

Nem vagyok túl nagy szám

(Bon Bon)

Vallomással és köszönettel kell kezdenünk. Vallomással, mert mint az olvasó is észrevehette, nem hagyományos műfajú cikket írtunk, hanem tudományos kiáltványt. Fő célunk az volt, hogy felhívjuk a figyelmet arra a kihívásra és lehetőségre, amit a Big Data jelensége tartogat a szociológia, illetve tágabban a társadalomtudományok számára. E zsáner szabályai szerint a közlő forradalom jelentőségét, felforgatóerejét szerettük volna kiemelni – így értelem-szerűen jó néhány dolgot nem vagy nem kellő mélységben tárgyaltunk. Magát a fogalmat is nagyon lazán kezeltük a fenti cél érdekében: Big Datának tekintettünk minden olyan digitális adatot, amely méreteiben túlmegy a hagyományos survey-alapú adatfelvétel és statisztikai feldolgozás keretein.

Köszönettel tartozunk a *Replika* szerkesztőinek, akik írásunk köré egész blokkot szerveztek. De még inkább és őszintén hálásak lehetünk a tanulmányok szerzőinek, akik különböző tudományos háttérrel, eltérő megközelítések alapján formailag is sokszínűen osztották meg nézeteiket az olvasókkal. Ez a sokszólamúság nem csak a téma komplexitását ragadja meg jobban, hanem azt is illusztrálja, hogy a szélsőségektől mentes, technorealizmus fősodrán belül maradó megközelítések mennyire eltérőek lehetnek. Gondolatébresztő, provokáló, egymással is felelő hozzászólások születtek, amelyek sok utat nyitnak meg a további töprengések előtt. Így nem csak a szerzők értékes hozzászólásait köszönjük, hanem azt is, hogy az írásaikból kibomló problémagazdagság és a lehetséges megközelítések sokasága gyakran pótolta, kiegészítette azon kérdések kifejtését, amelyeket nem vagy nem kellő alaposággal tárgyaltunk a kiáltványunkban.

Épp a rengeteg felvetett izgalmas kérdés miatt a részletes reagálás túlfeszítené egy ilyen összefoglaló kereteit. Nem vesszük végig tételesen az egyes tanulmányok állításait, csak néhány olyan fogalmi csomópontot jelölnénk meg, amelyeket az írások alapján különösen fontosnak találunk – ha nem kiáltványt írtunk volna, ezeket talán mi is jobban hangsúlyozzuk. Nyolc továbbgondolásra is alkalmas, főbb tematikus csomópontot látunk.

1. *Veszélyek és gyengeségek.* Székely Iván a SWOT-analízis alapfogalmaihoz nyúl vissza, amikor a Big Data „sötét”, vagy legalábbis aggályos oldaláról beszél – és igaza van. Tegyük hozzá rögtön: Z. Karvalics is indokoltan hangsúlyozza az új jelenségek – számunkra különösen fontos módon a technológiai újítások – nyomán kitörő morális pánikok gyakoriságát a társadalmakban. Ám bármennyire is helytállóak és szórakoztatóak Karvalics e morális pánikot meggyőző példákkal illusztráló érvei, az NSA körüli, illetve a különböző üzleti (Facebook, Uber, Spotify), illetve a hackerek okozta (Wikileaks, Ashley Madison) botrányok után aligha cáfolható, hogy mindinkább bővül azoknak a szereplőknek a köre, akik egyre növekvő mennyiségű s többnél több életösszefüggésben gyűjtenek és tárolnak adatokat rólunk. Ezen adatok felhasználásának szándéka és képessége persze változó lehet, de a személyes adatok védelme, a privacy és az emberi méltóság fogalmai lényegileg újragondolandó átalakuláson mennek keresztül napjainkban – a Big Data következtében is.

