokban áramlanak. A hálózati áramlás nemcsak az egyik lehetséges működési módja a társadalmi rendszernek, hanem a legfontosabb szerkezeti forma, amely meghatározza a gazdasági, politikai és szimbolikus szféra működését. Az információs korszak tehát új domináns térbeli mintázatot hoz létre, a hálózatiságot, amelyen belül értelmezhető az áramlások tere.

Az időtlen idő az információs korszak új időfelfogására utal. Az időnek ez a konceptualizálása erősen kapcsolódik a hálózati működésmódhoz. Egyszerre jeleníti meg az időtlenséget és a pillanat gyors elmúlását. Az időt két végletbe formálja: a kiterjesztett jelen egyidejűségébe, ahol minden real time elérhető, másfelől az *örökkévalóság* időtlenségébe, ahol a digitális lábnyomok örökre megmaradhatnak.

E két alapidimenzió, a tér és idő új koncepciója mentén alakul ki a valós virtualitás kultúrája, amelyben a valóság maga immár egy virtuális rendszerben konstruálódik és áll össze mai késő modern világunkká (Castells 1996).

Castells trilógiája első kiadásának kéziratát 1996 márciusában adta nyomdába, alig másfél évvel azután, hogy a Netscape, az első könnyen kezelhető webes kereső megjelent – és ezzel laikusok számára is elérhetővé tette az internetet. 1996-ban, amikor az internetet még primitív design, szűkös tartalom és nehézkes hozzáférés jellemezte, egyáltalán nem volt magától értetődő, hogy Castells értelmezése nem science fiction, hanem szociológiai leírás. Azonban a Big Data forradalma után Castells tézisei, illetve az információs társadalomról adott jövődőlései a szemünk előtt válnak valósággá.

## II. Big Data, mint új ismeretelméleti paradigma

A hálózati térben minden kapcsolat, cselekvés, érdeklődés nyomot hagy, kutathatóvá válik. A megismerés és megértés új útjai nyílnak meg azáltal, hogy a világméretű hálózati közösség böngészésének, keresésének, navigációjának, kapcsolatainak, értékeléseinek és vásárlásainak beláthatatlan tömege nem vész el, hanem velünk marad és így kutatható. Ez egy olyan új ismeretelméleti paradigmát jelenít meg, amelyben a korábbi korszak magától értetődő alapvetései, módszertani standardjai kiüresednek, irrelevánssá válnak.

A jelen okfejtés szempontjából csak két dolgot kell kiemelnünk. Egyrészt, hogy a „szociológiai kánon” mindig kontextuskötött. Az, hogy mit tekintünk elfogadott tudásnak, mindig is a módszertani, ismeretelméleti előfeltevésekhez kötött volt. Másodsor, ez a kánon, „szociológia” soha nem volt homogén. Ebből fakadóan a szociológia jövőjére nézve nem az következik, hogy a Big Data új ismeretelméleti korszakában irrelevánssá vált. Inkább az történik majd, hogy ez a sokszínű hagyomány újragondolja, átrendezi önmagát szellemileg. Ahogy Tarde példája is mutatja, lehet, hogy új klasszikusok jelennek meg, „újrafelfedezik” azokat a gondolkodókat, akik jobban illeszkednek ebbe az új paradigmába.

A szociológusoknak újra kell gondolniuk hagyományos adatgyűjtési technikáikat és fogalomhasználatukat, és meg kell tanulniuk, hogy megértsék és használják az egyre növekvő adatfolyamot – vagy legalábbis képessé váljanak értelmes párbeszédet folytatni az adatbányászat szakembereivel. Az elmúlt években ugyanis az analitikus szakemberek egy új csoportja alakul ki, akik képesek az adatkezelés, az adatgyűjtés és -tisztítás, a -tárolás és -elemzés egész folyamatának lebonyolítására.

Mayer-Schönberger és Kenneth Cukier ezeket az új szakembereket „algoritmistáknak” hívja, utalva az alkimistákra. Az algoritmista az, akit azért alkalmaznak, hogy strukturálja, formálja és vallassa az adathalmazokat. A szociológusoknak korántsem kell ebbe a csoportba tartozniuk, mert ők azok, akik analitikus eszközeikkel megfelelően értelmezhetik a klasztereket, mintázatot, kapcsolatokat és oksági viszonyokat.

Az átmenet az elmélet által vezérelt kutatásokból az adatok által vezérelt vizsgálatok felé egyben kísértést jelent, hogy azt gondoljuk, nincs szükség az elméletre többé – amint azt Chris Anderson programadó, *Az elmélet vége* című cikkében meg is hirdette (Anderson 2007). Anderson elképzelése szerint a tudomány hagyományos, elméletekből levezetett hipotéziseket empirikusan igazoló vagy elvető modellje elveszítette a relevanciáját. Az új helyzetben, amikor az adatok és adatfeldolgozási kapacitások szinte korlátlaná válnak, a korrelációk és mintázatok adatbányászati eszközökkel történő feltárása lesz a domináns, vagyis teoretikus előfeltevésekre nincs szükség, mert az adatok megdolgozása ezek nélkül is eredményhez vezet. Az „elmélet vége”-tézis azonban abszurd. A digitalizált, Big Data által uralt világban az elméletre új szerep vár. Bár valóban vannak esetek, amikor az adatbányászat meglepő korrelációkat mutat meg, ám a legtöbb problémát nem lehet megoldani kizárólag a korrelációk, mintázatok bemutatásával.<sup>2</sup> Az összefüggések értelmezése, azokból a tanulságok levonása igencsak földhözragadt maradna elmélet nélkül. A jelentés több, mint algoritmus. Világunk megértéséhez a Big Data által kínált új látásmód szükséges, de az eszköz önmagában nem elégséges.

