

Székely Iván

Az adatmentes zónák szükségessége és esélye

Helytelen reflexió Dessewffy Tibor és Láng László írására

Azért helytelen ez a reflexió, mert nem az eredeti írás logikáját és felépítését követi, ugyanakkor mégis Dessewffy Tibor és Láng László (2015) vitaindító tanulmánya váltotta ki megírását. Szándéka szerint problémafelvető jellegű (s ebben követi az eredeti tanulmány műfaját, amennyiben a „probléma” kifejezést értékmentesen, nemcsak „gond”, hanem „lehetőség” értelemben is tekintjük), ráadásul a hozzászólás műfajának sokszínűségét kihasználva némiképp esszészzerű is.

A „Big Data” a 2010-es évek egyik divatszava vagy inkább divatkifejezése az üzleti elemzők és az információs szuperhatalmak szakértői körében, de népszerűsége a politikusok és a tudomány képviselői között is nő. E tekintetben szervesen illeszkedik azon korábbi metaforikus elnevezések sorába, amelyek új adatelemzési módszerek és eszközök létét (és nélkülözhetetlenségét) vannak hivatva tudatni a potenciális felhasználói körrel: ilyen elnevezés az adatbányászat vagy az adattárház is. Természetesen a mennyiségi változások minőségivé válhatnak, és ez történik – nemcsak a tanulmányban felvázolt alkalmazási területen – a Big Data esetében is. Ezért nem is vitatom az elnevezés hasznosságát sem az üzleti és politikai célú elemzések, sem az empirikus szociológiai vizsgálatok terén, de lehet, hogy néhány év múlva már mosolygunk a „Big Data” megnevezésén. Ha visszatekintünk az információkezelés demokratizálásában fontos szerepet játszó, ma már afféle kőkori eszköznek tűnő floppy-lemezek tárhelykapacitásának jelölésére, az 1980-as, 90-es években általánosan elterjedt 5¼-es formátumú lemezek eredetileg 170 kB kapacitását 360 kB-ra növelve már dupla sűrűségű (*double density*) adattárolásról, 720 kB-nál nagy sűrűségű (*high density*), 1,2 MB-nál már kiterjesztett sűrűségű (*extended density*) adattárolásról beszéltünk, az 1,76 MB-os kapacitás már szinte Nagy Adattömegnek tűnt, és hasonló megítélésbeli változást látunk a processzor-teljesítmények fejlődésében vagy a „nagyszámítógépek” és „szuperszámítógépek” számítási kapacitásának növekedésében, és mindenhol, ahol akár Moore törvényét, akár a hálózatalmélet törvényszerűségeit látjuk érvényesülni – amikor pedig az extra, ultra, szuper, hiper előtagok listája kimerül, új elnevezést kell találni és előről kezdődhet a fokozás. Nyilván a Big Data is nagyságrendekkel nagyobb lesz a mainál a mindenütt jelenlévő szenzorok elterjedésével, az intelligens tárgyak, az „internet of things” általánossá válásával, és az új minő-

séget meg kell majd különböztetni a maitól. Ezt nemcsak a fejlesztők üzleti érdekei diktálják, hanem az az elvárásunk is, hogy szinte minden napra jusson egy technológiai forradalom, áttörés, ami majd megváltoztatja a világot – ennek ébren tartásáról nemcsak a bulvársajtó, hanem a tudományos és szakmai fórumok szóhasználata is gondoskodik. Ezt táplálja az az – egyébként vitatható – elképzelés is, hogy annál jobb, ha valamiből minél több van, különösen, ha információról van szó. A mostani adattömeg és az adatelemzési kapacitás tehát a közeljövőben nem is lesz annyira „Nagy”, és gondolatban már most hátra kell lépniünk, hogy egyáltalán megbecsülhessük a mennyiségi változások minőségi következményeit.

Mindez persze nemigen befolyásolja a jelenkori kutató hozzáállását, aki számára a mai lehetőségek is forradalmiak, ezért nézzük először is a szociológiai kutatások gyakorlatias szemszögéből a mai Big Data perspektíváit.

A tanulmány összeveti az „adathiányos”, vagyis a mostanáig tartó korszak vizsgálati módszereit a nagy adattömegeken végzett vizsgálatokéival. Valóban, megváltozott a minta fogalma és szerepe az empirikus társadalomtudományi vizsgálatok perspektívájában, azonban ne felejtjük el, hogy a korábbi korszakok csak a mai nézőpontunkból tűnhetnek „tökéletlenek”, hiányosnak: minden korszak szerves egységet képezett a saját céljaival, lehetőségeivel, nézetrendszereivel – köztük társadalomfelfogásával –, problémáival és azok értelmezésével, megoldásaival. Ez az egység természetesen nem csupán látványos paradigmaváltásokkal szegmentált idilli, statikus korszakokból állt, hanem dinamikusan változó szakaszokból, amelyekben hol a problémák koncipiálása, hol a megoldásukra kitalált módszerek szaladtak előre és mutatták a másikat elégtelennek, különösen utólag nézve. Közlebebről, hol túl kevés adat állt rendelkezésre egy-egy ambiciózus empirikus vizsgálat céljára, mint például az időjárás soktényezős elemzésére és ebből következően előrejelzésére, hol pedig túl sok, mint például a népszámlálási adattömeg akár évtizedeket is igénylő feldolgozására. Ez azonban nem azt jelenti, hogy éppen most jött el a Kánaán, amikor a célok és az eszközök végre egyensúlyba kerültek – épp ellenkezőleg: az új eszközök természetesen új célokat, nézetrendszereket – köztük társadalomfelfogásokat –, problémákat és azok értelmezését, megoldásait fogják generálni. Ma a Big Data lehetőségei iránt vonzódó kutatóknak éppen ezt az új egységet kell megteremteniük, illetve tisztázniuk kell néhány alapvető, ebben a közegben nem triviális kérdést a maguk számára is: egyáltalán mit akarnak mérni, milyen következtetések akarnak levonni, voltaképpen mire jó és mi is az, amit csinálnak.

Tisztázandó kérdések

Természetesen vannak a fentieknél jóval gyakorlatiasabb kérdések is, amelyekkel az új lehetőség iránt érdeklődő társadalomkutatóknak (és egyben megbízóiknak és a kutatási eredmények alkalmazóinak is) foglalkozniuk kell – nevezhetjük ezeket virtuális SWOT-analízisünk Gyengeségeinek és Veszélyeinek. Tehetjük ezt azért is, mert az eredeti tanulmány bőséges listáját nyújtotta az Erősségeknek és Lehetőségeknek.

Datafication

Az egyik ilyen kérdés az adatközpontúság, angol szakzsargonnal a *datafication* vagy *dataism* növekvő szerepe a társadalom vizsgálatában, megértésében. A nagy adattömegek gyors és

hatékony elemzése minőségileg megváltoztathatja a kvantitatív vizsgálatok módszertanát és hatékonyságát, de vajon átveheti-e a kvalitatív vizsgálatok szerepét, megkérdőjelezheti-e a kvalitatív megközelítés létjogosultságát?

Szélsőségesen adatközpontú megközelítéssel akár azt is prognosztizálhatjuk, hogy a jövőben már nem lesz szükség személyes kontaktusra a kutató és a kutatás alanyai között, nem lesz szükség mélyinterjúk költséges és időigényes készítésére, hiszen a tetszőleges granularitású adatokból teljesen világossá válik a kutató számára, hogy az alanyok miről mit gondolnak és mit cselekszenek, sőt az is, hogy miről mit fognak gondolni és mit fognak cselekedni – egyszerűen olyan tulajdonságaikat is megismerhetjük az adatok elemzésével, amelyeknek az alanyok maguk sincsenek tudatában. Minek akkor megkérdézni őket? Sokkal tisztább, jobban strukturált adatokhoz juthatunk a távolból, az emberek viselkedésének megfigyelésével és elemzésével. Úgy gondolom, hogy ezzel a megközelítéssel a társadalomtudományok módszertani spektrumának, sőt önmeghatározásának fontos eleme, mondhatnám, ethosza veszne el. Ha pedig nem engedjük elveszni, akkor frissíteni kell a viszonyukat, nehogy véglegesen szétváljanak a „korszerű” adatelemzéssel és a „hagyományos” kvalitatív módszerekkel végzett vizsgálatok útjai, eredményeik értelmezésének keretei.