2. *Kutatás és személyes adatok.* Ám nem csak állami szervezetek, titkosszolgálatok és az üzleti hírszerzés gyűjtik digitális adatainkat, hanem a társadalomtudományok kutatói is belépnek a potenciális adatgyűjtők sorába. Gayer Zoltán helyesen mutat rá, hogy míg a hagyományos survey-technikákon alapuló elemzések lényegéből fakad az anonimitás, addig a Big Data konkrét személyek viselkedéselemeinek megismerését és elemzését is lehetővé teszi. Nem kell hosszasan fejtegetnünk, hogy a „nagyvárosi egyetemista fiatal lányokról” tett állításokhoz képest a privacy szempontjából mennyivel aggályosabbak azok az információk, amelyekre az „XY Facebook-oldaláról begyűjtött elérhetőségek, kedvelések és barátok” adatbázisokba rendezése révén tehetünk szert. Az anonimizálás lehetősége itt is adott, és talán nem a társadalomtudományi megismeréstől kell a legjobban félni a személyes adatok alanyát. Elvégre milyen kutatási kérdés az, ha ezek a személyes szintű adatok relevánsak? Nos, lehetnek ilyenek (pl. a hálózatok egyes elemeinek azonosítása), de azért azt is láthatjuk, hogy ezek nem minden kutatás számára lényegesek. Ezzel együtt a különböző nagy adatbázisok beszerzésére, tárolására és anonimizálására vonatkozó diskurzus különösen fontos lesz a társadalomtudományok számára is.

3. *Az episztemológiai problémák.* Németh Renáta okkal veti fel, hogy a Big Data által kínált tudás sem objektív és abszolút. Az összegyűjtött adat ugyanúgy esetleges, a mérőeszköz sajátosságainak kitett, mint a survey esetében. Sőt. Míg a survey-ek hagyományokkal rendelkeznek és leülepedett eljárásokkal s megoldásokkal dolgoznak, addig a Big Data-elemzéseknek még nincsenek ilyen bejáratott protokolljai. De nem csak a módszertan kiforratlansága az, ami bonyolítja a képet, hanem az adatok – Cukier és Mayer-Schönberger kifejezésével – „maszatossága” is. A strukturált és nem strukturált adatok harmonizálása és értelmezése, a valószínűség-számítással megtippelt adathiányos cellák kitöltése számtalan problémát vet fel. De ha ezeken a „maszatosság”-ból fakadó problémákon túllépünk, még mindig ott van a valóságkonstrukció redukcionizmusa. Ahogy mondani szokták, „kalapácsos ember mindenütt a szöveget keresi”. Vegyük például az adatvizualizáció általunk nem tárgyalt, de a Big Data szempontjából lényeges részterületét. Egy-egy szerkezet, hálózat bemutatása valóban heurisztikus lehet. De egyúttal azt a veszélyt is magában rejt, hogy – maradva a hálózati példánál – ott is véleményvezérek, centrális szereplőket, hálózati csúcsokat keresünk, ahol ennek nincs igazán jelentősége. A megmutatás gesztusa mindig magában hordja az eltakarást, mint ahogy az emlékezést és a felejtést is – és eszünk ágban sincs azt állítani, hogy a Big Data-elemzés mentes lenne ettől a problémától.

