

DESSEWFFY TIBOR  
**DIGITÁLIS SZOCIOLÓGIA**

DESSEWFFY TIBOR

# **DIGITÁLIS SZOCIOLÓGIA**

SZOCIOLÓGIAI KÉPZELET A DIGITÁLIS KORBAN



**TYPOTEX**

A könyv kiadását a Magyar Tudományos Akadémia támogatta.



© Dessewffy Tibor, Typotex, Budapest, 2019  
Engedély nélkül semmilyen formában nem másolható!

ISBN 978 963 493 064 8

Kedves Olvasó!

Köszönjük, hogy kínálatunkból választott olvasnivalót!  
Újabb kiadványainkról, akcióinkról a [www.typotex.hu](http://www.typotex.hu)  
és a [facebook.com/typotexkiado](https://facebook.com/typotexkiado) oldalon értesülhet.

Typotex Kiadó

Alapította Votisky Zsuzsa, 1989

A kiadó az 1795-ben alapított Magyar Könyvkiadók  
és Könyvterjesztők Egyesülésének tagja.

Felelős kiadó: Németh Kinga

Főszerkesztő: Horváth Balázs

Felelős szerkesztő: Kovács Zoltán

Műszaki szerkesztő: Teklits Tamás

Borítóterv: Gödöny Tamás

Nyomta és kötötte: Generál Nyomda Kft., Szeged

Felelős vezető: Hunya Ágnes

# TARTALOM

<i>Előszó</i>	7
<b>1. ELMÉLETEK</b>	13
A Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon	15
Ápoljuk (digitális) kertjeinket	30
Digitális Pareto. Az online közösségeken belüli megoszlás mintázatairól	33
Teória és praxis az információs korban	46
Egyetem 2.0? Felsőoktatás az információs korban	57
<b>2. VALÓSÁGDARABKÁK</b>	73
Hírnév, rajongók és Facebook. Magyar hírességek és rajongóik	75
„Nélküled az élet elképzelhetetlen számomra! A példaképem vagy!”	
Mikrocelebek rajongói Magyarországon	99
Veszedelemes viszonyok? Egy onlife botrány elemzése	119
Hol lakoznak belső angyalaink? Szubaltern ellennyilvánosságok a menekültválság idején Magyarországon	145
Facebookon született. A menekültválság és a közösségi önszerveződések Magyarországon	168
Harry Potter, avagy a politikai bölcsek köve?	190
Kedvelik-e a nemzeti rockerek a McDonald'sot?	213
Rész és Egész, avagy kik alkotják a nemzetirock-szubkultúrát	232
„A hálózat, te buta!” Social networking, a millenniumi generáció és az Obama-kampány	248
Levélféle az ökocídium lovasához	264
<i>Hivatkozott irodalom</i>	268

# ELŐSZÓ

## *A kíváncsi dinoszauros*

Egy új területtel, a digitális szociológiával való ismerkedési folyamat stációit kínálja ez a könyv. Arra a brutális és mindent elérő átalakulásra próbál reflektálni, amit a digitális adatok robbanásszerű növekedéséből vezetek le ezekben a tanulmányokban, és amely messze nem csak, sőt elsősorban nem mennyiségi kérdés. Épp ellenkezőleg, mindent, a gazdaságot, kultúrát és politikát is átalakító folyamatokat indít el vagy erősít fel. Ebből a változásból, átalakulásból a társadalomtudományok sem maradnak ki. Bár látom azokat a veszélyeket, fenyegetéseket és kihívásokat, amelyeket a digitális adatrobbanás, az adatbázisok és az új modellezési technikák jelentenek, mégis azt hiszem, hogy ez a társadalomtudományok számára új, most megnyíló horizontokat is jelent. Nem mennék bele abba, hogy mi a társadalomtudományok hivatása úgy általában, erről jelenleg épp elég sok vita zajlik ahhoz, hogy ne érezzem azt, hogy ebben nekem állást kellene foglalnom. Inkább csak azt tudom, hogy mit jelentett és mit jelent a szociológia: a körülöttünk dübörgő társadalmi folyamatok és változások értelmezésének, kicsit ambiciózusabban a megértés igényét.

Ebben az összefüggésben az új korszak nem is annyira új. Attól, hogy a digitális adatrobbanás átalakítja az életviszonyok szerkezetét, illetve hogy a szociológiai megismerés számára a digitális lábnyomok elemzése új utakat nyit meg, a társadalmi létbe ágyazott egyén alapproblémái ugyanazok maradnak. Megszületünk, szenvedünk, szeretünk, keressük a helyünket, próbálunk értelmesen élni, hinni valamiben és megnyugvással meghalni. Az emberi létezés lényege nem változik, még ha a közeg, a kompetenciák elosztása és szerkezete, az előrelépés feltételei ebben az új kontextusban teljesen átalakulnak is.