Természetesen a személyes kontaktussal folytatott empirikus vizsgálatok is elvben könnyen „datafikálhatók” lennének, hiszen a technológia rendelkezésre áll: az alanyok testén elhelyezett érzékelők jeleiből, agyhullámaik felfogásából, a stimulálásukra adott válaszreakciók digitális adataiból könnyen megítélhető lenne az alanyok gondolkodása, közléseik őszintesége és megbízhatósága. A kvalitatív módszerek azonban éppen az ilyen scénáriók elkerülését, a humán kontaktus, az empátia lényegi elemként való megtartását célozzák, s ezzel az adatközpontú megközelítésnek mintegy módszertani ellentétét, vagy inkább kiegészítőjét alkotják. Kissé olyan a viszonyuk, mint az agy-elve kutatói dichotómiájának képviselői: mindkét diszciplína vizsgálatainak ugyanaz a tárgya, és egyre több az egymásra rímelő kutatási eredmény, feltárt összefüggés a két terület között, de a kutatások sohasem érnek teljesen össze, mert alapvető megközelítésük eltérő.

Adatok és kutatók

A növekvő adatközpontúság a társadalomtudós, kiemelten az empirikus kutató hozzáállását is megváltoztathatja: nagyobb lehet a kísértése arra, hogy divatosan ugyan, de kritikátlanul használja a nagy adatbővséget anélkül, hogy megértené az elemzési módszerek lényegét és az adatok jelentését. Ez a kísértés ciklikusan vissza-visszatér a korszerű adatelemzési eszközök és módszerek fejlődésével, használatuk kényelmesebbé válásával és automatizálhatóságával. Amikor például megszűnt a survey típusú adatfelvételek állományainak kizárólagosan nagy-számítógépes környezetben végezhető feldolgozása – a mintegy élő tűzfalként a kutató és a gép között álló operátorokkal együtt –, a személyi számítógépeken végezhető analízis új szabadsága azokat is a kvantitatív elemzések műfajához csábította, akiket addig nemcsak az adatelemzés bürokratikus nehézségei, hanem a társadalom működésének megértéséről alkotott felfogása is visszatartott a matematikai statisztika alkalmazásától. Amikor pedig a sokváltozós statisztikai elemzések, például a klaszter- és faktorelemzések már a PC-ken is végezhetővé váltak, „a gép” mintegy gombnyomásra produkálta a publikálható grafikus eredményeket, de nem biztos, hogy az eredmények akkori interpretálását mai szemmel minden esetben megalapozottnak tekintenénk.

Most ismét fennáll a lehetősége annak, hogy a felhasználóbarát vizualizáló programok a kvalitatív szemléletű kutatókat is az adatok kiterjedt használatára csábítsák, anélkül, hogy azok jelentését magukévá tennék. A „világtörténelem egy percben” típusú – egyébként impresszív – vizuális megjelenítő alkalmazások hatalmas adattömeg feldolgozásával érdekes rétegeket adhatnak hozzá eddigi ismereteinkhez, de nem biztos, hogy közelebb hoznak a világhoz, az emberi viselkedés lényegi megértéséhez. Sőt az az illúzió is kialakulhat, hogy a nagy adattömegek társadalomtudományi felhasználásához nem is kell érteni magukhoz az adatokhoz. A kísértés többretegű: ahogy azt Mayer-Schönberger és Cukier (2013) is hangsúlyozzák a tanulmány szerzői által is hivatkozott könyvükben, a Big Data eleve nem a „miért” feltárására irányul. Más szóval, ebben a közegben az okozati összefüggések helyett a feltárt korrelációk foglalják el a vizsgálat eredményeinek, hozzáadott értékének szerepét. Ez, tágabb értelemben, egybevág a társadalom igazgatásának divatos felfogásával (ellenőrző társadalom, aktuárius társadalom) vagy a devianciakezelés korszerűnek tekintett módjával,¹ amelyben a deviáns cselekedetek *oka* már nem érdekli sem a kutatót, sem a rendvédelmi szervezetet, sem az igazságszolgáltatást, csupán azok bekövetkeztének valószínűsége, megjósolhatósága és korrelációja az egyének és csoportok nagy adattömeg által leírható jellemzőivel.

Adatok és alanyok

Big Data és surveillance – különösen a mindenütt jelenlévő (*ubiquitous*) és rendszerszerű megfigyelés (*surveillance assemblage*)² – egymással szoros kapcsolatban álló, sőt egymást kölcsönösen inspiráló fogalmak. David Lyon (2014) szerint a kapcsolat három fő tényezője az összekapcsolt megfigyelési célú adatállományok és elemző eszközök használatának intenzifikálódása, a megfigyelés egyre erősödő jövőorientáltsága, valamint a megfigyelés alanyainak kettős szerepe, amely egyszerre táplálja és háritja a megfigyelést.

A nagy adattömegek elemzésével végzett megfigyelési tevékenység célja elsősorban min-tázatok keresése, mint ahogy erre irányul az ilyen eszköztárral végzett társadalomkutatás is. A két tevékenység elkülönítéséhez többek között az szükséges, hogy a kutató megkülönböztesse az akár teljes mintán ($n=all$) alapuló, de a társadalom egyes szegmenseire vonatkozó vizsgálatokat azoktól az esetektől, amikor a teljes mintából egyénekre, családokra, azonosítható kis csoportokra vonatkozó következtetéseket von le és tesz felhasználhatóvá.

A személyes azonosíthatóság kulcsfogalom a humán vonatkozású adatelemzésekben: az volt az „adathiányos” kor mintavételen alapuló vizsgálataiban is, és még inkább az a Big Data közegében. A 20. század második felében alakultak ki azok az elvek és eljárások, amelyek már nemcsak az empirikus szociológiai vizsgálatok egyedi adatainak bizalmas kezelését, hanem az azonosíthatóság mielőbbi megszüntetését, vagyis az adat és az alany közötti kapcsolat helyreállíthatatlan törlését célozzák. Ezt nem csupán a kutatói etika vagy később a kodifikált jogi előírások, esetleg a visszaélések elkerülésének pragmatikus szempontjai követelték meg, hanem az alanyok bizalmának megteremtése és fenntartása is, ami az adatminőség fontos előfeltétele. Ugyanilyen jelentőségűek azok a szabályok, amelyek az alany azonosítására

¹ Lásd az új büntetéstani (*new penology*) elméletét, klasszikus forrása Feeley és Simon tanulmánya (1992).

² A *Replika* 89. számában a megfigyeléstudomány, a *surveillance studies* témaköréből készített tematikus összeállítást, amelyben többek között a mindenütt jelenlévő és rendszerszerű megfigyelésről közölt tanulmányokat.

alkalmas adatok kutatási és adminisztratív célú kezelésének szigorú elválasztását írják elő: az alanyok által például a vagyoni helyzetükre vonatkozóan nyújtott információk átadása az adóhatóságnak a kutatói etika súlyos megsértését jelentené.

A Big Data esetében az alanyok bizalmának megszerzése vagy megtartása már elveszítette funkcióját, s ezzel egy etikailag fontos fék esett ki a kutató szempontrendszeréből. Ha nem is tételezzük fel, hogy a nagy adattömegek társadalomtudományi elemzőinek nem lesznek kutatásetikai megfontolásai, a gyakorlatban egyre inkább csak a jogszabályi rendelkezések, pontosabban azok érvényesíthetősége és szankcionálhatósága jelentenek korlátot a kutatási célú elemzések eredményének személyazonosításra alkalmas formában (például üzleti vagy politikai marketing céljára) történő felhasználásában.

Ma a közismert információs szuperhatalmak (Google, Facebook és társaik) komoly erőfeszítéseket tesznek annak bizonyítására, hogy nem tudják, sőt nem is akarják azonosítani felhasználóikat, hogy így mentesüljenek az adatkezelés jogi felelőssége alól – holott nyilvánvaló, hogy üzleti modelljük a felhasználóik minél jobb, minél egyénibb profilírozásán alapul. A társadalomkutató sem mondhatja azt, hogy „nem tudjuk, ki vagy, mert nem rögzítjük a TAJ-számodat” (csupán a származásodat, a lakhelyedet, a kapcsolataidat, a fogyasztói előlekedet és a közlekedési kihágásaidat stb.). A másik típusú, kevésbé szem előtt lévő információs szuperhatalom, a titkosszolgálati szektor ugyancsak a személyazonosításra alkalmatlan mintázatok elemzését hangsúlyozza, holott a titkosszolgálatok működésének deklarált célja, hogy azonosítsák a nemzet-, köz-, humán- vagy közelebről meg nem határozott biztonság szempontjából fontos egyéneket. További elterjedt (és alapvetően hamis) érv, hogy a szolgálatok csak „ártalmatlan” kommunikációs metaadatokat rögzítenek, nem pedig magát a kommunikáció esetlegesen szenzitív tartalmát.³ Az adatelemzési technológia azonban lényegét tekintve azonos a két szektorban, és azonos a nagy adattömegeken végzett társadalomtudományi kutatások perspektivikus adatelemzési technológiájával is. A Big Data-adatelemzést használó társadalomkutatónak tehát alapvetően tisztáznia kell a viszonyát a kutatás alanyai-val és kutatási eredményeinek felhasználóival.

