

Koltay Tibor

## Egy elméleti és gyakorlati kézikönyv a tudományos kutatási adatok menedzseléséről

A közel hatszáz oldalas *Praxishandbuch Forschungsdatenmanagement* (Gyakorlati kézikönyv a kutatási adatkezeléshez) című kötet bevezető tanulmányában arról olvashatunk, hogy a digitális átalakulás – jelentős potenciált és politikai akaratot teremtve – a tudományra és a kutatási adatokra irányítja a figyelmet. A kutatási adatok reflektorfénybe kerülésével pedig világszerte megnőtt az igény arra, hogy világos elméleti és gyakorlati iránymutatást kapjunk a kutatási adatok megbízhatóan tervezhető kezelésére. Ezt figyelembe véve, az egyes írások sorrendjében válogattunk a tanulmányok többségében megjelent gondolatokból.

Az öt fejezet közül az első az *adatok ökoszisztémáját* középpontba állítva, a könyvben tárgyalandó témák elméleti hátterét adja meg. Erről az ökoszisztémáról megtudhatjuk, hogy az formatív és holisztikus, amelyben a különféle érintettek összefognak egymással, mégpedig azzal a céllal, hogy adatokat állítsanak elő, kínáljanak és megtalálhatóvá tegyenek, és hogy feldolgozzák, gazdagítsák, archiválják, publikálják vagy felhasználják ezeket az adatokat. Az adatökoszisztéma keretet és szabályozási feltételeket ad az adatok menedzseléséhez, és hatással van az adatok életciklusának minden szakaszára.

Egy-egy ország adatökoszisztémájának viszont olyan eltérő alrendszerei vannak, mint a kutatási adatok, a közsférában keletkező (adminisztratív), az iparból és a gazdaság más szektoraiból származó (üzleti) és az állampolgárokhoz kötődő adatok. Számos, eddig megválaszolatlan kérdés merül fel ezekkel kapcsolatban.

Ezek közül említsük meg a következőket:

- Kinek a tulajdonát képezik az adatok, ki fér hozzá, és ki használhatja azokat kereskedelmi célokra?
- Hogyan lehet megakadályozni az adatokkal való visszaéléseket, különös tekintettel az érzékeny és személyes adatokra?
- Hogyan tudunk adatokat gyűjteni, elemezni, az emberek számára érthető módon közzé tenni, vagy éppen elavulásuk és más körülmények felmerülése esetén törölni?

Szóba kerül az is, hogy hiba lenne azt hinni, hogy maguk az adatok az igazságot hordozzák, és hogy a modern mesterségesintelligencia-technikák, vagy a gépi tanulás automatikusan megadnák helyes értelmezésüket.

Az adatok megfelelő menedzsmentjének társadalmi mozgatórugói között mindenesetre ott van annak megértése, hogy az adatintenzív kutatás elengedhetetlen a rejtett minták összetett összefüggésekben történő azonosításához, valamint olyan új felismerésekhez, amelyek segíthetnek megbirkózni a „nagy kihívásokkal”. Azt is látnunk kell mindeközben, hogy az adatok értékes árucikkek, amelyek birtoklásáért nemzetközi verseny alakult ki, amely egy-egy országban akár az adatok feletti szuverenitás elvesztéséhez is vezethet.

Természetesen szó esik a kötetben az Európai Nyílt Tudományos Együttműködés (European Open Science Cloud, EOSC) keretében történő kooperáció megvalósításáról is. Gondoljunk csak arra, hogy Németországban, kizárólag az EOSC és az NFDI (Nationale Forschungsdateninfrastruktur – Nemzeti Kutatási Adatinfrastruktúra) programok révén évente mintegy 90 millió eurót költenek az adatinfrastruktúra ökoszisztémájának fejlesztésére. A következő, nagy fontosságú téma a kutatás jogi környezete, amely meghatározza az adatok kezelésével és közzétételével kapcsolatos jogi kereteket, különös tekintettel a szerzői jogi előírásokra, a helyes tudományos gyakorlatra, a szolgáltatási és munkajogi előírásokra, valamint az adatvédelmi törvényekre.