Az állandóság és változás furcsa, dinamikus ölelkezése ezt a könyvet is áthatja. Tematikusan számtalan olyan témával foglalkozom, amelyekkel korábban is. Reználtan veszem észre, hogy már nem években, hanem évtizedekben kell mérnem a korábbi írásaimra, könyveimre való visszatekintést. Mivel nem vagyok

önmagam bibliográfusa, ebben a vonatkozásban kevésbé érdekel a múlt. Csak a jelen kötet összeállítása kapcsán tűnt fel, hogy az itt tárgyalt témák milyen komoly átfedésben vannak első, 1997-ben megjelent, *Kedélyes labirintus* című könyvemmel. Ebben tárgyaltam az idegenség kérdését, Friderikusz Sándorhoz írt levelek elemzésén keresztül a rajongás természetrajzát vagy a hétköznapi fogyasztás jellegzetességeit az államszocializmus időszakában, a népi hős fogalmának kialakulását, valamint a tömegkultúra és politika viszonyát.

Ezek a témák, ha más köntösben is, de mind felbukkannak ebben a könyvben. Ami a különbséget jelenti, hogy húsz éve ezek elsősorban spekulatív esszék voltak esetenként kvalitatív módszertannal megtámogatva. Ma a digitális módszerek lehetővé teszik, és remélem, az Olvasó is osztja majd ezt a vélekedésemet, a korábbiaknál merészebb kérdések megfogalmazását és megválaszolását. Számomra ezt jelenti az adatalapú társadalomtudomány: a szociológiai képzelet kiteljesedésének lehetőségét. Ez egyelőre valóban csupán lehetőség – túlságosan is gyakran találkozunk olyan tanulmányokkal, amelyek hatalmas adatapparátust mozgatva érdektelen, közhelyes megállapításokhoz lyukadnak ki. Ez azonban csak gyermekbetegség, amelynek tüneteit, mint oly gyakran máshol is, majd kiheverjük. A technológia által megnyitott lehetőségek belakásához, értelemmel telítéséhez nekünk, gyarló embereknek sok időre van szükségünk.

Miközben tehát a kötet írásait egyben tartja a digitális lábnyomokból építkező módszertan határainak kitapogatása, témáikban az írások szerteágazóak. Tudom, hogy ez szigorúan vett szakmai, professzionális értelemben öngyilkos stratégia. A standardizált és paraméterezett tudomány felfogásában az előrejutás és siker receptje az, ha az ember idejekorán lecövekel egy téma mellett, és abban mélyed el, publikál, a szakterület számára releváns konferenciákon vesz részt, és épít networköt. Hankiss Elemér egy kései művében megbánással és hiányérzettel beszélt saját életútja cikkcakkjairól, amelyek az ő értelmezésében megakadályozták azt, hogy igazán komoly életművet hozzon létre. Őszintén felkavart, amikor ezeket a gondolatait olvastam. Egyrészt azért, mert Hankiss számomra mindig is inspiráló volt, és úgy gondolom, hogy életműve a magyar szociológia egyik legfontosabb darabja. Mint az emlékezetére szerkesztett kötetből is tudható, ezzel a véleményemmel nem vagyok egyedül.<sup>1</sup> Hankiss hatásának az egyik oka, hogy a megismerés és megértés iránti vágy folyamatosan új területekre vezérelte. Gondolatait olyan nyelven fogalmazta meg, amellyel nem a legszűkebb szakmai nyilvánosságot, hanem a nagyközönséget célozta meg.

---

<sup>1</sup> *Hankiss Elemér-émlékkönyv*, Helikon, 2018

De ebben a vonatkozásban Hankiss csalódása és vágyai tőlem mégis távol állnak. Az utókor ítélete, az örökség, a fennmaradó életmű darabjai – számomra nem motiváló fogalmak. Nem érdekel, hogy fennmarad-e életművem, túlélnek-e engem írásaim – az ilyen irányú ambíciók hiánya segít elviselni az intézményi kívülállásból és a csapodár intellektuális érdeklődésből fakadó hátrányokat. Ebben az értelemben a szellemi munka önmagában hordja a jutalmát, egy probléma körüljárása, megfogalmazása, egy sajátos megfejtés megtalálása és megírása – maga az elnyert díj. Azt persze remélem, hogy a rejtvényfejtés örömeiből, ebből a „nyereményből” az Olvasó is részesedik majd e könyvet olvasva.

A digitális szociológia egyik sajátossága, hogy olyan komplex tudás mozgását kívánja meg, amely különböző szerzői kooperációkban ölt testet. Ezért a szokásos illetmeni köszönetnyilvánításokon túl szeretném megköszönni mindazok segítségét, akik az e kötetben található tanulmányok megírásában közreműködtek.