Adatok és felhasználásuk

Noha a Big Data üzleti és politikai célú felhasználásának hatékonysága és a fejlesztéséhez fűződő érdekek messze meghaladják a tudományos kutatás eszköztárát és perspektíváit, a két világ kapcsolata ugyancsak óvatosságra int. Egy üzleti vagy politikai adatelemzés a pszichológusok, szociológusok, szociálpszichológusok szakértelmével és tudásával kombinálva könnyen aprópénzre váltható, más szóval, e tudás birtokában még inkább megnő a befolyásoló képesség a társadalom szegmensei, csoportjai és egyénei felett, és ez a befolyás, értékrendtől és nézetrendszeredtől függően, „jó” és „rossz” egyaránt lehet. A Big Data közegében, ahol a kutató egyre kevésbé tudja megbecsülni a kutatási eredmények visszacsatolásából, üzleti és politikai célú felhasználásából eredő következményeket, még az etikus hozzáállású kutatók is nehezen tudják kizárni a „*dual use*”-t, vagyis azt, hogy a kutatási eredményeket a

3 A metaadat, Bruce Schneier szavaival, „maga a megfigyelés” (https://www.schneier.com/blog/archives/2013/09/metadata_equals.html), de a Snowden-botrány után elég arra gondolnunk, hogy csupán a napi mintegy ötmilliárd rögzített mobiltelefon-adat segítségével miként követhető az emberek tartózkodási helye, mozgása, kapcsolati hálójá, életmódja az Egyesült Államok titkosszolgálatai számára.

kutatás eredeti céljától és a kutató meggyőződésétől eltérő vagy egyenesen ellenkező célra használják fel.⁴

A nagy adattömegek inherens jellemzője a „rendetlenség”, vagyis a pontatlanság és hiányosság, azonban olyan nagyszámú az adat (és a következtetések is korrelációkon, nem pedig pontos megfeleltetésen alapulnak), hogy ez nem okoz gondot az elemzésben. Pontosabban, nem okoz gondot a kutatónak, de okozhat az alanyaknak, ha az elemzés eredménye közvetlenül csatolódik vissza a végrehajtó hatalom beavatkozó intézkedéseibe. A 95%-os korreláció nem vigasztalja a maradék 5%-ba tartozó egyént, akit a többség viselkedési mintázatai alapján valamilyen nemkívánatos viselkedésűnek ítélnék vagy csupán annak prognosztizálnak, és így akár egy el sem követett kihágás miatt „proaktívan” szankcionálnak, rosszabb esetben a kategorikus gyanú⁵ tárgyává tesznek. Természetesen a mintavételen alapuló vizsgálatok esetén is van hibahatár, de itt a visszacsatolás a kutatási eredmények és az adminisztratív intézkedések között korántsem ennyire szoros. A nem tudományos érdeklődésből végzett (szintén „n=all” típusú) népszámlálási adatfelvételeknél és vizsgálatoknál pedig az egyedi adatok nevesített személyes adatként való igazgatási felhasználása a statisztikai szakma jogi és etikai normáinak legsúlyosabb megsértését jelenti.

A fentiek fényében kérdéses, hogy a Big Data közegében növekszik-e vagy csökken a távolság a társadalomtudományi alapkutatások és az alkalmazott kutatások között, egyáltalán lehetséges-e alapkutatásokat végezni az eredmények közvetlen üzleti, politikai, bünyöldözési felhasználása nélkül, avagy a jövőben már csakis alkalmazott kutatásokról beszélhetünk?

A minta a teljes populáció?

Természetesen tisztáznunk kell azt is, hogy tulajdonképpen milyen társadalmat, milyen populációt mérünk a nagy adathalmazokkal. Az n=all perspektíváitól és a hálózati penetráció fejlődési ütemének grafikonjaitól elkápráztatva hajlamosak vagyunk azt hinni, hogy az emberiség kizárólag internethasználókból, és azon belül az internetet az üzleti és politikai szereplők – kiemelten az információs szuperhatalmak – szempontjai szerint használókból áll.

Ennek a problémakörnek a tárgyalása feltűnően alulreprezentált, nemcsak a Big Data-elemzések, hanem a ma elterjedt „mintavételes” online kutatások módszertanában is. Az utóbbi esetében az olcsóság, a gyorsaság, a könnyű elemezhetőség elfedi a kutatások megrendelői elől (de még a kutatók elől is) azt, hogy nemcsak elveszítik a társadalom egyes rétegeit a vizsgálat látóteréből, de azon belül is csak azok véleményét vizsgálják, akik hajlandók válaszolni az ilyen megkeresésekre, és hogy egy lelkes unokával rendelkező tanyasi nyugdíjas véleményének ezerszeres felszorozása nem pótolja a népesség ezen szegmensének reprezentatív vizsgálatát.

A nagy adattömeg alapú elemzések a válaszadási hajlandóság problémáját részben megoldják, mivel az adatok többnyire az alanyok tudta és beleegyezése nélkül kerülnek rögzítésre és elemzésre. Azonban ezekben az adatokban nincsenek benne, vagy torzítva vannak

⁴ Például – valós esetek nyomán – elképzelhetjük, hogy egy kutatás célja az, hogy megvizsgálja két, a politikai marketing által összekapcsolt érték összefüggését az emberek gondolkodásában, és ha a vizsgálat eredménye a függetlenségüket bizonyítja, a marketerek ennek birtokában még jobb hatásfokú „árukapcsolást” tudnak elérni a politikai kommunikációban.

⁵ Például a radikális muszlim fórumokat látogatókat eleve terroristagyanús elemeknek tekintik és repülési tiltólistára helyezik.

benne azoknak az embereknek az adatai, akik szándékosan igyekeznek elkerülni a profilírozást, már amennyire észszerű energiárfordítással ez egyáltalán lehetséges: hirdetésblokkoló vagy éppen ellenkezőleg, „mindenre rákattintó” alkalmazásokat⁶ telepítenek, anonim böngészőket,⁷ webpoloskairtókat,⁸ TOR-hálózatot,⁹ Blackphone-t¹⁰ és hasonló eszközöket¹¹ használnak. Ez a nem elhanyagolható szegmens ugyan nem befolyásolja a „hagyományos” Big Data statisztikai elemzések végzését, mint például a Google keresőkérdések megoszlásának vizsgálatát, de befolyásolhatja például a földrajzi megoszlással, így a járványok terjedésével való korrelációját.

Mit is mérünk?

Mit mér, illetve mit fog mérni a jövőben a Big Data alapú társadalomkutatás? Elsősorban az emberek életviszonyainak, kommunikációjának a hálózati világba települő mintázatait, a résztvevők digitális lábnyomait, a „virtualizálódó társadalmi lét” jellemzőit, és csak másodsorban az offline életviszonyok digitalizált, hálózaton elérhetővé tett adatait. Azonban az internetes kommunikáció jelenleg sem elhanyagolható, de a jövőben várhatóan tovább növekvő részét már ma is gépek, robotok generálják. A chatbotok¹² gondoskodnak arról, hogy csökkenő kommunikációs aktivitás esetén is megfelelő forgalmat generáljanak a szolgáltatók számára a közösségi hálózatokban, és nem kizárt, hogy a humán felhasználók nevében egymással is kommunikálnak; az anonimizáló kommunikációs protokollok gyakran álforgalmat generálnak a valós résztvevők között a forgalomanalízisen alapuló megfigyelés nehezítésére; a spambotok¹³ egyes becslések szerint az e-mail-forgalom kilencven százalékáért felelősek, és üzeneteiket gyakran a címzettek ismerőseinek nevében küldik; az intelligens ágensok a humán felhasználók nevében kutatnak fel információkat a hálózaton, adnak és vesznek értékpapírokat, licitálnak árveréseken; az avatárok pedig, vizuális megjelenésüktől függetlenül, afféle virtuális személyiségekként önálló cselekvésre, interakciókra képesek a virtuális térben. Ezeknek az entitásoknak az aktivitása ugyancsak részét képezi az elemzendő hálózati adattömegnek. Noha időnként éppen a gépi partner kéri tőlünk képfelismeré-

6 Ilyen az AdNauseam elnevezésű Firefox kiterjesztés, amely együttműködik a hirdetésblokkolókkal: automatikusan és véletlenszerűen minden blokkolt hirdetésre rákattint (ezzel még növeli is a kattintásalapú hirdetési bevételüket, de összezavarja a profilírozást végző hirdetőket), <http://dhowe.github.io/AdNauseam/>.

