

Könyvszemle

SIPOS JÚLIA GONDOZÁSÁBAN

ADATBŐSÉG – A TÁRSADALOMKUTATÓK SZEMÉVEL

Az *Educatio* folyóirat – deklarált missziója szerint – interdiszciplináris szemle azok számára, akik az oktatás társadalmi összefüggéseit keresik. *Adatbőség* címmel megjelent tematikus számában a kifejezetten oktatással foglalkozó témák a megszokotthoz képest kisebb súllyal szerepelnek, a kötet jórészt a társadalomkutatás szélesebb spektrumát érintő kérdéseket, módszertani és metodológiai problémákat jár körül.

A kötet szerkesztője, Veroszta Zsuzsanna indító tanulmányában (*Adminisztratív adatok társadalomkutatási kezelése*) azt a témát veszi górcső alá, amely az adatbőség, mondhatni adatrobbanás napjainkban tapasztalható jelenségének talán legnagyobb szeletét képviseli. A nagy tömegű és szédületes gyorsasággal növekvő adminisztratív, nem kutatási célból begyűjtött adatok felhasználása a társadalomkutatásban igen vonzó lehetőség (már csak azért is, mert a kutató számára lényegében „ingyen” állnak rendelkezésre). Ugyanakkor nem szabad figyelmen kívül hagyni, hogy az adminisztratív adatokat, adatbázisokat számos speciális validálási eljárásnak kell alávetni ahhoz, hogy kutatásra alkalmas adathalmaz álljon elő. Ezt kiegészítik azok a teendők, amelyek az adatkapcsolásoknál, főleg az adminisztratív és survey adatbázisok közti összekapcsolásoknál jelentkeznek. Mindehhez nem áll még rendelkezésre olyan mennyiségű tapasztalat, olyan mélységű elméleti és módszertani tudás, mint a kutatási célú adatgyűjtések területén. Napjainkban zajlik ez a tanulási folyamat, amelyhez jelentős mértékben járul hozzá az írás. Elsősorban azzal, hogy tudatosítja a kérdés jelentőségét, és áttekinti a konkrét problématerületeket, teendőket, amelyeket figyelembe kell venni az adatkezelési eljárás során.

Az adatok társadalomtudományi felhasználásáról a történelmi idő dimenziójában is érdemes gondolkodni, amelynek során további kérdések merülnek fel. Ezek megfogalmazására vállalkozik Nagy Péter Tibor tanulmánya (*A régi és új történelmi adatok társadalmi természetét*). Abból indul ki, hogy sor kerülhet például régi típusú adatok új szempontokból, új módszerekkel történő feldolgozására, elemzésére, régi adatokból újjáalkotott metaadatok konstruálására, régi és újabb típusú adatok összevetésére, más, hasonló témájú kutatások adataival való összevetésre. Előfordulhat, hogy régebbi korokról eddig még fel nem tárt dokumentumok ke-

rülnek elő, a régi dolgokról új adatok keletkezhetnek, tehát meghosszabbodhat az az időtáv, amelyre adataink vonatkoznak, a jelen világ történéseiről viszont egyre szélesedő tematikában születnek strukturálatlan adattömegek. Mindez bonyolult helyzetet teremt az adatkezelésben. A szerző ezt úgy fogalmazza meg, hogy az adatfeltárás, adatrendezés és adatértelmezés jelenségvilága egy komplex tudomány-szociológiai összefüggésrendszer része. Az idő dimenziójában pedig azt a folyamatot jeleníti meg, amelynek elemei az új dolgokról szóló, de még strukturálatlan adatok keletkezése, a strukturált, illetve tudományos adatok születése, majd visszatöltődése az adatok körforgásába.

A további tanulmányok az adatgyűjtés, adatfeldolgozás és -felhasználás gyakorlatának néhány területét járják körül. Gárdos Éva munkája (*Adatok és kezelésük a hivatalos statisztikában*) áttekinti a hivatalos statisztikai szolgálat által használt adatforrástípusokat, felhasználási területeiket és módjukat. Eligazítja az olvasót a statisztikai adatgyűjtések rendszerében, amit a gyakorló kutató és más felhasználó általában a maga számára átláthatatlan dzsungelként él meg. Megérthetjük az elsődleges és másodlagos adatforrások, a Big Data szakszerű fogalmát, a vállalkozások és a kutatóhelyek adatai kezelésének sajátosságait. Kiemelten foglalkozik a tanulmány az adminisztratív adatok kérdésével, ezen adatok statisztikai felhasználhatóságával, a vonatkozó nemzetközi ajánlásokkal és a hazai gyakorlattal.