A kutatási adatok kapcsán számos megoldandó problémával kell szembenéznünk. Ezek kutatásetikai megközelítése azonban még gyermekcipőben jár. Ez a megközelítés egyaránt fontos az adatgyűjtés és az adatkommunikáció szempontjából, továbbá szerepet játszik benne a – már fentebb említett – helyes tudományos gyakorlat, valamint az érintettek és a társadalom egészének felelősségvállalása. Ez megköveteli, hogy egy folyamatos etikai diskurzus eredményeként olyan szabványok jöjjenek létre, amelyek a tudományág-specifikus és helyi igényekhez is hozzáigazíthatóvá válnak. Ez magában foglalja továbbá azt a felelősséget, amelyet a kutatási folyamatban részt vevő emberekért, intézményekért, állatokért és tárgyakért kell vállalnunk. Hasonlóan igaz ez a társadalommal és környezetünkkel kapcsolatos viszonyulásunkra is.

*Az adatpiacokkal* foglalkozó fejezet egyike azoknak, amelyek olyan témákkal foglalkoznak, amelyekkel viszonylag ritkán találkozunk. Itt olvashatunk ugyanis

a kereskedelmi és a tudományos adatpiacokról, valamint a városi adatterek és digitális közkezesek kérdéseiről. Az adatok ugyanis olyan erőforrások, amelyek értékét sokan még mindig nagyon magasra becsülik, sőt néha túlbecsülik. Ennek fényében az adatokért folytatott verseny teljes lendülettel folyik. Bár az adatpiacokat a kereslet és a kínálat határozza meg, az adatok értéke az adatok tényleges tartalmától és minőségi jellemzőitől függ, mint például a helyesség, az időszerűség, a pontosság vagy a konzisztencia, és fontos tulajdonságuk, hogy illeszkednek-e az adott üzleti modellekhez.

Jelenleg az üzleti, a tudományos és a közszféra, valamint a polgárok adatpiacai és piacterei elkülönülnek egymástól, azonban a kutatók és az „utca embere” számára egyaránt átláthatatlanok. Bár az adatpiacok adta lehetőségektől függetlenül is nagy, tudománypolitikai szinten is hangsúlyozott igény van az adatok újrafelhasználására, az (főként jogi természetű) akadályokba ütközik ezeken a piactereken. Az adatokkal való kereskedés ugyanakkor egyre fontosabb gazdasági ágazattá kezd válni, aminek hatására új üzleti modellek vannak kialakulóban. A tudományos adatok piacán azonban a magánszektor játszik domináns szerepet. Az elmúlt években sokféle megközelítés született a városok és közösségek digitális infrastruktúráinak továbbfejlesztésére, és ebben a nyilvánosan elérhető adatoknak és metaadatoknak központi szerepet kell kapniuk. Ezek az adatok, amelyeknek egy része köztulajdon, részben vagy egészben felhasználhatók olyan közösségi célokra, mint a társadalmi kohézió és a gazdasági kohézió erősítése.

A kötet harmadik fejezetének tárgya az *adatcultúra*, amely a tudományos kutatási adatok menedzsmentjének összefüggésrendszerében megjelenő kulturális vonatkozásokra koncentrál. Itt főleg a nyitottság és a megosztás új adatcultúráját kellene megteremtünk, mégpedig úgy, hogy az az adatok teljes életciklusát át tudja fogni.

A kutatóknak ugyanakkor tudniuk kell, hogy a digitális világ kínálta lehetőségekkel az eddiginél felelősségteljesebben kell bánniuk, amiben segítségükre lehetnek a digitális kompetenciák és az adatcultúra azon változásai, amelyek a megbízható digitális tudomány irányába mutatnak.