Elsőként is volt diákjaimnak, Mezei Mikesnek és Váry Dánielnek, akikben megvan az intellektuális kalandvágy, a járatlan utak bejárásához kellő kíváncsiság, valamint az empirikus adatfeldolgozáshoz és értelmezéshez szükséges készségek birtoklása. Mikes és Dani a jövő, olyan szociológusok, akik megtestesítik az új módszerek ismeretét a tudományos problémák iránti fogékonysággal. Idesorolom Naszály Natáliát is, aki Gurály Sárával egyetemben fontos közreműködője volt egyes fejezetek elméleti megalapozásának. Ami közös mindannyiukban, és beszédes a magyar társadalomtudományok helyzetével kapcsolatban, hogy a piaci szereplők rögvést felszippantottak tudásukat. Hálás vagyok azért, hogy napi munkájuk mellett módot találtak arra, hogy ezekben a kutatásokban aktívan részt vegyenek.

Mára elismert kollégámmá vált Nagy Zsófia, a kötet két írásának a társszerzője – bár ennél ambícióink egykor nagyra törőbbek voltak. De Zsófi megértette velem, hogy az együtt munkálkodás különös adomány – örülni kell annak, amikor lehetséges, de nem kell szomorkodással tölteni az időt, ha valami miatt éppen nem működik.

A kötetben helyet kapott két korábbi írásom, amelyek fontos háttéréül szolgálnak az adatforradalomnak. Ezek társszerzői, Ravasz Ábel és Ságvári Bence már különböző életutakon haladva sikeres szereplői a maguk világának (Bence szociológus, Ábel e sorok írásakor épp a szlovák kormány tagja mint romaügyi kormánybiztos). Az ő példájuk is megélhetővé teszi azt az örömet, amit az ember egy fiatal tehetség beérésének és későbbi sikereinek kapcsán érez – csak remélni tudom, hogy a velem való munkakapcsolatukból több intellektuális előnyük származik, mint amennyi egzisztenciális hátrányuk.

Solymos András matematikus a klaszterezési eljárásokhoz nyújtott elengedhetetlen segítséget. A közgazdász Láng Lászlóval folytatott eszmecsereimből sokat tanultam, és ezeket a meglátásokat több tanulmányba is beépítettem. Szeretnék végül köszönetet mondani Kovács Zoltánnak, aki heroikus munkával simította ki a kézirat gyakran összegubancolódott szálait.

Nem a fatalizmus, hanem a szellemi produktumok természete és az élet hiábavalóságának elfogadása tesz rezisztenssé saját intellektuális lábnyomom fennmaradására vonatkozóan. Ebből fakad az a zavar is, ami akkor fog el, ha az engem ért intellektuális hatásokat kellene itt felsorolnom. Az évek során ki tudja már, hogy kollégáimmal és barátaimmal folytatott beszélgetéseink és vitáink éppen milyen módon befolyásolták gondolkodásomat. Ha felismerik egy-egy beszélgetésünk hozadékát ezekben az írásokban, kérem, gondoljanak rám megértéssel, és ha lehet, szeretettel.

A személyes hatások reménytelen listázása helyett hadd említsek három olyan intézményt, amely meghatározó volt e könyv kapcsán. Az első az ELTE Digitális Szociológia Kutatóközpontja, amelyet partizánakcióként működtetünk Nagy Zsófi-val és a tehetséges diákokkal már évek óta. Az itt folyó kutatások, beszélgetések és rendezvények bizonyosan formálták gondolataimat.

A másik fontos hatás az ELTE Társadalomtudományi Karán tartott kurzusaimhoz kapcsolódik. Ezek témáikban többnyire a digitális átalakulás, a digitális kultúra és a diszrupció témakörével foglalkoznak. Ahogy a bevezető előadások alkalmával mindig elmondom, ezek perverz struktúrájú kurzusok: a hallgatók ugyanis sokkal többet tudnak nálam az adott témákról benne élve a közösségi médiában és követve a legújabb fejleményeket. Így nem túlzás azt mondani, hogy visszajelzéseikből és dolgozataikból rengeteget tanulok.

A harmadik intézmény, amely elengedhetetlen támogatást nyújtott a kötet empirikus alapjaihoz, a Datalyze digitális marketingügynökség, amely egyben az ELTE egyik első társadalomtudományos *spin of* cége. Az itt kifejlesztett speciális adatgyűjtő és kategorizációs eljárások tették lehetővé a digitális lábnyomok feldolgozását. Biztos vagyok benne, hogy az ilyen kutató és piaci cégek közötti együttműködés a jövő. Külön köszönet a céget vezető Sütő Annának, amiért támogatta a kutatással járó erőforrás-lekötést.