Másfelől a világ adatokkal való megragadása egyidős az emberi közösségek történetével. Schütz fenomenológiai munkáiban az ember alapvető jellegzetességének tartja, hogy valamely értelmezhető rendet kell vetítenie a kozmológiai, természeti és társadalmi káoszba. A káoszon úrrá lenni törekvő ember egyik leghatékonyabb „rendteremtő” eszköze az adat, az információ. Így ettől a kényszertől és vágytól vezérelve kerülnek már az első barlangrajzokra az elejteni kívánt állatok képei és az elejtett példányok számára utaló karistolások. Az adatok előállítása és gyűjtése fontos komponens volt már a korai civilizációk fejlődése számára (Rifkin 2010). A mezopotámiai és egyiptomi kultúrákban összegyűjtött és rögzített (csillagászati, meteorológiai, vízmennyiségi és mezőgazdasági) információk tették lehetővé, hogy az ókori birodalmak kielégítésék a termelési, a közellátásra vonatkozó, a kereskedelmi és a kormányzati adminisztratív igényeket. Ez az adatgyűjtés korántsem ártatlan vállalkozás. Ahogy a politikai közgazdász, James C. Scott meggyőzően érvelt, a felvilágosodás óta a nemzetállamok bürokratikus kontrolljának előfeltétele volt az adatkezelés standardizációja és a folyamatok adatosítása (*datafication*).

2 A Wall-mart, a világ legnagyobb kiskereskedelmi bolthálózatának és a Big Data lelkes alkalmazójának informatikai vezetője, Linda M. Dillman pl. arról a meglepő tényről számolt be, hogy hurrikánok idején hirtelen megugrik a kereslet az eperlekvárral töltött ropogós rágcálnivaló iránt. A céget nem érdeklik az okok, számukra épp elég azt tudni, hogy a hurrikánszezonban a zseblámpák, ablakvédő fatáblák és szögek mellé látványos helyen kell felhalmozniuk a Strawberry pop tart dobozokat – a tapasztalatok szerint ilyenkor a kereslet nem ismer határokat (<http://digital-stats.blogspot.hu/2013/03/wal-mart-customers-buy-pop-tarts-when.html>).

A hüvelyk, a rőf, az egynapi járőföld helyébe a metrikus rendszer lépett a maga standard és így könnyebben kezelhető, átváltható és illeszkedő adatrezsimjével. Az idő tekintetében is a „pitymallat”, „napestig” és a „vecsernye idején” lassan feledésbe merült, és felváltotta a pontosan felbontható, mérhető és mindenki számára globálisan is összehangolt azonos „homogén idő” (Anderson 1989; Hunt 2008). Tér és idő transzparens, személytelen és univerzális fogalmakba szorítása teremtette meg a valóság adminisztratív-bürokratikus eszközökkel való uralásának lehetőségét. E rendkívül centralizált folyamatok alapját az a meggyőződés adta, hogy a tudományos beavatkozás javíthatja minden szempontból az emberi életet. Ehhez pedig adatokra, lehetőleg minél több és minél pontosabb adatra van szükség. A modern Leviatán adatéhsége kielégíthetetlen. Ez a vonás a modernitás lényegi sajátossága, amely minden részterületet meghatároz. James kiemeli, hogy az adatosítás a modern mezőgazdaságban, az erdőgazdálkodásban, a városi fejlődésben, a társadalmi mozgalmak és az ipari termelés esetében hogyan vezetett a sokoldalú, szerves, helyi, hétköznapi tudás visszaszorulásához és gyakran eltűnéséhez. Az organikus emberi látásmód helyett kerültek előtérbe az adatalapú, bürokratikus és standardizált eljárások, s kezdtünk el Scott könyvének címével élve „úgy látni, mint az állam” (Scott 2000).

### **Big Data, mint módszertani paradigma**

A szociológia és más empirikus társadalomtudományok az elmúlt száz évben nagyrészt abból nyerték legitimitásukat, hogy ők szállítottak „tényeket”, „kemény adatokat” a politikai és üzleti döntéshozók, valamint az érdeklődő nagyközönség számára (Savage és Burrows 2007). Ez azt is jelentette, hogy módszertanilag a mintavételen alapuló empirikus felmérés lett az ilyen elemzések legáltalánosabb eszköze. Ezért nevezte Jean Baudrillard az ezerfős reprezentatív mintán alapuló kutatásokat a modernitás kristálygömbjének.

Indokolt tehát, hogy a Big Data felhasználásának sajátosságait a reprezentatív mintavétel szemben próbáljuk feltérképezni. Az alábbiakban semmiképpen sem amellett kívánunk érvelni, hogy a Big Data alapú megközelítések, kutatások módszertanilag problémamentesek lennének. Épp ellenkezőleg: mivel az adatok jelenleg többnyire rendszerezetlenek, „maszatosak”, hiányoznak a bejártatott standardizált megoldások és eljárások, valamint a módszertant hasznosító elméletek sem erősödtek még meg, napjainkban számos probléma jelentkezik ebben a megközelítésben is. Azonban úgy véljük, érdemes végiggondolni azokat az eltéréseket, amelyek a rohamosan fejlődő Big Data-kutatásokat meghatározzák majd. Mayer-Schönberger és Kennett Cukier „A Big Data-forrdalom”-ról szóló könyvükben a reprezentatív mintára utalva nevezik az új paradigmát „*n=all*”-nak, ahol a minta mérete gyakran egybeesik az adott populáció nagyságával (Mayer-Schönberger és Cukier 2013).