7 Az anonim böngészők elrejtik az illetéktelen harmadik felek elől (így a társadalomkutatók elől is) azt, hogy ki milyen weboldalakat látogat.

8 A webpoloskák a felhasználó képernyőjén megjelenő észrevehetetlen elemek, amelyek célja, hogy a felhasználó és a szolgáltató közötti kommunikációt egy harmadik fél megfigyelhesse. Ilyen technikával működnek egyébként a látogatottságmérő szolgáltatások is.

9 A TOR (The Onion Router) egy elosztott elven működő anonim kommunikációs hálózat, amely elrejtja a kommunikációban részt vevő felek kilétét, a kommunikáció tartalmát és az egyes kommunikációs aktusok összefüggéseit az illetéktelen harmadik felek elől (<https://torproject.org>).

10 A Blackphone egy megfigyelésgátló szolgáltatáscsomag részeként kifejlesztett, kereskedelmi forgalomban kapható „okostelefon”, <https://blackphone.ch/>.

11 Összefoglaló néven Privátszférát Erősítő Technológiák (Privacy Enhancing Technologies, PET); színvonalas magyar és angol nyelvű fóruma a Nemzetközi Pet Portal és Blog, <http://pet-portal.eu/>.

12 Intelligens beszélgetőágensok (*chat robots*), amelyeket már bárki alkalmazhat akár saját beszélgetőpartnernek, akár saját nevében kommunikáló virtuális személyiségként, akár csupán a kommunikációs szolgáltatás forgalmának növelése céljából.

13 Kéretlen üzeneteket (spameket) küldő programok: van e-mail-küldő, blogokon és közösségi hálózatokon posztoló, wikik és hasonló fórumok bejegyzéseit generáló válfajuk.

si képességet igénylő feladatok teljesítését („ha nem robot vagy, írd be ide a képen látható ellenőrző kódot”),¹⁴ humán aktorként egyre kevésbé tudjuk megkülönböztetni a valós és mesterséges kommunikációs partnert, és kérdéses, hogy a Big Data elemzői meg tudják-e őket különböztetni.

Ehhez járul a *filter bubble* jelenségének hatása is: egyéni szinten az, hogy korábbi hálózati tevékenységeink egyre inkább bezárnak egy testre szabott buborékba, amelyben a világnak csak egy szűrt nézetét látjuk, és ez behatárolja a világról alkotott képünket, cselekvési lehetőségeinket; csoportszinten pedig az, hogy hálózati tevékenységeink során eleve azt kapjuk, amit „mindenki” keres, és így egyre kevesebb az esélyünk arra, hogy új, alternatív gondolatokat, termékeket, szolgáltatásokat ismerjünk meg, vagy hogy a mi alternatív elképzeléseinkből valaha mainstream gondolatok válhassanak. Ez lehet, hogy jó hír az üzleti és politikai döntéshozóknak, de kevésbé jó hír a társadalomkutatóknak, hacsak nem éppen ennek a jelenségnek a hatásait akarják vizsgálni.

Következésképpen, lehetnek bármily prediktív erejűek is a Big Data-elemzések, nem a virtuális „valóságot” fogják tükrözni, hanem annak a virtuális lét jellegzetességei által módosított képét.

Kellenek-e adatmentes zónák?

Azt gondolom, hogy éppen az adatosítás, a datafikáció korlátlan terjeszkedésének reakciójaként egyes egyének, csoportok körében megnő az „adatmentesség” presztízse és igénye, és ennek a társadalomtudományok szempontjából sem elhanyagolható következményei lesznek.

Korai példák

Noha a társadalom adatosítását a modernitás jellemzőjének és egyben előfeltételének szokás tartani, az emberek megszámlalhatóságának és megszámozhatóságának¹⁵ igénye, csakúgy, mint az ezzel szembeni tiltakozás, voltaképpen a bibliai kortól kezdve nyomon követhető.¹⁶ A minőségi változást ebben az esetben is a mennyiségi, pontosabban kapacitásbeli változások indukálták: a második világháborút megelőző, de főleg a háborút követő időszakban a korai nagyszámítógépek segítségével a népesség nyilvántartása, igazgatása, ellenőrzése és kategorizálása minőségileg új lehetőségeket és távlatokat nyitott az államhatalom számára.

14 Ez a számítástechnikában elterjedt nevén a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) eljárás, amelynek lényege, hogy a számítógép automatikusan generál egy feladványt (többnyire egy grafikus eltorzított betű- és számkombinációt), amit csak egy emberi kommunikációs partner tud helyesen megválaszolni (átírni a billentyűzeten), de a válasz helyességét a gép is el tudja dönteni. (Voltaképpen egy fordított Turing-tesztről van szó: az eredeti tesztben egy embernek kell eldöntenie kommunikációs partneréről, hogy az gép-e.)

15 A kettő közötti különbség szakmai szempontból az, hogy a megszámlalhatóság csoport- vagy társadalmi szintű adatok, a megszámozhatóság egyéni szintű adatok gyűjtésére és feldolgozására irányul. Lásd még mai magyar kontextusban Zádori (2000).

16 Ha ez a kérdés nem is tartozik a korábban „történeti informatika”, majd „információelvű történelem”, ma leginkább „információtörténelem” néven ismert történettudományi irányzat (Z. Karvalics 2004) vizsgálati fókuszába, kutatói számos fontos adatot és összefüggést szolgáltathatnak erről a kérdésről.

„Az emberek megszámozása”, vagyis az adatosítás elleni ideológiák és mozgalmak kialakulása az 1960-as és 70-es évekre esett Nyugat-Európában és Észak-Amerikában. Ebben az időszakban a számítógépek, számítóközpontok még fizikailag koncentrált helyen működtek, vagyis láthatóak, vagy legalábbis elképzelhetőek voltak a laikusok számára; a Nagy Testvér is beazonosíthatóan, centralizáltan működött. Ekkor születtek a társadalom adatosításának és összehuzalozásának nagy ívű tervei, például a Nagy Hesseni Terv vagy az amerikai National Data Center terve, amelyek propagátorai deklarált célokkal és értéktartalommal töltötték meg vízióikat, és még nem volt divatos az az illúzió, amely szerint a technika voltaképpen semleges, és minden technikai újdonság jóra és rosszra egyformán felhasználható.

A 70-es években az európai lakosság egy részének emlékeiben még élt az emberek bőrébe tetovált azonosítószám képe, a tájékozottabbak pedig még emlékeztek a Hollerith-kártyák németországi felhasználására is.¹⁷ Magyarországon a hatvanas évekig tartott az eredetileg statisztikai célra felvett népszámlálási ívek adatai alapján kitelepített németek részbeni visszaköltözése a sváb falvakba. Nyugaton ebben az időszakban születtek azok az átfogó jelentések,¹⁸ amelyekben jogászok, társadalomtudósok és a számítástudomány képviselői próbáltak választ adni arra a kérdésre, hogy vajon ártalmas-e a számítógépesítés (voltaképpen az adatosítás) a nyugati típusú társadalomra, és ekkor kezdték publikálni tanulmányaikat a számítógépes jövő radikális optimistái és radikális pesszimistái.¹⁹

Harvey Matusow, a McCarthy-korszak kommunista aktivistájából lett antikommunista besúgója és hamis tanúja, később avantgárd művész, viharos életének ebben a szakaszában alakította meg az Adatfeldolgozó Gépek Megsemmisítéséért Küzdő Nemzetközi Társaságot és írta meg *A gerilla-hadviselés kézikönyve a számítógépek elleni viszontcsapásra* című művét (Matusow 1968). Szervezetének sajátos jellemzője volt, hogy kétezer fő körüli tagságának nem elhanyagolható részét a számítástechnikai szakma képviselőiből toborozta, ami némi-képp luddita színezetet adott mozgalmuknak.

Franciaországban a hetvenes évek közepén a sajtó által is tüzelt lakossági tiltakozás bontakozott ki a központosított állami nyilvántartásokkal szemben: a SAFARI (Système Automatisé pour les Fichiers Administratifs et le Répertoire des Individus) a felnőtt lakosság egységes azonosítón alapuló nyilvántartását és adatainak központosított tárolását célozta, a GAMIN (Gestion Automatisée de la Médecine Infantile) pedig a társadalmilag és egészségileg hátrányos helyzetűnek ítélt gyermekek nyilvántartását. Ez utóbbi tervről a szülők határozott fellépése miatt le is mondtak a szervezők, 1978-ra pedig megszületett „az informatikáról és a szabadságjogokról” szóló törvény és a jogok érvényesülését ellenőrző független testület, a CNIL.