Végül köszönöm Áginak és Dedének, hogy elviselték, amikor olyan sok közös játékból vontam ki magam a dolgozószoba ajtaját magamra zárva. Egy család élete mindig bonyolult társasjáték – a nélkül a szeretetáramlás nélkül, amit mi kreáltunk ebben az „igen igen kemény, kemény világ”-ban, én nem maradtam volna fenn, ergo ezt a könyvet sem tarthatná kezében a nyájas Olvasó.



A kötet fő üzenete, hogy a digitális adatrobbanás és az azt kísérő módszertani megújulás a társadalomtudományi megismerés számára is teljesen új lehetőségeket nyit meg.

A könyv két részre tagolódik.

Az első részben a Big Data elméleti, episztemológiai és társadalmi kérdéseit tárgyalja a szerző: különös tekintettel a korábbi társadalomtudományi kánonhoz való kapcsolódás szempontjára.

A második részben az első rész elméleti tételeit empirikus tanulmányok illusztrálják. E tanulmányok a *social media* (közösségi média) lábnyomaiból építkező innovatív módszertan alapján szerteágazó, de a hagyományos eszközökkel nehezen megválaszolható kérdésekre prezentálnak empirikus eredményeket.

Van-e a magyar celebek rajongói között eltérés a környezettudatosság szempontjából?

Az ún. mikrocelebek követői milyen klaszterekre bonthatók, és ezeket mi jellemzi?

Milyen közéleti preferenciákat találunk egyes Facebook-csoportoknál?

A Harry Potter-rajongók milyen értékmintázatokat mutatnak, és tapasztalható-e jelentősebb közéleti érdeklődés a körükben?

Találunk e „kozmpopolita” fogyasztási mintázatokat a nemzeti rock követői körében?

Melyek voltak a 2008-as Obama-kampány közösségi médiában szerveződő sajátosságai?

Vannak-e a digitális korszakváltásnak ökológiai következményei?

Kevés lejáratosabb klisé van annál, mint hogy szülőként az ember mindent máshogy lát. Igaz ez a digitális adatrobbanás nyomán végbemenő változásokra is – bizonyos megszorításokkal. Mert nyilván nagyobb szorongással figyeli az ember ennek a világnak a változásait, ha tudja, hogy amit ma elképzelsz vagy megsejtesz, az holnapra gyermeke számára magától értetődő valóság lesz. Ugyanakkor, ha úgy tetszik, a hankissi szabadságeszmény követőjeként abban biztos vagyok, hogy nincsenek társadalmi, technológiai determinizmusok. Legyen szó mesterséges intelligenciáról, prediktív modellekről, algoritmikus hatalomról, társadalmi *scoring*ről, a privát szféra határainak leomlásáról vagy a platformok korlátlan adathalászatáról – ezek mind nyitott végű történetek. Nyitott végűek abban az értelemben, hogy emberi cselekvések, innovációk és közösségi érdekképviseletek megvívott konfliktusai nyomán fog az új helyzet előállni. Más szempontból ezért sokan sokféle árat fizetnek majd. Ám a szociológiai képzelet számára, amely eredetileg a körülöttünk levő világ előítéletek és szokványos kliséktől mentes megértésére volt

hivatott, ez az átalakulás, az új és régi szereplők viselkedésének megértése, a konfliktusok feltérképezése és elemzése egy új Aranykor eljövételének ígérését hordozza. Ez még akkor is így van, ha a metrikák és indikátorok kalodájába zárt Szellem számára ez a kíváncsiság nem éppen kifizetődő. Meglehet, ez a szakma, ahogy én elképzeltem, ismertem és szerettem, éppen eltűnőben van – de ha kihálásra ítélt dinoszauruszként élünk is, a világ változásaira való nyitottság és kíváncsiság, hűség egy korábbi ethosz értékeihez ebben az elsöprő erejű változásokkal leírható világban még mindig a legautentikusabbnak gondolható Lét – igaz, jelenleg meglehetősen szűkös menüből választva.

*Budapest, 2019. július 19.*

*Dessewffy Tibor*

1.

# ELMÉLETEK

# A BIG DATA ÉS A TÁRSADALOMTUDOMÁNYOK VÉLETLEN TALÁLKOZÁSA A MŰTŐASZTALON

*Társszerző: Láng László*

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 dialektikáról, 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 viszonyoknak a második elemére és a társadalomtudományok számára jelentkező kihívásokra koncentrálnunk. De nem árt észben tartani, a társadalmi folyamatok és gyakorlatok fokozódó digitalizációja nyomán a megértés új útjai nem egyszerűen választható alternatívák, hanem egyre inkább a megismerés adekvát módozatai lesznek. Lehetséges a fiatalok kulturális fogyasztásairól vagy az idősök betegségekkel 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 exabájtnyi adat keletkezett, addig ma hét nap alatt jön létre ennyi információ. Ha a pontos számok kérdések is lehetnek, az exponenciális növekedés ténye vitathatatlan.