Mindazonáltal szociológiai szempontból nem pusztán a minta nagysága az, ami igazi új-donságot hoz a kutatásba. Véleményünk szerint öt további különbséget érdemes szem előtt tartani, amikor összehasonlítjuk a mintavételen alapuló és a Big Data-megközelítéseket. Ezek: az időbeliség dimenziója, a viselkedési információk szerepe, a strukturált és a nem strukturált adatok kombinálásának lehetősége, a globális elérhetőség és az egyének közötti kapcsolatokra fókuszálás lehetősége.

1. *Az időbeliség dimenziója.* A hagyományos módon végrehajtott adatgyűjtés fárasztó, költséges és időigényes volt. Mindig is jelentős eltérés volt az adatgyűjtés időpontjában megfigyelt és a feldolgozott adatok közzétételekor uralkodó állapotok között. Többnyire évek teltek el az adatgyűjtéstől, így a közzététel időpontjában a tanulmányok elemzése jóval korábban összegyűjtött adatok alapján történt (Csepeli és Prazsák 2004).

A digitális technológiák alkalmazása révén az adatfelvétel sebessége is nagymértékben növekedett.

Azzal, hogy a korábban analóg információ digitális formában lesz elérhető, a szociológiai megismerés hatékonyabb és gyorsabb lehet. A valós idejű, real-time viselkedés mérése megoldást kínálhat azoknak a mutatóknak a mérésében, ahol problémát jelent a mérés időbeli elcsúszása (Csepeli és Dessewffy 2015).

2. *Vélemény vs. viselkedés.* A kérdőívekből véleményeket ismerünk meg. Ezen vélemények többnyire ismeretelméletileg „maszatosak”, a belőlük kirajzolódó kép valósághoz fűződő viszonya esetleges.

Nyilvánvaló, hogy a kérdőívekre adott válaszok tartalmát torzíthatják a politikai fenntartások, esetleg félelem (például az 1980-as években a „hogyan vélekedik a Szovjetunióról?” kérdés esetében). De többről van szó, mint hogy a politikailag, gazdaságilag, szexuálisan érzékeny kérdések esetében a válaszok nem a valós vélekedéseket tükrözik. Schütz fenomenológiai tudásszociológiájának alaptétele, hogy a mindennapi tudás kiemelt jelentőségű, mely a mindennapok magától értetődő, „elsődleges valósága” értelmezésének lerakódott eredménye. Ez a mindennapi tudás azonban sajátos szabályokat követ, lényegénél fogva inkoherens, összefüggéstelen és nem reflexív. Ez az alapvetés bomlik ki később azokban a szociálpszichológiai művekben, amelyek az öncsalásra való szinte határtalan képességünket mutatják be (Tavris és Aronson 2008). Ez azt jelenti, hogy egy kérdőíves felvétel esetében még a legártatlanabb vizsgálatoknál is komoly ismeretelméleti kétségeink lehetnek a válaszok érvényességére vonatkozóan. Nem azért, mert félre akarjuk vezetni a kutatókat, hanem mert legtöbbször nincs kikristályosodott és tiszta álláspontunk az adott kérdésekről. A szociológiai kutatás számtalan területéről lehetne hozni példákat: az időmérleg-vizsgálatoktól az attitűdskálákkal végzett felmérésekig lehet találkozni ezzel a problémával. A kérdezési helyzetben készült felmérések hajlamosak olyan válaszokat produkálni, amelyek tükrözik a válaszadók vágyát, hogy megfeleljenek a vélt elvárásoknak. A survey-vizsgálatok eredménye azért is ennyire ingatag, mert homályos és gyakran nem létező nézeteket tükröz. Ezzel szemben a Big Data alapú kutatás alapvetően a digitális lábnyomokat, azaz a valódi viselkedés nyomait használja. Ha például valakinek az internetes vásárlási szokásai, preferenciái érdekelnek, akkor nem a „Böngészés közben felkeresi-e az alábbi oldalakat?” kérdést kell megválaszolni, mint egy kérőíves kutatásban, hanem a valódi viselkedés alapján az oldalak látogatottságát, az ott töltött időt, a tovább kattintások irányát és a kapcsolatokat is tisztán láthatjuk. Így a kutatók sokkal pontosabb képet szerezhetnek a valós preferenciákról, döntésekről és cselekvésről, mint a hagyományos módszerekkel.

3. *A Big Data adatbázisai strukturált és nem strukturált adatokat tartalmazhatnak.* A „strukturált adaton” az olyan információt kell érteni, amelynek világos és tisztán meghatározott a hosszúsága és formátuma. A strukturált adatok közé tartoznak a számok, dátumok, szavak és számok csoportjai. A legtöbb szakértő egyetért abban, hogy ezek a strukturált

adatok az összes adat mintegy 20%-át teszik ki. E strukturált adatok származhatnak szenzorokból, GPS-ekből, intelligens mérőórákból, orvosi és más eszközökből. Az internethasználat során a webnaplók, a rögzített kattintások és a tranzakciók strukturált adatok egyre növekvő és hatalmas forrását képezik. A strukturálatlan adatok az előbbiekből következően a becslések szerint mintegy 80%-át teszik ki az összes adat mennyiségének. Ezek az adatok, amelyeket hagyományosan tartalomnak tekintünk: a képek, videók, szövegek, webes tartalmak stb. Mivel ezek a strukturálatlan adatok digitális formátumban érhetőek el, számos új kifinomult tartalomelemzés, pl. szövegelemzés tárgyává tehető (Hurwitz, Nugent, Fern és Kaufman 2013). Ezen túlmenően a két különböző típusú, a strukturált és strukturálatlan adat kombinálása gyökeresen új kutatási területeket nyit meg a társadalomtudósok számára. A strukturált és strukturálatlan adatok kombinálásakor nemcsak egy adott tartalom változásait figyelhetjük meg, hanem mondjuk egy Facebook-vita esetében azt is rekonstruálhatjuk, hogy az online közvélemény pontosan milyen interakciókon keresztül változott.