Ezekben a mozgalmakban és akciókban több nézőpont keveredett: szerepet játszott a központi államhatalom ellenőrző képességének növekedésével szembeni gyanakvás, a magánélet és az önrendelkezés féltése, az emberek megszámozása, lélektelen adatokká változtatása

17 Lásd erről Edwin Black (2002) könyvét vagy röviden Andrassy (2014) összeállítását.

18 Nagy-Britanniában a Younger-jelentés (HMSO 1972), Franciaországban a Tricot-jelentés (Tricot 1975), de számos nyugati országban alakult tudós bizottság vagy kutatóhely a kérdéskör különböző aspektusainak vizsgálatára, például a Royal Commission on Publicity and Secrecy Svédországban (1969), vagy a Norwegian Research Center for Computers and Law (1971).

19 Elemző összefoglalását lásd Donk és Tops (1995).

elleni tiltakozás, a félelem attól, hogy a gépek veszik át az uralmat az emberek fölött – közös azonban bennük az adatokkal, az adatosítással szembeni ellenérzés.²⁰

Mai példák

A társadalom adatosításának mai szintje alig összemérhető a korábbi évtizedekével, mint ahogy a jövőbeli szintje is új minőségeket fog jelenteni. Annak ellenére, hogy az adatosítás alanyai, vagyis a társadalom széles rétegei számára a róluk szóló adatok gyűjtése a korábbi időszakhoz képest sokkal nehezebben észlelhető, feldolgozásuk módja és következményei sokkal nehezebben felfoghatók; hogy az elviccelt Nagy Testvér helyett sok szorgalmas Kis Testvér (és persze néhány óriásira nőtt Testvér) figyeli a viselkedésünket, kommunikációnkat, találja ki gondolatainkat; hogy a fegyelmező információs hatalom helyett legtöbbször az ajándékokat kínáló, tömeges testre szabást ígérő információs hatalommal találkozunk; és annak ellenére, hogy a legtöbb, a felszínen semlegesnek vagy egyenesen vonzóknak tűnő, de a társadalom alapvető értékeit erodáló változást afféle megfővő békaként²¹ éljük át, a fent említett nézőpontok, mozgalmak, akciók utódai mégis markánsan jelen vannak a ma információs társadalmában.

Az adattakarékosság és adatelkerülés elve (*data minimization principle*) ma már nemcsak tudományos diskurzusokban, hanem kötelező erejű jogi dokumentumokban is szerepel. Ez az elv az azonosítható személyekkel kapcsolatba hozható adatok kezelésére vonatkozik, és azt mondja ki, hogy csakis az adatkezelés céljához szükséges és elégséges mennyiségű adat kezelhető és csakis a cél által indokolt ideig őrizhető meg, ahol pedig a cél személyes adatok kezelése nélkül is megvalósítható, ott személyes adatok egyáltalán nem kezelhetők. Az adattakarékosság és az adatelkerülés látszólag ellentétes az üzleti és igazgatási célú adatkezelési gyakorlattal, amely szerint a cél minél több személyes adat rögzítése, elemzése és minél hosszabb ideig tartó megőrzése, és az eredeti céltől függetlenül törekedni kell az új összefüggések feltárására és hasznosítására. (A megoldást az adatok és az alanyok közötti kapcsolat megszüntetése, vagyis az adatok anonimizálása jelentheti.)

Az információs jogok koncepciójáról és a digitális identitásról folytatott szakmai és laikus diskurzusokban ismételtel felbukkan „a jog ahhoz, hogy ne kategorizáljanak” (*the right not to be categorized*) igénye. Ez az igény annak felismeréséből született, hogy a nagy adattömegek elemzéséből levont következtetések társadalmi osztályozáshoz (*social sorting*) vezetnek, és ez nemcsak csoportszinten, hanem személyes szinten is korlátozza az egyén mozgásterét, szabadságát.²² A „kategorizálatlan” állapot, mint az információs önrendelkezés egyik aspektusa felvetődött a privacy fogalma korszerű tipológiájának kialakításában is, de végül nem került be a jelenleg általánosan elfogadott tipológiába (Finn, Wright és Friedewald 2013).

20 Jól szemléltetik ezt a hozzáállást azok a korabeli karikatúrák, amelyek az emberek számítógépek általi „bedarálását”, viselkedésük gépies egységesítését, személyiségük számokká alakítását ábrázolják, vagy az ember-gép kapcsolat visszajára fordulását: amikor az operátor bemutatkozik a számítógépnek, egy megjegyezhetetlenül hosszú számsorral azonosítja magát, a gép viszont egyszerűen csak ennyit mond: „Alfie vagyok”.

21 A megfővő béka metaforája szerint, ahogyan a forró vízbe dobott béka ösztönösen kiugrik a vízből, de a lassan felmelegített vízben végül megfő, ugyanúgy az ember is hajlamos megszokni és elfogadni a számára előnytelen változásokat, ha azok lassan és fokozatosan következnek be.

22 Ennek a korlátozottságnak sajátos vetületét tükrözi a kulturális és etnikai diverzitás és a többgyökerű identitás témájával foglalkozó online fórum, a Mixed Roots Stories cikke (Salole 2014).

A civil szektor reakcióinak paradox példája a 2015 elején formálódó *We Are Not Numbers* (Nem Vagyunk Számok) projekt és közösség, amely ugyan voltaképpen kevesli az általuk képviselt gázai áldozatokról rögzített és közölt adatokat, és technikailag még több adat publikálását akarja elérni, hogy ezzel statisztikai adatok helyett emberi sorsok bemutatásával hozzá közelebb a történeteket honlapjuk látogatóihoz – a kezdeményezésben azonban nehéz nem észrevenni az adatok és az emberi lények, a statisztika és az egyéni sorsok szembeállítását, az adatosítással szembeni attitűdöt.

Ez az attitűd a világ szerencsésebb, jómódú szegmenseiben is megnyilvánul, ahogy azt a jóga, a környezettudatosság, a spirituális lét kultúráját hirdető népszerű amerikai életmód-magazinban megjelent szabadvers érzékletesen illusztrálja (Salole 2014). Néhány sorát nem művészi értéke, hanem a benne kifejeződő életérzés miatt idézem:²³

I am not a number.
I am not the number of friends I have (or haven't).
I am not the number of likes on my cheeky status update or just-posted profile picture.
I am not the number of text messages that make my phone spring to life and jingle.
[...]
I am a beautiful mess.
I can't be summed up.
None of us can.
We are not numbers.²⁴

Számos befolyásos civil szervezet tevékenységének középpontjában az információs jogok állnak, egyesek a minél több adat elérhetőségéért, nyilvánosságáért küzdenek, mások viszont az emberek információs magánéletének védelméért, adatosításuk és elemzésük korlátok közé szorításáért. Az utóbbi csoportból kiemelkedik az Electronic Privacy Information Center (EPIC), amely a testszkennerek használatának korlátozásától a keresőgépek adatvédelmi implikációig, az információs jogalkotás fejlesztésétől a Big Data-elemzések kockázatainak vizsgálatáig számos területen aktív, és amelynek elnök-vezérigazgatója, Marc Rotenberg az internet legbefolyásosabb ötven szereplője közé tartozik a *Newsweek* rangsora szerint.

Ugyancsak említést kíván a Privacy International,²⁵ amely radikális jogvédő és tanácsadói tevékenységét a világ számos országában folytatja, többek között a titkosszolgálati adatgyűjtések témakörében.

23 Nyersfordításban:

Nem vagyok szám.
Nem vagyok a barátaim száma (vagy annak hiánya).
Nem vagyok a pimaszkodásomat vagy éppen posztolt profilképemet lájkolók száma.
Nem vagyok a mobilomat felélesztő csilingelő SMS-ek száma.
[...]
Szépséges zűrzavar vagyok.
Összegezni nem lehet.
Egyikünket sem lehet.
Nem vagyunk számok.

24 Elgondolkodtató, hogy az adatosítással, a megszámozással, az elgépiesítéssel szembeni ellenérzések miért a nyugati kultúrkör sajátosságai – a keleti társadalmakban talán azért nincsenek hagyományai, mert ott az élő és élettelen, a lelkes és lelketlen között nincs olyan dichotóm kettősség, mint nyugaton.

25 <https://privacyinternational.org/>.