A másik meghatározó folyamat az adattermelés és az adatok digitalizációja. 2000-ben még a világ adatvagyonának 75%-át analóg formában, papíron, filmeken vagy kazettákon tároltuk. 2013-ban az információtömegnek a 98%-a digitális formában érhető el. *Tehát nem pusztán a mennyiségből, hanem az adatok digitális voltából fakadnak a megismerés új lehetőségei.* A digitális jel többet mutat meg ott is, ahol korábban nem értük utol a reprezentációt. Egy nyomtatott újság esetében korábban maximum az eladott példányszámot ismertük, a Big Data esetében viszont akár azt is tudhatjuk, hogy az olvasók mely cikkeket olvasták, mennyi időt töltöttek az egyes írásokkal, és ez milyen mintázatokat rajzol ki, például érdekl-e a külpolitikai cikkek olvasóit a gasztronómia. A digitális jel korlátlanul *időtálló* – évekkal 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 emberi tevékenység* nyomán jön létre, facebookos feltöltéseink és lájkolásaink mellett okostelefonunk különösebb 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.

## BIG DATA – SZOCIOLÓGIAELMÉLETI ELŐKÉPEK

Ez az újszerűség azonban nem szociológiai előképek nélkül való. E helyütt nem aspirálhatunk arra, hogy a társadalomtudományok történetét áttekintsük a Big Data szempontjából. Csupán a múlt század két végpontjáról mutatunk be egy-egy példát, amelyek 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 vita zajlott a szociológia természetéről Gabriel Tarde és Émile Durkheim között. A szociológia kanonizált története szerint ezt a vitát Durkheim, a tudomány alapító atyáinak 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 magyarázhatók (Durkheim 1895). Vagyis az egyéni, individuális és 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, az egyéni sajátosságok megértése és magyarázata a szociológia feladata.

„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 képekben terjeszthessük külföldön a napisajtóban” (Tarde 1962: 167–168, idézi: Latour 2010: 159).

Bruno Latour úgy véli, hogy ez a megközelítés nemcsak sokkal hasznosabb a társadalmi jelenségek megértésében, mint Durkheimé, de napjainkban Tarde – különösen a Big Data szempontjából – 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 „datascape” lehet, hogy egy évszázaddal később, igazolta Tarde meglátásait (...) Ma, hála a digitális médiumnak, tanúi vagyunk a nyomon követhetőség elve lenyűgöző kiterjesztésének. Ez ma nemcsak 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 leképeződései, telefonhívások és így tovább. Ami korábban csak a tudományos tevékenység számára volt lehetséges – hogy megmaradjon az aggregátumok káposztája, és az individuális kontribúciók kecskéje is jóllakjon –, 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 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 mérvadó: a tér és az idő kezelése. Azért is jogosult 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 időfel fogására 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* (valós időben) elérhető, másfelől az örökkévalóság időtlenségébe, ahol a digitális lábnyomok örökre megmaradhatnak.

E két alapidimenzió, a tér és idő új koncepciója alapján 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, ami a laikusok számára is elérhetővé tette az internetet. 1996-ban, amikor az internetet még primitív dizájn, szűkös tartalom és nehézkes hozzáférés jellemezte, egyáltalán nem volt magától értetődő világunknak az az értelmezése, amit Castells látott, és nem a science fiction, hanem a szociológia eszközeivel leírt. Azonban a Big Data forradalma után Castells tézisei, illetve az információs társadalomról adott jövendölései a szemünk előtt válnak valósággá.

## A BIG DATA MINT ÚJ ISMERETELMÉLETI PARADIGMA

A hálózati térben minden kapcsolat, cselekvés vagy érdeklődés nyomot hagy, és kutathatóvá válik. A megismerés és megértés új útjai nyílnak meg azáltal, hogy a világméretű hálózati közösség böngészésének, keresésének, navigációjának, a 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 és módszertani sztenderdjei kiüresednek, irrelevánsá válnak.

A jelen okfejtés szempontjából csak két dolgot kell kiemelnünk. Egyrészt, hogy a „szociológiai kánon” kontextustól függ. Az, hogy mit tekintünk elfogadott tudásnak, mindig is a módszertani, ismeretelméleti előfeltevésekhez kötött volt. Másodsor, ez a kánon soha nem volt homogén. Ebből fakadóan a szociológia jövőjére nézve nem az következik, hogy a Big Data új ismeretelméleti korszakában irrelevánsá 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, „újra felfedezik” 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 alakult ki, akik képesek az adatkezelés, az adatgyűjtés és -tisztítás, a tárolás és elemzés egész folyamatának lebonyolítására.