4. *Hozzáférés.* A Big Data – mivel az üzleti vállalkozások egy része nagy hangsúlyt fektet arra, hogy adataihoz ne lehessen hozzáférni – látszólag ijesztő távolságba kerül a társadalomkutatóktól.<sup>3</sup> Ám a helyzet ennél komplikáltabb, de kedvezőbb is. Ha valaki elsajátítja az adatgyűjtés és -feldolgozás módszereit – vagy együttműködést alakít ki erre nyitott algoritmistákkal –, hirtelen szinte beláthatatlan lehetőségek nyílnak meg előtte vagy akár a félperiféria kutatói számára is. A survey a nemzetállami keretekre szabott kutatási módszer volt: az 1000 fős minta a nemzetállami határok között volt értelmezhető és reprezentatív. A nemzetközi összehasonlító vizsgálatok rendkívüli költségeket, szervezési erőfeszítéseket igényeltek – számos ismeretelméleti probléma szőnyeg alá söprésével. Arra pedig, hogy valaki egy másik országban a saját elképzelései alapján végezzen kutatásokat, szinte esély sem volt. Ez a helyzet megváltozott a nyitott kutatói adatbázisok korában, ám itt is a már felvett adatbázisokból választhattak a kutatók, amelyek nem a saját elképzeléseik és kérdésfeltevésük alapján lettek kialakítva. Ebben a tekintetben az Open Data-mozgalom, illetve a Big Data tereket és megközelítéseket nyit meg. Ha valakit például a magyar és amerikai olvasási szokások, könyves trendek, fogyasztói affinitás érdekel, a magyar moly.hu több mint száz-ezer és az amerikai goodreads több mint 43 millió felhasználójának anonimizált preferenciáit és értékeléseit is letöltheti.

5. *Az egyes elemek közötti kapcsolatok láthatóvá válnak.* Mint már említettük, az „ $n=all$ ” paradigmában a mintának nincsenek méretbeli korlátai, vagyis a minta lehet olyan nagy, mint a vizsgált közösség maga. Ez szociológiai szempontból nem csak azért releváns, mert így módon ki lehet küszöbölni a statisztikai hibákat, hanem azért is, mert a szegmentáció leképezheti a valóság sokszínűségét. A számosságbeli korlátok miatt a reprezentatív min-

---

3 Ebben a tanulmányunkban a Big Data által megnyitott lehetőségekre koncentrálunk. Ez természetesen nem jelenti azt, hogy a lehetséges aggályok és veszélyek ne léteznének az adatvédelmi megfontolásoktól az episztemológiai beszűkülésig, amikor a csak digitális adatokból kinyerhető válaszok korlátok közé szorítják már a kérdésfeltevést is (boyd és Crawford 2012). Egy fenyegető elemre azonban itt mégis fel kell hívunk a figyelmet: ez pedig az adatokhoz való hozzáférés kérdése. Nemcsak arról van szó, hogy az elmúlt években egyes cégeknél példátlan mennyiségű és komplexitású adat termelődött, hanem arról a tendenciáról is, hogy a cégek, mint a Facebook, felismerve az értéküket, szűkebbre szabják és egyre nehezebben teszik elérhetővé az adatokat. A folyamat végét még nem látjuk, és bár tanulmányunkban ezt nem taglaljuk, reális lehetőség van egy új adatbirtokló elit kialakulására, amellyel szemben páriává válnak a hagyományos társadalomtudományok képviselői is.



tán bizonyos fokú összetettséget tudunk elérni – így az összefüggések feltárása is részleges marad. Ha valamilyen oknál fogva a szélsőjobboldali, vegetáriánus 18–35 év közötti nők klasszikus zenei preferenciáira lennének kíváncsiak, az ezerfős minta sok dimenzióra tört keresztábráinak celláiba valószínűleg egy nagyon alacsony elemszám kerülne. A Big Data gyakorlatilag korlátlan, vagyis a valósággal megegyező méretű mintájában biztosan megkapjuk a helyes eredményt, már ha van ilyen eredmény egyáltalán.

Fogalmi szinten tehát a Big Data valódi jelentősége – mint arra már a Durkheim–Tardé-vita ismertetése során utaltunk – az, hogy a holisztikus társadalmi fogalmak dekonstrukcióját kínálja. Ezek az átfogó, absztrakt fogalmak általában meghatározzák a gondolkodásmódunkat. Pedig az olyan holisztikus fogalmak, mint a „társadalom”, a „fogyasztó”, a „nők”, a „választópolgár”, nem is beszélve a társadalmi tudatról, leegyszerűsítők, mert eltakarják az általuk reprezentált valóság sokszínűségét. Nem sokat javít a helyzeten, ha a hagyományos fogalmakkal és technikákkal tovább bontjuk ezeket a terminusokat. Amikor „városi idősekről”, „bizonytalan szavazókról”, „középosztálybeli nőkről” vagy „egyetemista fiatalokról” beszélünk, e tovább bontott kategóriák éppoly esetlegesek és üresek, mint azok a kategóriák, amelyekből kiindultunk.

Nyilvánvaló, hogy ezek a címkék meglehetősen különböző emberek konglomerátumait, hálózatait takarják. Ideális esetben és ha jól használjuk, a Big Data segíthet megragadni, értelmezni és megjeleníteni a kapcsolatoknak és végső soron az életnek a sokszínűségét és összetettségét azáltal, hogy lehetővé teszi a nagy számosságú szereplő azonosítását és a közöttük levő kapcsolatok feltárását.<sup>4</sup>

Érdekes ezeket az absztrakt sajátosságokat egy konkrét példán szemléltetni. 2012 májusában az Amerikában élő egyiptomi újságíró, Mona El Tahavi Miért gyűlölnék minket címmel publikált egy cikket a nők helyzetéről az arab országokban (El Tahavi 2012). A cikk nagy vihart kavart, ezúttal arra használnánk, hogy bemutassuk a Big Data alapú megközelítés sajátosságait, előnyeit.