Meglehetősen kényes és mind a kutatók, szakemberek, mind a hétköznapiak szintjén sokat emlegetett probléma tekintetében ad igen hasznos segítséget a tájékozódáshoz Székely Iván tanulmánya (*Közadatok és nyilvános adatbázisok: a hozzáférés kérdései*). Az információs szabadságról van szó, ami a szerző szerint „röviden azt jelenti, hogy mindenki szabadon hozzáférhet a közszféra információihoz, és ezt a jogát csak szűk körben, törvény korlátozhatja”. Nem lehet megspórolni az alapvető fogalmak definíciójának, értelmezésének, gyakorlati használatának rögzítését, bemutatását (közinformáció, közadat, nyilvánosság, a személyes adatoknál az önrendelkezés). Kiemelt helyet kap az ingyenesség kérdése, ami értelemszerűen a felhasználókat leginkább érdeklő téma. Az információs piac érdekei és a más típusú szempontok érvényesítésének küzdelmében figyelemre méltó európai példa a nyílt hozzáférésű adatok EU-portálja, amely úgy ad ingyenes hozzáférést az adatokhoz, hogy nem vizsgálja felhasználásuk célját. A felhasználás célja pedig igazán sarkalatos kérdés (tekintettel az egyre gyakoribb bűncselekményekre, az adatokkal való visszaélésre). A szerző szerint azonban miközben ma már meglehetősen egyszerű a közadatokhoz való hozzáférés, az igazi privilégium az adatelemzési lehetőségek birtoklása, ami nemesak tudást, hanem megfelelő számítógépes kapacitást is jelent.

A közadatok felhasználásának egyik jelentős területéről ad képet Halász Gábor tanulmánya (*Adatokkal történő kormányzás: dilemmák és perspektívák*). A szerző nemzetközi kontextusban vizsgálja a kérdést, elsősorban az OECD és az EU

gyakorlatának (kiemelten oktatási programjainak – INES, PISA EU oktatási indikátorfejlesztés) bemutatásával teszi jól érthetővé, hogy mennyiben jelent újat az ún. adatokkal történő kormányzás, amely szorosan összefügg a tényeken alapuló döntéshozatallal. Ezek a módszerek annak a fordulatnak a megjelenési formái, eszközei, amelynek során a beavatkozó állam (kormányzat) értékelő, reflexív, kommunikatív állammá (kormányzattá) válik, ahol a szakpolitika a tényekkel és adatokkal való szembesítéssel „noszogatja” a szereplőket a kitűzött célok elérése érdekében. A szerző szerint a legnagyobb dilemma az, vajon a szakpolitika, a döntéshozatal „hol vonja meg azon adatok körét, amelyek begyűjtése lehetséges a rendelkezésre álló erőforrások keretei között és az adatszolgáltatók, és közvetlen begyűjtők túlzott terhelése nélkül”.

Ugyancsak nemzetközi megközelítést képvisel Széll Krisztián tanulmánya, amely már kifejezetten az oktatással kapcsolatos adatok kérdésére irányul (*Az adathozzáférés és felhasználás nemzetközi gyakorlatai*). Két olyan ország – Anglia és Svédország – példáját mutatja be, amely egységes adatgyűjtési és felhasználási alapelvekre épülő, több évtizedre visszatekintő adatkultúrával rendelkezik. Bár oktatási rendszerük struktúrája – képzésük eltérő történeti, társadalmi jellemzői következtében – különböző, a szóban forgó téma tekintetében több hasonlóságot mutatnak. Először is szinte minden információ egyéni szinten áll rendelkezésre, ami lehetőséget ad a többcélú felhasználásra a tanügyigazgatástól a költségvetés tervezéséig, a szakpolitikai elemzéstől az oktatáskutatásig. Igen fejlett a tanulók, a tanárok és az iskolák teljesítményének értékelése. Kiemelt szempont a transzparencia és elszámoltathatóság, és sikerült kellő egyensúlyt kialakítaniuk az adatvédelem és az adatfelhasználás között.