Meyer-Schönberger és Kenneth Cukier ezeket az új szakembereket „algoritmistáknak” hívja, utalva az alkímistákra. 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, mintázatokat, 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 korlátlanul válnak, a korrelációk és mintázatok



adattányászati 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, a Big Data által uralt valóságban az elméletre új szerep vár. Bár valóban vannak esetek, amikor az adattányászat meglepő korrelációkat mutat meg, ám a legtöbb problémát nem lehet megoldani kizárólag a korrelációk, mintázatok bemutatásával.<sup>1</sup> Az összefüggések értelmezése, azokból a tanulságok levonása igencsak földhözragadt maradna elmélet nélkül. A jelentés több mint algoritmus. Világunk megértéshez 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 komponense volt már a korai civilizációknak is (Rifkin 2010). A mezopotámiai és egyiptomi kultúrákban összegyűjtött és rögzített (csillagászati, meteorológiai, vízmennyiségi és mezőgazdasági) információk tették lehetővé, hogy az ókori birodalmak kielégítsék a termelési, a közellátásra vonatkozó, a kereskedelmi és a kormányzati adminisztratív igényeket. Ez az adatgyűjtés korántsem ártatlan vállalkozás. Mint 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*).

A hüvelyk, a rőf, az egy napi járőföld helyébe a metrikus rendszer lépett a maga standard és így könnyebben kezelhető, átváltható és illeszkedő adatrezimjével. 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,

<sup>1</sup> A Walmart, 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 rágcseknivaló 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 az eperlekváros sütit tartalmazó 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>).

Hunt 2008). A 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. Scott kiemeli, hogy az adatosítás a modern mezőgazdaságban, az erdőgazdálkodásban, a városi fejlődésben, a társadalmi mozgalmak és az ipari termelés esetében hogyan vezettek a sokoldalú, szerves, helyi és hétköznapi tudás visszaszorulásához, illetve 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).

## A 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 mintavétellel szemben próbáljuk feltérképezni. Az alábbiakban semmiképpen sem amellet kívánunk érvelni, hogy a Big Data-alapú megközelítések és kutatások módszertanilag problémamentesek lennének. Épp ellenkezőleg: mivel az adatok jelenleg többnyire rendszerezetlenek, „maszatosak”, hiányoznak a bejáratott 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 Cukier „A bigdata-forradalom”-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 újdonságot hoz a kutatásba. Véleményünk szerint *öt* további különbséget érdemes szem előtt tartani, amikor a mintavételen alapuló és

a Big Data-megközelítéseket összehasonlítjuk. Ezek az időbeliség dimenziója, a viselkedési információk szerepe, a strukturált és 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 mutatkozott az adatgyűjtés időpontjában megfigyelt és a néha évekkal később feldolgozott adatok közzétételekor uralkodó állapotok között (Csepeli és Prazsák 2004). A digitális technológiák alkalmazása révén az adatfelvétel sebessége is nagymértékben gyorsult.

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). Ugyanakkor ez a gyorsuló idő hatalmas módszertani és episztemológiai kihívást is jelent. A platformok folyamatosan változó adatelérési szabályai vagy az elérhető adatok körének rapid változása állandó reflexiót és innovációt igényel a kutatóktól.

2. *Vélemény vs. viselkedés.* 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 a félelem (pl. az 1980-as években a „hogyan vélekedik a Szovjetunióról?” kérdés esetében). De többről van szó, mint hogy a politikailag, gazdaságilag vagy szexuálisan érzékeny kérdések esetében a válaszok nem a valós vélekedéseket tükrözik. Schütz fenomenológiai tudásszociológiájának alaptétele, hogy a mindennapi tudás kiemelt jelentőségű, ez adja az hétköznapi életben magától értetődőnek tekintett „elsődleges valóságot”. 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: időmérleg-vizsgálatoktól az attitűdskálákkal végzett felmérésekig lehet találkozni ezzel a problémával. A kérdéses helyzetben készült felmérések hajlamosak olyan válaszokat produkálni, amelyek tükrözik a válaszadók vágyát, hogy megfeleljenek a vélt elvárásoknak. A survey-vizsgálatok eredménye ezért is ennyire ingatag, mert homályos és

gyakran nem létező nézetek tükröznek. Ezzel szemben a Big Data-alapú kutatás alapvetően a digitális lábnyomokat, azaz a valódi viselkedés nyomait használja. Ha például valakinek az internetes vásárlási preferenciái érdekelnek, akkor nem a „böngészés közben felkeresi-e az alábbi oldalakat?” kérdést kell megválaszolni, mint egy kérdőíves kutatásban, hanem a valódi viselkedés alapján az oldalak látogatottságát, az ott töltött időt, a továbbskattintások irányát és a kapcsolatokat is tisztán láthatjuk. Így a kutatók a hagyományos módszerekhez viszonyítva sokkal pontosabb képet szerezhetnek a valós preferenciákról, döntésekről és cselekvésről.