A hagyományos megközelítésben tartalomelemzésnek vethettük volna alá a kiválasztott arab nyelvű hírforrásokat, szerencsés esetben és bőséges anyagi források birtokában beemelve ezek mellé az angol és más releváns nyelvű kiadványokat. Ennek eredményeképp valószínűleg egy olyan tipizációt kaptunk volna, amely a különböző, általunk alkotott kategóriák előfordulási gyakoriságát mutatta volna. Ezt követően, ha igazán szerencsésnek voltunk a fundraisingben, egy survey-jel vizsgálhattuk volna az egyes országokban a cikk által kiváltott érzelmeket, esetleges attitűdváltozásokat. Maga a folyamat jó pár évet vett volna igénybe, és több szempontból is bizonytalan, hiányos eredményeket hozott volna. Megértési lehetőségeinket beszűkítették volna az adatfelvétel korlátai, eredményeink pedig csak a vizsgált országokra – nem pedig a globális közvéleményre – vonatkoztak volna. Egy rendkívül szenzitív témában a különböző normákhoz való igazodás bonyolította volna a válaszadások nyomán kialakult képet. Nagyon kevés kutatóintézet rendelkezik olyan anyagi erőforrásokkal, hogy egy ilyen kutatást megvalósíthasson – és a kapott eredmény legalábbis kérdéses lenne.

---

<sup>4</sup> A korábbi kommunikációs paradigmában gyakorlatilag semmit nem tudtunk arról, hogy üzeneteink mennyiben érték el a kívánt célcsoportot. A Big Data alapú megközelítés megkönnyíti az üzenetek nyomon követését. Ha szükséges, a hatékonyság növelése érdekében megváltoztathatjuk a terjedés útvonulatát, sőt menet közben akár az üzenet tartalmát is személyre szabhatjuk.

Ezzel szemben mit kínál a Big Data alapú megközelítés? El Tahavi cikke hatalmas érzelmi hullámokat vert, szerencsére a Twitteren is. Azért nevezhetjük ezt szerencsének, mert a Twitter nyitott struktúrája (jelenleg még) lehetővé teszi az itt zajló folyamatok kutatását (Lieberman 2014; Hannah 2012).

A blogger Alex Hannah-t szintén a cikk által kiváltott reakciók érdekelték. Ezért, bár nem rendelkezett különösebb erőforrásokkal, legyűjtötte a cikk hashtagjéhez kapcsolódó adatforgalmat. Félrevinné a jelen gondolatmenetet, ha itt részletesen értekezni az arab országok női szerepekhez kapcsolódó gyakorlatáról. Itt pusztán az előbbieken ismertetett módszertani sajátosságok szemléltetésére szorítkozom.

1. *Időbeliség.* A Twitter-vihart a maga azonnalóságában lehetett vizsgálni: Hannah például 2012. április 24-én délután 12:21 és 15:33 között gyűjtötte az adatokat, amelyeket blogján azonnal publikált is.

2. *Viselkedés vs. vélemény.* Habár ezek az adatok is textuális reprezentációkat dolgoznak fel, mármint tweeteket, de nem survey-kérdésekre adott válaszok, hanem cselekvések abban az értelemben, hogy az üzenetek küldői és továbbküldői bizonyosan szándékaik szerint cselekedtek. Vagyis lehetséges, hogy az üzenetek tartalma itt is homályos és inkonzisztens, de abban biztosak lehetünk, hogy az üzenet megfogalmazásának és elküldésének szándéka valós.

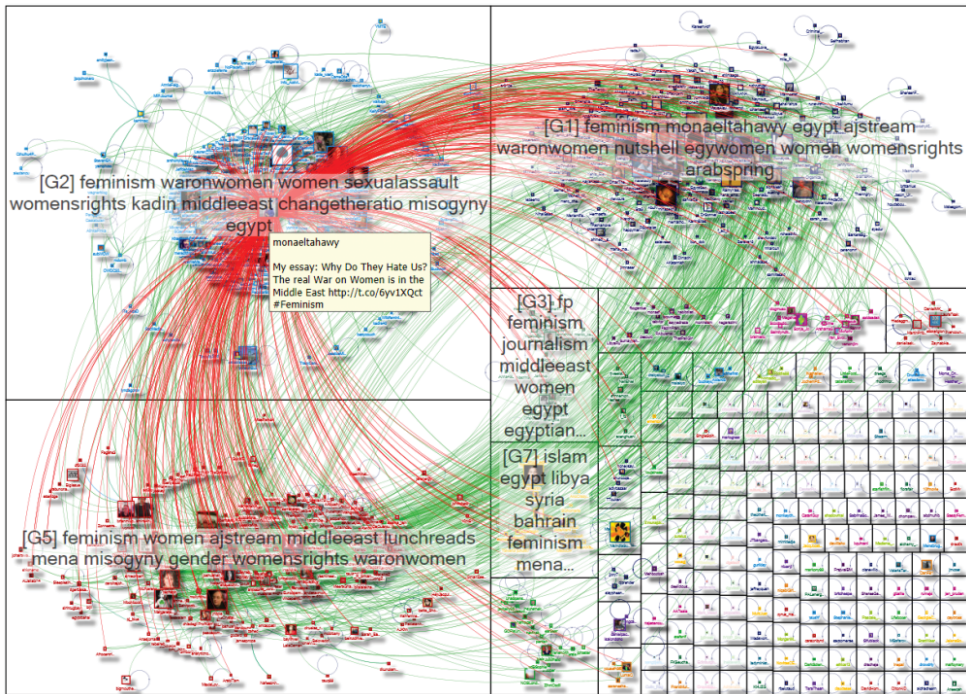
3. *Strukturált és nem strukturált adatok.* A Twitter-üzenetek estében is strukturált (idő, hely) és nem strukturált (szöveges tartalom) adatokat találunk. Ezáltal lehetőségessé válik a diskurzus dinamikájának a rekonstruálása is: kinek milyen típusú üzenetére hogyan és mikor reagáltak különböző csoportok.