4. *Mit reprezentálnak a digitális világ adatai?* Többeknél előkerül az „n=all” képletének kritikája, abban a formában, hogy pl. a közösségi médiában mért vélemények és adatok mennyiben feleltethetők meg az állásfoglalás nézetek tagoltságának. Valóban, számos lehangoló példát hozhatunk arra, ahogy Facebook-adatok elemzésével jut valaki a „társadalomra” vonatkozó megállapításokra, vagy a Tumblr-kommentfolyamok elemzésével próbálnak érvényes állításokat tenni a „fiatalokról”. Márpedig a részből az egészre való ugrásokat legalábbis problematizálni kellene. Először is érvelhetünk amellett, hogy ezek az adatok az elsődleges adatgyűjtési csoportról szólnak (vagyis a facebookozók és a Tumblrt használók csoportjáról). Németh Renátának igaza van abban, hogy a véletlenszerű mintavételi eljárás statisztikai garanciáinak hiánya miatt az értelmezési tartomány kiterjesztésével csínján kell bánni. Annál is inkább, mert a digitális lábnyomokat számolatlanul maguk mögött hagyók sokasága nem csak „lefele” hiányos. Vagyis nem csak arról van szó, hogy vannak, és vélhetőleg mindig is lesznek olyan társadalmi csoportok, akiket, példánkban maradván, a Facebook és a Tumblr nem ér el. De szembe kell néznünk a jelenséggel is, amit Székely Iván jelez, hogy a digitális bennszülöttek egy részénél nagyon erős az adatvédelmi tudatosság – vagy nem használják, vagy lehetőség szerint korlátozzák az adataikra vadászó applikációkat. És így némileg módosítva visszajuthatunk a hagyományos adatfelvételi problémához: a társadalomban legfelül és legalul levőket nem igazán érjük el.

5. *A Big Data-metonímia.* A Big Data azonban nem csak az elérés/reprezentáció estében veti fel a „miről is beszélünk” kérdését, hanem az egységesnek tételezett jelölő is rendkívül sokszínű valóságokat takar. Nyilván egészen más következtetések levonására és értékítéletek hozzárendelésére ad alkalmat, ha egy nanomikroszkóp vagy a vezető nélküli autó Big Data-val dolgozó és azt újratermelő eszközeiről beszélünk. A hagyományos szociológiai szótárral egész más hatásmechanizmusokat és mintázatokat írhatunk le, mint ha az Ok Cupidhoz hasonló társkeresők működését vizsgáljuk, vagy mint amikor egy tornádó áldozatai megtalálásának és segítésének digitális filantrópiához sorolható Big Data-alkalmazásait vizsgáljuk (Rudder 2014; Meier 2015). Karvalics nagyon pontosan írja le az értelmezések eltérő szintjei közötti csúszkálás veszélyeit. A korábbi „szétziláló” (*disruptive*) innovációk esetében is ezzel a jelenséggel találkozunk (Christensen et al. 2008). Az internet terjedéstörténetét a mai napig erős hiedelmek kísérik. Ezek többnyire a fenti csúsztatási mechanizmuson, a részterületeken megállapított részvizsgálatok reflektálatlan kiterjesztésén alapulnak (pl. internetfüggő játékos, addikció s az elmagányosító internet mítosza). Ha e jelenséget tudásszociológiailag nézzük, a folyamat nem igazán meglepő. A radikálisan új jelenségekkel találkozáskor általában azok bizonyos részeire rezonálunk, azokat próbáljuk megérteni és azokból általánosítunk. Ebből akkor fakadnak problémák, ha az általánosításokból kizárjuk a valóság visszacsatolásának, kontrolljának lehetőségét. Ennyiben tehát nem csupán a Big Data problémájáról van szó – de ezen a területen, épp, mert átláthatatlan és sokszínű spektrumról van szó, e hiba elkerülésére fokozottan kell figyelni.

6. Osztjuk a Csepeli György esszéjében is megfogalmazott gondolatot, hogy a Big Data a szociológiai megismerés számára hatalmas lehetőséget tartogat. De azt is látnunk kell, hogy ezek a projektek ma még gyermekcipőben járnak. Ha biztosra vehetjük is, hogy az andersoni provokáció az „elmélet végéről” lekerül a napirendről, azt nem tudjuk, hogy mely területeken, mely elméleti irányzatokkal kombinálva lesz igazán erőteljes és revelatív ez a kísérlet. Gayer tanulmánya megerősít minket abban, hogy az izlés, és tágabban a kultúrszociológia,

sokat profitálhat egy ilyen alapállásból. Az mondjuk, hogy a szegénység újratermelődésének kutatása mennyiben hasznosíthatja a Big Data által kínált lehetőségeket, ma még nyitott kérdés. Ennek megválaszolása, a technológiai lehetőségek radikális bővülése mellett, a kutatók új generációjának kreativitásán és lelkesedésén, valamint a tágabb tudománypolitikai mező erőterének változásain múlik majd.