3. *A Big Data adatbázisok strukturált és nem strukturált adatokat tartalmazhatnak.* A „strukturált adat” olyan információ, amelynek világos és tisztán meghatározott a hosszúsága és formátuma. A strukturált adatok közé tartoznak a dátumok, szavak és számok csoportjai. A legtöbb szakértő egyetért abban, hogy ezek a strukturált adatok az összes adat mintegy 20%-át jelentik (Hurwitz et al. 2013). E strukturált adatok származhatnak szenzorokból, GPS-ekből, intelligens mérőórákból, valamint orvosi és más eszközökből. Az internethasználat során a webnaplók, a rögzített kattintások és a tranzakciók a strukturált adatok egyre növekvő és hatalmas forrását képezik. A strukturálatlan adatok a becslések szerint mintegy 80%-át teszik ki az összes adat mennyiségének. Ezeket az adatokat hagyományosan tartalomnak tekintjük: képek, videók, szövegek vagy webes tartalmak stb. Mivel a strukturálatlan adatok digitális formátumban érhetők el, számos új kifinomult tartalomelemzés, például szövegelemzés tárgyává tehetők (uo.). Ezen túl a strukturált és a strukturálatlan adat kombinálása gyökeresen új kutatási területeket nyit meg a társadalomtudósok számára. A két különböző típusú adat kombinálása esetén 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 adataikhoz ne lehessen hozzáférni – látszólag ijesztő távolságba kerül a társadalomkutatóktól.<sup>2</sup> Ám a helyzet ennél komplikáltabb, de

<sup>2</sup> 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, Crawford 2012). Egy fenyegető elemre azonban mégis fel kell hívnunk 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 is, felismerve az értéküket, egyre nehezebben teszik elérhetővé az adatokat. A folyamat végét még nem látjuk, és bár ezt itt nem taglaljuk, reális lehetősége van egy új adatbirtokló elit kialakulásának, amellyel szemben páriává válnak a hagyományos társadalomtudományok képviselői is.

kedvezőbb is. Ha valaki elsajátítja az adatgyűjtés és -feldolgozás módszereit – vagy együttműködést alakít ki erre nyitott algoritmistákkal –, hirtelen szinte belát-hatatlan lehetőségek nyílnak meg 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 és 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 és fogyasztói affinitás érdekel, a magyar moly.hu több mint százezer, valamint az amerikai goodsreads 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 nemcsak azért releváns, mert ily módon ki lehet küszöbölni a statisztikai hibákat, hanem azért is, mert a szegmentáció leképezheti a valóság sokszínűségét. A számosságbeli korlátok miatt a reprezentatív mintán 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 kereszt-tábláinak celláiba valószínűleg 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-Tarde-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 tudat”-ról, leegyszerűsítőek, 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 köztük levő kapcsolatok feltárását.<sup>3</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 és 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 szerencsések voltunk a *fundraising*ben, egy *survey*vel 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.

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

---

<sup>3</sup> A korábbi kommunikációs paradigmában gyakorlatilag semmit nem tudtunk arról, hogy üzeneteink mennyiben érték el a kívánt célcsoportokat. A Big Data-alapú megközelítés megkönnyíti az üzenetek nyomon követését. Ha szükséges, a hatékonyság növelése érdekében megváltoztathatjuk a terjedés útvonalát, sőt menet közben akár az üzenet tartalmát is személyre szabhatjuk.

kapcsolódó adatforgalmat. Félrevinné a jelen gondolatmenetet, ha itt részletesen értekeznenék az arab országokban a női szerephez kapcsolódó gyakorlatokról. Itt pusztán az előbbieken ismertetett módszertani sajátosságok szemléltetésére szorítkozunk.

1. *Időbeliség.* A Twitter-vihart a maga azonnaliságában 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 bárholnan hozzáférhetőek. 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 és ezek egymáshoz való viszonya. Ennek alapján definiálhatjuk a véleményvezéreket, a körjük szerveződő csoportokat, illetve ezen csoportok egymáshoz fűződő viszonyát és kapcsolataik rendszerét. Ezáltal láthatóvá válnak olyan nem triviális szerkezetek is, mint például, hogy az Occupy-mozgalom egyes alakjai vagy a kurd emberi jogi mozgalomhoz kapcsolódó csoportok milyen álláspontot képviseltek ebben a vitában (bővebben lásd Hannah 2012).