4. *Hozzáférés.* A tweetekre való fókuszálás lehetővé teszi, hogy a kutató – kilépve a nemzetállam kereteibe záró megközelítésből – az áramlások globális terében zajló információterjedést rekonstruálja. Ráadásul a Twitter-adatok bárhonnán hozzáférhetőek, s ahogy lehetséges Amerikából elemezni egyiptomi adatokat, ennek esélye fordítva is adott, jóval kisebb belépési, marginális költségek mellett, mint a hagyományos kutatások esetében.

5. *Valódi struktúrák.* Az adatgyűjtés és -vizualizáció révén pontosan láthatóvá válnak a személyek, illetve ezek egymáshoz való viszonya. Ennek alapján definiálhatjuk a véleményvezéreket, a köréjük szerveződő csoportokat, illetve ezen csoportok egymáshoz fűződő viszonyát, kapcsolataik rendszerét. Ezáltal láthatóvá válnak olyan nem triviális szerkezetek is, mint például az, hogy az Occupy-mozgalom egyes alakjai, illetve a kurd emberi jogi mozgalomhoz kapcsolódó csoportok milyen álláspontot képviseltek ebben a vitában (bővebben lásd Hannah 2012).

Ezen eredmények interpretációja kapcsán fokozott óvatossággal kell eljárunk. Ahogy arra a social média forradalmak kapcsán sokan felhívták a figyelmet, az arab országokban jelenleg a lakosság egy elenyésző hányada használja a Twittert (Gladwell 2010; Morozov 2011; Fuchs 2012). Ezért pontosan kell látnunk az ilyen kutatások értelmezési tartományát, illetve érdemes fokozott gondossággal megfogalmazni kutatási kérdéseinket. A Twitter-alapú kutatások nem reprezentálják az adott társadalom nézeteit, de – mint az előbbieken igyekeztünk bizonyítani – leképezik az adott vitában részt vevők véleményét és magatartását. Ha nem keverjük össze a második állítást az elsővel, érvényes, fontos és újszerű eredményeket kapunk.

1. ábra. Az El Tahavi cikke nyomán kirobbant Twitter-„vihár” vizualizációja



Forrás: Blog – Alex Hanna<sup>5</sup>

### Láthatjuk-e a jövőt?

A Big Data előbbi öt szociológiai jellemzője vezet az új paradigma egy hatodik, talán legmeghökkenőbb eredményéhez: a jóslatok, megalapozott előrejelzések képességéhez. A Big Data alapú prediktív analitika része lett napjaink valóságának. A Target nevű üzletlánc közismert és gyakran idézett példája tökéletes metaforaként mutatja az új helyzetet. A minnesotai Target analitikus algoritmistái egy tizenéves lány vásárlási szokásainak elemzéséből statisztikai módszerek segítségével már az első trimeszterben kimutatták, hogy a vásárló terhes.

Egy misztikus formula segítségével, mert magasabb arányban vásárolt illatmentes testápolót, ásványi étrend-kiegészítőket és nagyzacskós vattagolyókat. A Target ezért elkezdett babafelszerelések vásárlásához kedvezményt biztosító kuponokat küldözgetni vásárlójának, apja legnagyobb megdöbbenésére, aki az ő elenyésző emberi következtetési hatalmával még mit sem tudott a várható családi eseményről (Ellenberg 2014).

5 <http://blog.alex-hanna.com/2012/04/visualizing-the-polarized-discourse-of-why-do-they-hate-us>

Nyilvánvaló, hogy az üzleti intelligencia új szintre ért el a Big Data módszereinek alkalmazásával. Számunkra azonban sokkal fontosabb kérdés az, hogy a szociológia képes lesz-e kihasználni ezt a potenciált.

Azáltal, hogy az adatgyűjtés, -feldolgozás és az összefüggések feltárásának költségei a digitalizációval drámaian csökkentek, lehetségessé vált nagyszámú változó hatásainak vizsgálata hatalmas populációkon is. Ennek nyomán megnyílik az esélye annak, hogy eddig nem ismert mintázatokat (megnövekedett vattagolyó-vásárlás – teresség) kutassunk, és ez jelentősen megnöveli a predikció, az emberi viselkedés előrejelzésének esélyét is. A fizikus Barabási Albert László *Villanások* (2010) című könyvének alcímében egyenesen azt állítja: a jövő kiszámítható. Ez az állítás nem kis kihívást jelent mindazoknak, akik hisznek az ember döntési szabadságában és e döntések egyediségében, kiszámíthatatlanságában. Mégis a két álláspont nem összeegyeztethetetlen: szabadon és egyénileg döntünk, de meglehetősen kiszámíthatóan (Duhigg 2014).