A legfontosabb indikátor azonban a Big Data típusú társadalomtudományi vizsgálatok alanyainak, az egyéneknek, családoknak és kis közösségeknek a véleménye, viselkedése. Ez eddig leginkább csak mintavételes empirikus vizsgálatokkal volt elemezhető; a szakmai és módszertani szempontból színvonalas Global Privacy of Data-vizsgálata, amelyet az Ipsos Reid közvélemény-kutató intézet a Queen's Egyetem megfigyeléstudományi központjának koncepciója alapján készített 2006-ban kilenc országban,²⁶ kimutatta, hogy az internethasználók között nem elhanyagolható arányban tapasztalható az a viselkedés, hogy nem vesznek igénybe egyes online szolgáltatásokat azok túlzó adatigénye miatt, vagy hogy hamis adatokat adnak meg a szolgáltató megtévesztése céljából (Zureik et al. 2010).

Nem elhanyagolható tehát azoknak a száma, akik valamilyen módon korlátozni akarják az adatosítás mértékét és annak saját személyükre, környezetükre gyakorolt következményeit. Ennek egyik módja a már említett PET-technológiák valamelyikének használata,²⁷ vásárlásaik, pénzügyi tranzakcióik például Bitcoinban bonyolítása, a kommunikációs eszközök és alkalmazások „adathalász” gyári beállításainak kedvezőbbre állítása, másik módja az a próbálkozás, hogy hamis adatokkal tévesszék meg az adatelemző, profilépítő automatákat.²⁸

Az a digitális bennszülött, aki ad a magánéletére, a mindennapi gyakorlatban a laptopján vagy táblagépén letakarja a beépített kamerát, kikapcsolja a mikrofont, amikor nincs rá szüksége; időnként „butatelefont” használ, és csak olyan mobil készüléket vesz magának, amelynek kivehető az akkumulátora, különben kötelezően be van kapcsolva a hálózathoz. E réteg tagjainál egyfajta ambivalens viszonyulás figyelhető meg az adatosítással és a hálózatisággal szemben: egyfelől elvárják, hogy éjjel-nappal a hálózat részei legyenek, másfelől jogot formálnak arra, hogy korlátozzák ennek mértékét és hatásait. Ez a hozzáállás csak részben irányul a személyes megfigyeltség elkerülésére, vagyis arra, hogy az alany csak a saját magánéletét védje, de másokéhoz hozzáférjen, gyakran a csoportos és egyéni adatmentesség (reálisabban: a korlátozottabb adatosítás) elérése a célja: ha már eléggé kiélvezte vagy elszenvedte az adatosítás igazgatási és szolgáltatási következményeit, kiiratkozik az életét átalakító közösségi hálózatokból, alternatív valós és virtuális közösségeket keres, és igyekszik fenntartani a lehetőséget, hogy információs autonómiáját, ha korlátozott mértékben is, de érvényesíthesse.²⁹

26 Brazília, Egyesült Államok, Franciaország, Japán, Kanada, Kína, Magyarország, Mexikó, Spanyolország.

27 Vitatható módszertannal ugyan, de az Eurobarometer FL 225 számú vizsgálata (European Commission 2008) az EU átlagában a tizenöt éves és annál idősebb válaszadók 56 százalékánál, a listavezető Hollandia esetében 79 százalékánál, Magyarországon 36 százalékánál mutatta ki az ilyen technológiák használatát.

28 Valerie Steeves kanadai diákok között, elsősorban a nemi szerepek közösségi hálózatokban mutatott sztereotípiái szempontjából végzett vizsgálatait, némiképp ellentétben az előzetes elképzelésekkel, azt mutatták, hogy a nők inkább használják adataik védelmére a technológiai megoldásokat, míg a férfiak inkább próbálkoznak hamis adatok megadásával, vö. Steeves előadását a CPDP 2015 konferencián (<https://www.youtube.com/watch?v=wEVO3Lb9i7g>). Megjegyzendő, hogy a hamis adatok megadása többnyire nem téveszti meg a profilépítőket és a Big Data-elemzőket, legfeljebb azzal egészítik ki az alany profilját, hogy hamis adatokat ad meg magáról.

29 A csoportos adatmentesség intellektuális luxusát kicsiben magam is átéltem, noha időben és térben korlátozott módon, a St. Gallen-i Egyetem és a Harvard Egyetem Berkman Központjának egyik közös workshopján (amelynek témája éppen a megfigyelés fogalmának dekonstrukciója és rekonstrukciója volt): a játékszabályok szerint három napon át mindenkinek ki kellett kapcsolnia a telefonját és a számítógépét, így a beszélgetések alatt nem lehetett üveget intézni, leveleket írni, híreket olvasni és hírt adni, blogolni, tuitelni, lájkolni, profilt építeni, ehelyett egymás személyére, mondanójára kellett koncentrálnunk és reagálnunk. Lehet, hogy így adatokat és információkat veszítettünk egy tágabb kontextusban, de tartós élményt és koncentrált szakmai tartalmat nyertünk a szűkebb közösségben.

Könnyű lenne ebben a prognózisban az adatmentes zónák igényét pusztán a hatás-ellenhatás társadalmi ciklikusságára fogni: ahogy a fast food kitermelte a slow food, sőt a slow mozgalmakat, úgy a „datafikáció” is kitermeli majd az adatfóbiát, az emberi lét lényegi elemeinek folyamatos digitális rögzítésével, elemzésével és befolyásolásával szembeni életérzést, ideológiát, rezisztenciát és rezilienciát – persze csak mint időben és térben körülhatárolt virtuális enklávákat a megkerülhetetlen mainstream hatással szemben. Ezzel szemben az adatmentesség vagy inkább megfigyelésmentesség igénye tartós jelenség a modern társadalmakban, olyannyira, hogy küszöbön áll az ezt kodifikáló jog, közismert provokatív nevén „the right to be forgotten” (RTBF – a jog ahhoz, hogy elfelejtsenek) hatályának nemzetközi szintre emelése.³⁰ Az RTBF nem a datafikáció univerzalitásának korlátozását célozza, hanem az így keletkezett adatoknak az alanyuk általi törölthetőségét, s ezzel információs önrendelkezésük érvényesíthetőségét.

Az RTBF nemcsak egy absztrakt jog, hanem a digitális kor egyik alapkérdését feszegető koncepció. Viktor Mayer-Schönberger gondolatébresztő, sőt provokatív könyve, a *Delete*, amely a vitaindító tanulmány szerzői által idézett *Big Data* című könyvének megjelenése előtt vált ismertté, sorra veszi a digitális örök emlékezet ígéretének okait, esélyeit, várható következményeit és a lehetséges válaszlépéseket (Mayer-Schönberger 2009). A szerző saját javaslata az, hogy minden személyes adatnak legyen általános vagy egyedileg beállítható lejáratási ideje, amelynek elteltével, a fizikai világ elporladásához hasonlóan, az adat semmisüljön meg, amennyiben az adatot valaki – értéke, érdekessége vagy személyes kötődése miatt – nem archiválja. Ezzel voltaképpen az archívumok több ezer éves történetének egyik alaptévékenységét, az értékelést és szelektálást hozza vissza egy új kontextusban; egy olyan tevékenységet, amellyel küzdenek az új archívumi paradigma intézményei, és amelyet digitális polgárként igyekszünk figyelmen kívül hagyni (Székely 2007).

A töröltetéshez való jog, érthető módon, sérti az információs szuperhatalmak és a Big Data-elemzők üzleti érdekeit: főállású brüsszeli ellenlobbistáik sikerrel keltenek indokolatlan félelmeket „a történelem vége” vagy a szólásszabadság korlátozása mint potenciális következmények iránt, s ehhez megnyerték például a kultúra- és szakmaféltő francia levéltárosokat, akik – az RTBF valódi tartalmának megismerése helyett – petíciót írtak alá a tervezett jog törvénybe iktatása ellen. Érdekes módon a készülő EU-rendelet másik adatkorlátozó előírása, a „data protection by design and by default” (magyarul kb. beépített és beállított adatvédelem) nem találkozott ilyen ellenkezéssel, legalábbis nyílt formában. A data protection by design azt követeli meg, hogy a személyes adatokat kezelő eszközöket és rendszereket eleve úgy tervezzék meg, hogy használatuk teljes életciklusán át felelősen és nyomon követhetően érvényesítsék az adatvédelmi szabályokat, a data protection by default pedig azt, hogy az ilyen eszközök és rendszerek gyári beállítása mindig az alany védelmére nézve előnyösebb állapotú legyen, vagyis az alany maga dönthessen tevőlegesen a kapcsolódó szolgáltatások igénybevételéről, és ne csak kikapcsolni tudja ezeket a beállításokat vagy leiratkozni tudjon

30 Az átfogó európai adatvédelmi reform legfontosabb elemeként készül, a tagállamokban közvetlenül hatályosuló adatvédelmi rendelettervezet (European Commission 2012) tartalmazza az RTBF jogát, amely lényegét tekintve egyébként nem új jog. A tervezet jelenlegi verziójában az RTBF a törölthetőséghez való jogként (*the right to erasure*) szerepel.

az automatikusan létrejövő regisztrációiról (arról nem is szólva, hogy bizonyos esetekben a leiratkozás sem lehetséges).