7. *Módszertani felvetések.* Itt nem csak a jelenleg még aluldefiniált, nem kanonizált Big Data-elemzési technikák megismeréséről van szó – vagy szerényebben, olyan nyelv kialakításáról, amely a társadalomtudósok számára is értelmezhetővé teszi ezeket az analitikai eszközöket. Mivel írásunkban a survey-módszerrel szembeállítva próbáltuk bemutatni a Big Data jelentőségét, itt hadd pontosítsunk. Nem a survey végét vizionáljuk. Épp ellenkezőleg, a survey, és ami azt illeti, a kvalitatív módszertan kreatív reneszánszát prognosztizáljuk. Habár az üzleti életben valóban gyakran elég lehet a szimpla korreláció, az okságról, és tágabban a narratív értelmezésről nem mondhat le a reflektált megértés és a társadalomtudomány. Ezért módszertanilag, a Big Data eredendő maszatossága miatt, a predikciós képesség megalapozásához, illetve az eredmények értelmezéséhez elengedhetetlen lesz az olyan új integrált módszertan kialakítása, amelyben az online és offline, kvalitatív és kvantitatív módszerek új, az adott helyzethez igazodó konglomerátuma alakul ki. A digitális antropológia már megfogalmazta az igényt, hogy a Big Data helyett inkább a Thick Data-ra, a Sűrű Adatra fókuszáljunk. Az elnevezés Clifford Geertz klasszikus, programadó tanulmányára utal, amelynek fő üzenete ebben a kontextusban is értelmezhető. Az adatok soha nem „beszélnek önmagukért”, hanem az érvényességi határaikat, értelmezési tartományaikat lehatároló és az összefüggések jelentéseit feltáró interpretációban nyerik el jelentésüket (Wang 2013).

8. *Társadalomtudomány és Big Data.* Végül bármilyen fontosak is az elméleti kiáltványok és vitacikkek, nem általuk fog eldőlni, hogy a Big Data valóban forradalmasítja-e a társadalomtudományokat vagy csak múlt divatról van szó. Az lesz a valódi kérdés, amelyen a Big Data alkalmazása megmérettetik, hogy alkalmasabb-e a valóság magyarázatára, többet tud-e hozzátenni a világ megértéséhez, mint a korábbi paradigmák. Mint korábban említettük, ez jelenleg még nyitott kérdés. Egyelőre nem látunk elsöprő teljesítményeket. Ám e tekintetben sem vagyunk pesszimisták. A digitális térbe is beköltöző létformák növekvő száma és aránya miatt is okkal lehet feltételezni, hogy a Big Data-alapú megközelítések relevanciája is növekedni fog, s ez elvezethet a digitalizálódó, hálózati és hibrid valóság mélyebb, alaposabb megértéséhez. Hogy a régi kínai átokként érvényes a mondás, vagy, épp ellenkezőleg, áldást jelent a szociológia számára, még nem tudhatjuk; de izgalmas időket élünk.

Hivatkozott irodalom

- Christensen, Clayton M., Michael B. Horn és Curtis W. Johnson (2008): *Disrupting Class. How Disruptive Innovation Will Change the Way the World Learns.* New York, NY: McGraw-Hill.
- Mayer-Schönberger, Viktor és Kenneth Cukier (2013): *Big Data. A Revolution That Will Transform How We Live, Work, and Think.* London: John Murray.
- Meier, Patrick (2015): *Digital Humanitarians. How Big Data is Changing the Face of Humanitarianism.* Boca Raton, FL: CRC Press.
- Rudder, Christian (2014): *Dataclysm. Who We Are (When We Think No One's Looking).* New York: Random House.
- Wang, Tricia (2013): Big Data Needs Thick Data. *Ethnography Matters* (2013. május 13.). Interneten: <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>.