Arra a felvetésre, hogy mennyiben releváns a Big Data a szociológia hagyományos nagy kérdéseinek megválaszolására, többféle válasz is adható. A mi megközelítésünkben a Big Data kapcsán – tudásszociológiai analógiával – egy gyenge és egy erős program is meghirdethető. A Big Data gyenge programjában a digitális lábnyomokból nyilvánvalóan kirajzolódnak fő csapásirányok: a kulturális fogyasztás, közösségimédia-használat, ifjúsági szubkultúrák, társadalmi és politikai mozgalmak témáihoz kapcsolódó kutatásokban meglehetősen kézenfekvő a Big Data használata – már ha az adatok elérhetőek és rendelkezésre állnak. A Big Data, a mi elképzeléseinkhez közelebb álló, erős programja ennél többet mond, nagyobb magyarázóerővel és igényekkel lép fel. Ez a megközelítés nem az interneten és a közösségi hálózatokban generált adatokból indul ki, hanem a korábban Tarde és Latour nyomán ismertetett episztemológiai fordulatból. A Big Data erős programja szerint a digitalizált adatok felszabadítják gondolkodásunkat a korábban kényszerűségből használt leegyszerűsítő, homogenizáló és holisztikus fogalmak használata alól. Láthatóvá, megragadhatóvá és elemezhetővé teszik azokat az egyéni döntésekből és viselkedésből kialakuló hálózatokat és kapcsolat mintázatokat, amelyek ott vannak valamennyi leegyszerűsítő fogalmunk mögött. A Big Data megteremti a dinamikus, módszertani individualista és hálózati szemléleten alapuló elmélet gyakorlásának tárgyi, empirikus feltételeit. A Big Data erős programja tehát a társadalomtudományok módszertani, elméleti és gyakorlati paradigmaváltását vetíti előre. Ennek alapján nemcsak lehetséges a szociológia hagyományos nagy kérdéseinek, a társadalmi egyenlőtlenségek, a kirekesztés, a hatalom gyakorlásának vagy a társadalmi tőkének a vizsgálata a Big Data eszközeivel, hanem éppenséggel ezáltal nyílik meg a dinamikus szemléletű, a valódi kapcsolatokat és hálózatokat bemutató megközelítés előtt az út. Lehet, hogy ehhez az adatrobbanás ellenére sem rendelkezünk ma még elég adattal, de ennek a folyamatnak még az elején járunk, és a magunk részéről biztosak vagyunk abban, hogy ebbe az irányba haladunk.

Csak egyetlen példát felvillantva: a fiatalok radikalizálódásáról számos tanulmány jelent meg az utóbbi időben. Képzelnék el egy pillanatra, hogy ezt a kérdést nem hagyományos vagy online survey kérdőívek segítségével vizsgáljuk, hanem Big Data alapon összesítjük a nemzeti rockegyüttesek rajongói táborát. Itt nem csak azt tudhatjuk meg, hogy ennek a közösségnek milyen a szerkezete, a kommunikációs hálózata, földrajzi megoszlása, időbeli aktivitása, hogy milyen típusú információk milyen mintázatokban terjednek, hanem azt is, hogy az egyes aktivitásokhoz milyen affinitások társulnak (kik azok, akik a humorra reagál-

nak, kik azok, akik a brutalitásra, és milyen esetleges kapcsolat van e tematikák és szereplők között). Hálózati mintázatokat kiegészítve szövegelemzéssel az egyes tematikus csomópontok regisztereinek alakulását is megvizsgálhatjuk (milyen hálózatokban erős például a történelmi sérelmek említése, és hol dominánsak ehhez képest a rasszista tartalmak). Mindez kiegészíthető a további tartalmak, ízléspreferenciák és tevékenységek elemzésével (szeretik-e a fekete rapet vagy a kosárlabdát), illetve hogy mikor, milyen gyakran, kivel kommunikálnak, és kinek a véleményét tekintik más közösségekben mérvadónak. Mindezt nem egyszeri adatfelvétellel, hanem időbeliségében folyamatosan mérhetjük.<sup>6</sup>

Az egyre inkább virtualizálódó társadalmi lét digitális lábnyomai elemzésének eszköze a Big Data alapú kutatás. A szociológusoknak, akik a valóság különböző területeinek, az oktatás, a vallás, a gazdaság, a városi élet és rétegződés vizsgálatára szakosodtak, szembe kell néznünk azzal a kilátással, hogy elveszítjük a tudásmonopóliumunkat az új adatbányászokkal szemben, akik nem társadalomtudományi fogalomrendszerből indulnak ki, hanem hagyják, hogy az algoritmusok működése révén az adatokból bontakozzék ki a számukra megjelenő valóság. Azonban ezt a viszonyt nem csak kiszorításként, hanem egy új kooperáció és együttműködés kezdeteként is fel lehet fogni, ahol a szociológia nem a semmiből indul.

A Big Data megjelenése olyan változást jelent a társadalomtudományokban, mint a mikroszkóp felfedezése a biológiában. Ahogy Latour írja: „Ha megváltoztatjuk a megfigyelés eszközét, az egész társadalomelmélet átalakul” (Latour 2010: 155). Láthatóvá válnak olyan korábban rejtett hálózatok, mintázatok és működési elvek, amelyeket addig nem ismertünk, s mélyebben és pontosabban ismerhetjük meg a társadalmi valóságot. Ez a hagyományos intézményi szerkezetek működését is átalakítja, mint ahogy a vírusok, baktériumok és mikrotestek felfedezése is megváltoztatta az orvostudományt és a biológiát. A Big Data a nanogyógyászatától az űrkutatásig, a kultúrakutatástól a kereskedelemig, a politikától a párválasztás szabályainak megértéséig lehetőséget kínál a pontosabb megismerésre. Ezért a Big Data alighanem maga mögött hagyja a tinédzser szexet övező bizonytalanságok Ariely által leírt jellegzetességeit, és a huszonegyedik század egyik legnagyobb intellektuális kihívásként sok mindent megváltoztat majd – beleértve a társadalomtudományokat is.