Hogyan kapcsolódik az egyéni adattöröltethetőség joga és igénye, vagy a kommunikációs eszközök és szolgáltatások adatelkerülő tervezése a nagy adattömegeken végzett társadalomtudományi vizsgálatokhoz? Az elemzések nyersanyaga értelemszerűen a minél teljesebb körű adatállomány, és ha az egyes alanyokra vonatkozó adatok hiányoznak a mintából, avagy az alanyokra vonatkozó adatok hiányosak a többi alany adataihoz képest, akkor ez torzítja a mintát, hasonlóképpen a szándékosan hamis adatokat szolgáltató (ilyen informatikai alkalmazásokat használó) alanyok adataihoz. Addig, amíg az információs önrendelkezésükkel ily módon élni kívánók száma alacsony, a torzító hatás is alacsony, illetve bőven „belefér” a Big Data típusú adatok inherens pontatlanságába. Ha viszont a vizsgált területen az ilyen alanyok és adatok száma magas, akkor a korrelációk számításánál, illetve az eredmények adminisztratív vagy marketing célú felhasználásánál ezt a tényezőt is figyelembe kell venni.

A felsorolt példák arra utalnak, hogy a modern tömegtársadalmakban az adatosítás léte és elfogadása tekinthető normál életformának, az adatmentes törekvések pedig alternatív nézeteket fejeznek ki. Az alternatív nézetek – csakúgy, mint az alternatív életviszonyok, szubkultúrák és ezek érvényesülési lehetőségei – azonban igen fontosak a nyugati demokratikus társadalmakban, részben e társadalmak önmeghatározásából adódóan, részben pedig azért, mert maguk is segítik a mainstream életviszonyok de facto és de jure határainak kialakítását, formálását.

Az adatmentes zónák egy adatosított tömegtársadalomban természetesen nem lehetnek teljesek és homogének: csak időben, fizikai és virtuális térben elhatárolt módon érvényesülhetnek a mai kor sokrétű és párhuzamos valóságaiban. Egyéni szintű érvényesíthetőségük is jelentősen megnehezedett az elmúlt évtizedek során a technológiai fejlődés és a ráépült alkalmazások globális elterjedése következtében, bár a másik oldalról az érvényesíthetőségüket támogatják egyes új technológiai fejlesztések, civil mozgalmak, sőt a jogalkotás azon elemei is, amelyek az információs túlhatalommal bíró szereplők hatalmát vannak hivatva korlátozni.

Mit kezdhet a társadalomkutató az ilyen hipotetikus adatmentes zónákkal? Természetesen beléphet és vizsgálhatja is, de nyilvánvalóan nem Big Data típusú adatelemzéses módszerekkel (ha csak indirekt módon nem a kommunikációs „lyukakból” következtet az ilyen zónák létére). Tartok tőle azonban, hogy ezek a zónák elsősorban azok kiváltságai lesznek, akik más fizikai és információs javak birtoklásában is magasabb státuszúak a többségnél: a jogaikat és technikai lehetőségeiket egyaránt jól ismerő, erős jogérvényesítő képességgel rendelkező jómódú rétegek tagjaié, kiegészülve az alternatív életformák radikális híveivel és gyakorlóival.

Helyretevési kísérlet

Ha az olvasó mindezek után azt gondolja, hogy ez az írás egy klasszikus technoszociális disztópia aktuális tükröződése, téved. Igaz, hiányzik belőle az információs társadalom beláthatatlannak tűnő távlatainak fenntartás nélküli csodálata, és általánosságban hiányzik belőle a kritikátlan technooptimizmus. Természetesen nem állítom, hogy a szerzők ilyen szemléletűek lennének, de az írásuk interpretálása nyitva áll minden olvasó előtt, köztük azok

előtt is, akik korunk technoszociális hibrid valóságában a technológiát mint a társadalmi problémák felszabadító erejű megoldóját látják, vagy magukévá teszik azt az axiómát, amely szerint az információs társadalom eleve jobb a korábbi korok társadalmainál. Aki nem osztja a lelkesedést, könnyen az ellendrukkerek, a technopesszimisták, vagy az értetlen konzervatívok táborába toloncolva érezheti magát. Jómagam ezzel szemben technorealistán tartom magam, és ez a reflexió is ebből a hozzáállásból íródott.

Néhány olyan szempontra igyekeztem felhívni a figyelmet, amelyek a Big Data lehetőségeitől, távlataitól megérintett társadalomkutatók számára sem hagyhatók figyelmen kívül. Természetesen a nagy adattömegek társadalomtudományi felhasználásának lehetőségei és azok implikációi jóval sokszínűbbek annál, mint amilyenek akár a vitaindító tanulmány, akár ez a reflexió mutatja, ezért az itt említettekén kívül még számos, legalább ilyen fontos szempont felvethető: ezek kialakulófélben lévő leltárának jó gyűjtőhelye a nemrég létrejött *Big Data and Society* folyóirat, de releváns diskurzusok jelentek meg azon diszciplínák fórumain is,³¹ ahol a nagy adattömeg elemzése érdemi változást jelenthet a kutatásban, illetve magán a vizsgált területen.

Ezek a változások nem csupán a társadalomtudományok eszköztárát tágítják, hanem magukra a diszciplínákra is kihatnak, amelyek tartalma, kiterjedése, eszközei és voltaképpen létoka is képlékennyé válik általuk. Nem tudom, hogy a Big Data empirikus szociológiai felhasználása hányadik hullámnak tekinthető az adattömeg elemzésének titkosszolgálati, katonai, ipari, természettudományi és üzleti kiaknázási hullámai között, de nyilvánvalóan az egyéni kutatói ambíciókon, az új területek felfedezésének intellektuális izgalmain túlmenően a szociológia és rokon diszciplínái testületileg sem engedhetik meg maguknak, hogy ne terjesszék ki vizsgálati módszereiket a Big Data területére.

Azonban – amint láttuk – a nagy adattömegek elemzése és felhasználása nemcsak a kutató munkájára vagy magára a diszciplínára, hanem a kutatások alanyaira is kihat, és e hatások megítélése korántsem mentes az éles vitáktól, elég csak a Facebook 2014-ben nyilvánosságra került kísérletére³² gondolnunk, amelyben felhasználóinak érzelmi állapotát manipulálta, vagy a prediktív adatok visszacsatolására és alkalmazására a társadalom egységeinek és egyéneinek „szabályozására”. Kosinski, Stillwell és Graepel (2013) 58 000 önkéntesen végzett vizsgálata bizonyította, hogy a „lájkolás” adatai önmagukban is lehetővé teszik az alanyok személyes és szenzitív jellemzőinek, köztük szexuális orientációjuk, nemzetiségük, vallásos meggyőződésük, politikai nézeteik, személyiségjegyeik, intelligenciájuk, boldogságuk, függőségeik, szüleikkel való kapcsolatuk, koruk és nemük meglepően pontos és automatikus meghatározását. Messzebből nézve azt látjuk, hogy az „ $n=all$ ” minta tagjai számára egyre kevésbé átlátható adataik sorsa, elemzésük következményei, egyszerűen szólva: az, hogy mi irányítja az életüket az információs társadalomban, nemcsak ok-okozati, hanem korrelációs értelemben is. Ezzel párhuzamosan pedig a nagy adattömegek és az elemzésükhöz szük-

31 Például a megfigyeléstudomány, a kommunikáció, a kriminológia vagy a hálózatelmélet tudományos fórumain, de a nagy kiadók az elmúlt néhány évben önálló periodikákat is indítottak a Big Data témakörében.