A hazai pedagógiai célú adatgyűjtés és kezelés helyzetét a Köznevelési Információs Rendszer gyakorlata alapján mutatja be, elemzi Salomvári György tanulmánya (*Köznevelés az adatok bővületében – az adattermelés forradalma*). Azt a folyamatot írja le, amelynek során az adattömeg látványos növekedése mellett az aggregáltsági szint csökkentése válik egyre fontosabbá. Az egyéni szintű elemi adatok kezelése azonban komoly szabályok követését feltételezi a személyi azonosíthatóság kizárására. A szerző bemutatja, hogy milyen új lehetőségek rejlenek az új típusú adatvagyon felhasználására a pedagógiában, de felhívja a figyelmet arra a veszélyre, amelyet a komplex szemlélet elvesztése jelenthet az alacsony aggregáltsági szintű adatokból való kiindulás során. Adatkapcsolati háló létrehozásában gondolkodik, amely lehetővé teszi a különböző célú adatvagyon-felhasználással kapcsolatos igények kielégítését.

A kötet záró tanulmánya az előzőekhez képest gyökeresen más problémával foglalkozik, miközben az apropó ugyanaz. A nagy tömegben „készülő” adatok felhasználásának lehetősége nemcsak a kormányzatok, a kutatók, esetleg intézmények érdeklődését kelti fel, hanem az üzleti világot is. Bogáromi Eszter és Máth András munkája (*Sokforrású adatbázis-építés – buktatók, nehézségek, megoldási*

kísérletek) esettanulmány, amely leírja azt a munkafolyamatot, amelyet egy sajátos szemléletű településtipológia készítése során jártak be. A megrendelő magyar nagyvállalat termékportfóliójának átrendezésekor realizálta, hogy szüksége lenne a településszerkezet mélyebb megismerésére, hogy marketing erőforrása-it olyan területekre tudja koncentrálni, ahol azok a legnagyobb valószínűséggel hoznak növekvő piaci lehetőségeket. A szerzők a feladatot különböző adatbázisok összekapcsolásával oldották meg. Az olvasót megismertetik az egyes lépéseknél felmerülő problémákkal, a konkrét döntéseknél alkalmazott megfontolásokkal és a munka eredményével, a tipológiával.

A tanulmányok sorát a folyóirat *Valóság* rovatában két interjú szövegének közreadása követi. Kádárné Fülöp Judit az adminisztratív oktatási adatok felé irányuló növekvő figyelem háttéréről és következményeiről beszél. Arról, hogy még nincsenek kiaknázva az abban rejlő lehetőségek, nincs egységes stratégia az oktatási adatvagyon felhasználására, menedzselésére, fejlesztésére, adat-összekapcsolások végzésére. Rudas Tamás a TÁRKI-ban végzett adminisztratív alapú kutatásokról, a TÁRKI adatbank szerepéről, annak módosulásáról számol be. A bizalmatlanságot alapvető akadállynak tartja az adminisztratív adatok szélesebb körű felhasználásában, ami a lakossági óvatosság néha meglepő hiányával együtt van jelen. Felhívja a figyelmet arra az ellentmondásra, hogy az új eljárások kidolgozását általában nem társadalomkutatók, hanem mérnökök és fizikusok végzik, akik viszont kevésbé erősek az eredmények interpretálásában, mindazonáltal az elemzéseket is sokkal inkább ők készítik.

Tanúi, résztvevői vagyunk az „adattudomány” születésének, amely folyamat-hoz jelentős mértékben járul hozzá az *Educatio* folyóirat tematikus száma. Elsősorban azzal, ahogy a sokasodó adatok régi-új típusú keletkezésének jelenségét és felhasználási lehetőségeit a maguk konkrétságában és történetiségében vizsgálja.

(Educatio, 2015. 24. évfolyam 3. szám)

Hrubos Ildikó

professor emerita
Budapesti Corvinus Egyetem