## Hivatkozott irodalom

- Anderson, Chris (2007): The End of Theory. The Data Deluge Makes the Scientific Method Obsolete. *Wired* 16(7).  
Interneten: [http://archive.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory).
- Barabasi, Albert-Laszlo (2010): *Bursts. The Hidden Patterns Behind Everything We Do, from Your E-mail to Bloody Crusades*. New York: Dutton.
- boyd, danah és Kate Crawford (2012): Critical Questions For Big Data. *Information, Communication & Society* 15(5): 662–679.
- Castells, Manuel (1996): *The Rise of the Network Society. The Information Age: Economy, Society and Culture, Vol. I*. Cambridge, Mass. – Oxford, UK: Blackwell.
- Castells, Manuel (1997): *The Power of Identity, The Information Age: Economy, Society and Culture, Vol. II*. Cambridge, Mass. – Oxford, UK: Blackwell.

---

6 A jövő egyik legfontosabb kulcskérdése, hogy a szociológusoknak és társadalomkutatóknak lesz-e az ilyen adatokhoz hozzáférése, vagy az megmarad néhány amerikai nagyvállalat és a titkosszolgálatok ellenőrizhetetlen privilégiumának. Ezt a kérdést most nem kívánjuk megválaszolni, csak azt akartuk jelezni, hogy az adatok birtokában egy új epiztemológiai fordulatra nyílik mód.

- Castells, Manuel (1998): *End of Millennium. The Information Age. Economy, Society and Culture, Vol. III.* Cambridge, Mass. – Oxford, UK: Blackwell.
- Csepeli György és Prazsák Gergő (2004): Paradigm Change in Sociology. *Review of Sociology* 10(2): 39–46. Interneten: [http://www.szociologia.hu/dynamic/RevSoc\\_2004\\_2\\_CsepeliGy\\_Paradigm\\_change.pdf](http://www.szociologia.hu/dynamic/RevSoc_2004_2_CsepeliGy_Paradigm_change.pdf).
- Csepeli György és Dessewffy Tibor (2015): Big Data. A Technological Change That Will Fulfil Sociology. *Review of Sociology* (megjelenés előtt).
- Duhigg, Charles (2014): *A szokás hatalma.* Budapest: HVG.
- Durkheim, Émile (1978 [1895]): A szociológia módszertani szabályai. In uő *A társadalmi tények magyarázatához. Válogatott tanulmányok.* Budapest: KJK, 1978, 21–159.
- Ellenberg, Jordan (2014): What's Even Creepier Than Target Guessing That You're Pregnant? *Slate* (2014. 06. 09.). Interneten: [http://www.slate.com/blogs/how\\_not\\_to\\_be\\_wrong/2014/06/09/big\\_data\\_what\\_s\\_even\\_creepier\\_than\\_target\\_guessing\\_that\\_you\\_re\\_pregnant.html](http://www.slate.com/blogs/how_not_to_be_wrong/2014/06/09/big_data_what_s_even_creepier_than_target_guessing_that_you_re_pregnant.html).
- Fuchs, Christian (2012): Some Reflections on Manuel Castells' Book Networks of Outrage and Hope. *Social Movements in the Internet Age. tripleC* 10(2): 775–797. Interneten: <http://www.triple-c.at/index.php/tripleC/article/download/459/433> (letöltve: 2014. 12. 21).
- Gladwell, Malcolm (2010): SMALL CHANGE: Why the Revolution Will Not Be Tweeted. Twitter, Facebook, and Social Activism. *New Yorker* (október 4.).
- Hannah, Alex (2014): *Visualizing the Polarized Discourse of "Why Do They Hate US?"* Interneten: <http://blog.alex-hanna.com/2012/04/visualizing-the-polarized-discourse-of-why-do-they-hate-us/> (letöltve: 2014. 12. 28.).
- Hunt, Lynn (2008): *Measuring Time, Making History.* Budapest – New York: CEU Press.
- Hurwitz, Judith, Alan Nugent, Fern Halper és Marcia Kaufman (2013): *Big Data for Dummies.* Hoboken, NJ: Wiley.
- Latour Bruno (2002): Gabriel Tarde and the End of the Social. In *The Social in Question. New Bearings in History and the Social Sciences.* Patrick Joyce (szerk.). London: Routledge, 117–132.
- Latour, Bruno (2010): Tarde's Idea of Quantification In *The Social after Gabriel Tarde. Debates and Assessments.* Matei Candea (szerk.). New York: Routledge, 143–162.
- Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastien Grauwin és Dominique Boullier (2012): 'The Whole is Always Smaller Than its Parts'. A Digital Test of Gabriel Tarde's Monads. *The British Journal of Sociology* 63(4): 590–615.
- Lieberman, Michael (2014): *Visualizing Big Data. Social Network Analysis.* (Konferencia-előadás, Digital Research Conference, San Antonio, Texas, 2014 márciusa).
- Morozov, Evgeny (2011): *The Net Delusion. The Dark Side of Internet Freedom* New York: Public Affairs.
- Mayer-Schönberger, Viktor és Cukier Kenneth (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think.* London: John Murray.
- Rifkin, Jeremy (2010): *The Empathic Civilization.* New York: Jeremy P. Tarcher.
- Scott, James C. (1999): *Seeing Like a State. How Certain Schemes to Improve the Human Condition Have Failed.* New Haven, Conn.: Yale University Press.
- Savage, Mike és Roger Burrows (2007): The Coming Crisis of Empirical Sociology. *Sociology* 41(5): 885–899.
- Tarde, Gabriel (1898): *Social Laws. An Outline of Sociology.* New York: Macmillan.
- Tavris, Carol és Elliot Aronson (2008): *Mistakes Were Made (But Not by Me). Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts.* Boston: Mariner Books.