32 Az „érzelmi ragályok” vizsgálata (*emotional contagion study*) az alanyok tudta és beleegyezése nélkül történt, hasonlóképpen az 1950-es évek mozijaiban alkalmazott szubliminális üzenetekhez, csak itt a potenciális célközönség nem egy amerikai mozi közönsége, hanem a világ lakosságának számottevő része. A vizsgálatban egyetemeken is részt vettek az USA hadügyi köreinek – később cáfolt – pénzügyi támogatásával. Az eredeti (de már helyreigazításokat tartalmazó) tudományos közleményt lásd az Amerikai Tudományos Akadémia folyóiratában (Kramer et al. 2014), a nagyközönségnek szóló írások közül lásd például az ITcafé cikkét (Dajkó 2014), az etikai vonatkozásokról a *Guardian* írását (Chambers 2014).

séges eszközök és módszerek birtokában lévők (és nemcsak a titkosszolgálatok, hanem az alkalmazott társadalomkutatások végzői is) olyan információs előnybe kerülhetnek, amely tovább polarizálja az információs hatalmi viszonyokat, ez pedig ellentétes a hálózati világ látszólagos és valós demokratizálódási folyamataival. Mindez paradox módon azt eredményezi, hogy egyes „nyugati” típusú demokratikus értékeink és jogaink – köztük információs jogaink – a gyakorlatban egyre kevésbé érvényesíthetőek (a kulturális Keleten pedig ezek az értékek és jogok nem is „úgy” fontosak), vagyis a kutatások *közége* is megváltozik, bennük a kutató szerepével.

Ilyen távlatból nézve vajon mi lehet az, ami ebben a változó közegben valamilyen fogódzót ad a jelenre és a belátható jövőre nézve? Valamiféle biztosítékot a kutatások önazonos, jogszerű és etikus végzésére, a kutatási eredmények transzparens felhasználására? Nyilvánvalóan olyan alapvető elvek és értékek lennének alkalmasak erre, amelyek a gyors technológiai és társadalmi átalakulások közegében nem, vagy csak nagyon lassan változnak. Azonban a Big Data, a globalizáció, az aktuárius társadalom világában még az olyan alapvető elvek és értékek, mint a személyes szabadság, önrendelkezés, szabad akarat, emberi méltóság és egyebek is láthatóan erodálódnak, vagy legalábbis átalakulnak.

Egy informális beszélgetés során Antoinette Rouvroy-val, az információs társadalom és az algoritmikus kormányozhatóság kérdéseit vizsgáló filozófussal ezekről a társadalmi, jogi és filozófiai sarokpontokról és jövőjükről vitatkoztam, és azt állítottam, hogy legalább afféle „munkahipotézisként” ragaszkodnunk kell az ilyen alapvető elvekhez, értékekhez, gondolati tájékozódási pontokhoz. Rouvroy viszont azon az állásponton volt, hogy végső esetben egyetlen alapvető elvet tud csak változatlanul tekinteni egy elfogadható vagy éppen kívánatos jövőbeli társadalom szempontjából: a vitathatóságot.

Lehet, hogy igaza volt. Én mindenesetre remélem, hogy a Big Data-boom közegében és a nagy adattömegek jövőbeli humán vonatkozású alkalmazásánál legalább ez az egy elv tartósan érvényesíthető marad, és fennmaradásához ez a reflexió is hozzájárul.

Hivatkozott irodalom

- Andrássy Gábor (2004): *IBM és a Harmadik Birodalom*. Interneten: http://ritkanlathato.tortenelem.blog.hu/2014/12/03/ibm_es_a_harmadik_birodalom.
- Berta Sándor (2013): Dollármilliókat áldoz a magánélete védelmére Mark Zuckerberg. *Sg.hu* (2013. október 15.). Interneten: <http://sg.hu/cikkek/100535/dollarmilliokat-aldoz-a-maganelete-vedelmere-mark-zuckerberg>.
- Black, Edwin (2002): *Az IBM és a holokauszt*. Budapest: Athenaeum.
- Bulmer, Martin (szerk.) (1979): *Censuses, Surveys and Privacy*. London és Basingstoke: Macmillan Press.
- Chambers, Chris (2014): Facebook Fiasco. Was Cornell's Study of 'Emotional contagion' an Ethics Breach? Interneten: <http://www.theguardian.com/science/head-quarters/2014/jul/01/facebook-cornell-study-emotional-contagion-ethics-breach>.
- Dajkó Pál (2014): *A Nagy Facebook-manipuláció*. Interneten: http://itcafe.hu/hir/facebook_kiserlet.html.
- Deleuze, Gilles (1992 [1990]): Postscript on the Societies of Control. *October* 59 (Winter 1992): 3–7. Interneten: https://files.nyu.edu/dnm232/public/deleuze_postscript.pdf.
- Dessewffy Tibor és Láng László (2015): Big Data és a társadalomtudományok véletlen találkozása a mütőasztalon. *Replika* (92): 157–170.
- van de Donk, Wim B. H. J. és Pieter W. Tops (1995 [1992]): Informatizálás és demokrácia: Orwell vagy Athén? *Eszmélet* (27): 90–123. Interneten: http://eszmélet.hu/w_b_h_j_van_de_donk-informatizalas-es-demokracia-orwel/.
- European Commission (2008): Data Protection in the European Union. Citizens' perceptions. Analytical Report. *Flash Eurobarometer* 225. Interneten: http://ec.europa.eu/public_opinion/flash/fl_226_en.pdf.

- European Commission (2012): Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). COM(2012) 11 final, 2012/0011 (COD), Brussels, 25.1.2012. Interneten: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.
- Feeley, Malcolm M. és Jonathan Simon (1992): The New Penology. Notes on the Emerging Strategy of Corrections and its Implications. *Criminology* 30(4): 449–474. Interneten: <http://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=1717&context=facpubs>.
- Finn, Rachel L., David Wright és Michael Friedewald (2013): Seven Types of Privacy. In *European Data Protection. Coming of Age*. Serge Gutwirth, Ronald Leenes, Paul de Hert és Yves Pouillet (szerk.) Dordrecht: Springer Science + Business Media, 3–32.
- Harvey, Sarah (2015): I am Not a Number. *Elephant Journal* (2015. 03. 02.). Interneten: <http://www.elephantjournal.com/2015/03/i-am-not-a-number/>.
- HMSO (1972): *Report of the Committee on Privacy. Cmnd. 5012*. London: HMSO.
- Kosinski, Michal, David Stillwell és Thore Graepel (2013): Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110(15): 5802–5805. Interneten: <http://www.pnas.org/content/110/15/5802.full.pdf>.
- Kramer, Adam D. I., Jamie E. Guillory és Jeffrey T. Hancock (2014): Experimental Evidence of Massive-scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences of the United States of America* 111(24): 8788–8790. Interneten: www.pnas.org/cgi/doi/10.1073/pnas.1320040111 <http://www.pnas.org/content/111/24/8788.full.pdf>.
- Lyon, David (2014): Surveillance, Snowden, and Big Data. Capacities, Consequences, Critique. *Big Data and Society* 1(2): 1–13. Interneten: <http://bds.sagepub.com/content/spbds/1/2/2053951714541861.full.pdf>.
- Martin, James és Adrian Norman (1973): *The Computerised Society*. Bungay, Suffolk: The Chause Press Ltd.
- Mayer-Schönberger, Viktor (2009): *Delete. The Virtue of Forgetting in the Digital Age*. Princeton: Princeton University Press.
- Mayer-Schönberger, Viktor és Kenneth Cukier (2013): *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray Publishers.
- Matusow, Harvey (1968): *The Beast of Business. A Record of Computer Atrocities*. London: Wolfe Publishing.
- Pariser, Eli (2011): *The Filter Bubble. What the Internet is Hiding from You*. New York: Penguin Press.
- Salole, Lill (2014): What's Your Mix? *Mixed Roots Stories* (2014. 09. 02.). Interneten: <http://mixedrootstories.com/whats-your-mix/>.
- Székely Iván (2007): A négy archívumi világgép. *Információs Társadalom* 7(3): 15–46.
- Székely Iván (2013): Jog ahhoz, hogy elfelejtsenek és töröljenek. *Információs Társadalom* 13(3–4): 7–27.
- Tricot, Bernard (1975): *Rapport de la Commission Informatique et Libertés*. Paris: La Documentation Française.
- Z. Karvalics László (2004): *Bevezetés az információtörténelembe*. Budapest: Gondolat – Infonia.
- Zádori Zsolt (2000): Megszámolhatóság kontra megszámozhatóság. Kemény Istvánnal, Kertesi Gáborral, Majtényi Lászlóval, Székely Ivánnal és Zsigó Jenővel Zádori Zsolt beszélget. *Beszélő* 5(3): 8–18. Interneten: <http://beszelo.c3.hu/cikkek/megszamolhatosag-kontra-megszamozhatosag>.
- Zureik, Elia, L. Lynda Harling Stalker, Emily Smith, David Lyon és Yolande E. Chan (szerk.) (2010): *Surveillance, Privacy, and the Globalization of Personal Information*. Montreal–Kingston–London–Ithaca: McGill-Queen's University Press.