Az eredmények interpretációja során fokozott óvatossággal kell eljárunk. Ahogy arra a közösségi médiabeli forradalmak kapcsán sokan felhívták a figyelmet, az arab országokban jelenleg a lakosságnak csak egy elenyésző hányada használja a Twitter-t (Gladwell 2010, Morozov 2011, Fuchs 2012). Ezért pontosan kell látnunk az ilyen kutatások értelmezési tartományát, valamint 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.

## LÁTHATJUK-E A JÖVŐT?

A Big Data előbbi öt szociológiai jellemzője vezet az új paradigma egy hatodik, talán legmeghökkenőbb eredményéhez: a jóslatok és 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 nagy zacskós vattagolyókat. A Target ezért elkezdett babafelszerelések vásárlásához kedvezményt biztosító kuponokat küldözgetni vásárlójának, apja legnagyobb megdöbbenésére, aki az ő elenyésző emberi következtetési kapacitásaival még mit sem tudott a várható családi eseményről (Ellenberg 2014).

Nyilvánvaló, hogy az üzleti intelligencia új szintre ért el a Big Data-módszerek 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 mintázatok (több vattagolyó-vásárlás – terheség) kutassunk, és ez jelentősen megnöveli a predikció, az emberi viselkedés előrejelzésének esélyét is. A fizikus Barabási Albert László *Villanások* című könyvének alcímében egyenesen azt állítja: a jövő kiszámítható. Ez az állítás nem kis kihívást jelent mindazoknak, akik hisznek az ember döntési szabadságában és a 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 – a tudásszociológiai analógiával – egy gyenge és egy erős program is meghirdethető. A Big Data gyenge programjában a digi-



tális lábnyomokból nyilvánvalóan kirajzolódó főcsoportok: a kulturális fogyasztás, a közösségi média használata, az ifjúsági szubkultúrák vagy a társadalmi és politikai mozgalmak témáihoz kapcsolódó kutatásokban meglehetősen kézenfekvő a Big Data használata – már ha az adatok elérhetőek és rendelkezésre állnak. A Big Datának a mi elképzeléseinkhez közelebb álló, erős programja ennél többet mond, nagyobb magyarázó erővel 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 kapcsolati 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-gyakorlati paradigmaváltását vetíti előre. Ennek alapján nemcsak lehetséges a szociológia hagyományos nagy kérdéseinek, a társadalmi egyenlőtlenségek, a kirekesztés, a hatalomgyakorlás vagy a társadalmi tőke 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 – 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 nemzetirock-együttesek rajongói táborát. Itt nemcsak azt tudhatjuk meg, hogy ennek a közösségnek milyen a szerkezete, a kommunikációs hálózata, földrajzi megoszlása, időbeli aktivitása, vagy hogy milyen típusú információk milyen mintázatokban terjednek, hanem azt is, hogy az egyes aktivitásokhoz milyen affinitások társulnak (kik azok, akik a humorra reagálnak, kik azok, akik a brutalitásra, és milyen esetleges kapcsolat van e tematikák és a 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). Később mindez kiegészíthető további tartalmak, ízléspreferenciák és tevékenységek elemzésével (szeretik-e a fekete rapet vagy a

kosárlabdát), mikor, milyen gyakran és kivel kommunikálnak, és kinek a véleményét tekintik más közösségekben mérvadónak. Mindezt nem egyszeri adatfelvétellel, hanem időbeliségében folyamatosan mérhetjük.<sup>4</sup>

Az egyre inkább virtualizálódó társadalmi lét digitális lábnyomai elemzésének eszköze a Big Data-alapú kutatás. Nekünk, 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 tudásmonopóliumunkat az új adatbányászokkal szemben, akik nem a társadalomtudományi fogalomrendszerből indulnak ki, hanem hagyják, hogy az algoritmusok működése révén az adatokból bontakozzék ki a számukra megjelenő valóság. Azonban ezt a viszonyt nemcsak 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 eddig nem ismertünk; 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 tinédzserszexet övező bizonytalanságok Ariely által leírt jellegzetességeit, és a huszonegyedik század egyik legnagyobb intellektuális kihívásaként sok mindent megváltoztat majd – beleértve a társadalomtudományokat is.

*Eredeti megjelenés: Replika, 2015, 92–93, 157–170.*

---

<sup>4</sup> A jövő egyik legfontosabb kulcskérdése, hogy a szociológusoknak és társadalomkutatóknak lesz-e az ilyen adatokhoz hozzáférésük, 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 episztémológiai fordulatra nyílik mód.