A kutatási adatok és azok menedzselése az elmúlt években szinte minden tudományterületen terítékre került. A könyvtárosok és a kutatók egyrészt interdiszciplináris metaadatszabványok kidolgozására, másrészt az adatok szakterület-specifikus kezelésére törekszenek. Ezek az erőfeszítések megjelennek a könyvtárosképzés tanterveiben, és szilárdan beépültek a kutatók képzésébe is. Ennek kapcsán a német nyelvterületen folyó könyvtári és informatikai képzések jellemzőibe kapunk betekintést, hét németországi és svájci mesterképzés, és tizenkét alapképzés áttekintésével, ide értve a továbbképzést is.

Közben joggal merül fel az a kérdés, hogy miként küszöbölhetjük ki az olyan akadályokat, mint a kutatási infrastruktúra hiányosságai vagy a hagyományos viselkedés során kialakult és tanult gátlások. Nyilvánvaló, hogy ezek megváltoz-

tatásához szükségesek a kutatók egyéni kezdeményezései is, amelyek azonban megkövetelik a kutatókat segítő infrastruktúra fejlesztését. Kellenek pénzügyi és személyes ösztönzők is, továbbá meg kell könnyíteni a kutatási adatok menedzselését. Az olyan tudományterületeken, mint a klímakutatás, az asztrófizika vagy a részecskefizika nemcsak az adatok megosztásával és újrafelhasználásával kapcsolatos gátlások szintje alacsony, hanem kialakult gyakorlattal is számolhatunk. Ez azonban nincs mindenütt így.

Azt is látnunk kell, hogy a kutatási folyamat középpontjában mindig is a kutatók mint egyének álltak, akik motivációt éreztek arra, hogy válaszokat találjanak kutatási kérdéseikre, és az ezekből eredő megoldásokkal tudományos hírnevet szerezzenek maguknak. Ennek megfelelően a figyelem a legtöbb területen azokra az eredményekre irányult, amelyeket a kutatók elsőként publikáltak. Egy ilyen rendszerben, amely erősen a személyes teljesítményhez és a karrierhez igazodik, a nagy erőfeszítésekkel és hosszadalmas folyamatok során gyűjtött kutatási adatok olyan tökélet jelentenek, amely csak akkor válik láthatóvá, ha már átfogóan kiértékeltek és publikáltak. Ezért is kell nagy figyelmet kapnia a FAIR (Findable, Accessible, Interoperable, Reusable – Megtalálható, Hozzáférhető, Szabványos, Újrafelhasználható) alapelveknek, amelyek a kutatási adatok megtalálhatóságát, hozzáférhetőségét, szabványosságát és újrafelhasználhatóságát hivatottak előmozdítani.

Hozzájuk kapcsolódik a GO FAIR kezdeményezés, amelynek célja, hogy hatással legyen a német tudomány rendszerére, mégpedig úgy, hogy az összes kutatási terület és az EU minden tagállamának bevonásával, majd a jövőben határokon átnyúló formában interdiszciplináris kutatási adatokat tegyen újrafelhasználhatóvá. A FAIR alapelvek megvalósítása kultúraváltást igényel, amit a GO CHANGE program segít az ösztönzési rendszerek kialakításával, az adatszaktörők képzésével és továbbképzésével foglalkozó GO TRAIN programmal karöltve. E kettőhöz kapcsolódik a GO BUILD, amely szabványok, protokollok és szolgáltatások harmonizálásának útján az interoperabilitás és a közös adatinfrastruktúrák megteremtését célozza. Az adatmenedzsment iránti figyelem növekedésének hatására az sem meglepő, hogy az ehhez szükséges ismeretek és készségek elsajátítását elősegítő képzési és tanácsadási koncepciók és programok több, a kötetben található tanulmány tárgyát is képezik.

A következő, negyedik fejezet témája éppen a *kutatási adatok menedzselése*, tehát azokról az elvárásokról szól, amelyek a megfelelő adatokkal kapcsolatos feladatok szisztematikus tervezését és végrehajtását feltételezik, méghozzá úgy, hogy közben jogi, etikai és más kontextuális tényezőket is szem előtt kell tartanunk.

Az adatok menedzselése során számos, a tervezést és a végrehajtást befolyásoló tényezővel kell számolnunk. Ezek a tényezők sokfélék, és például az adatökoszisztémából, az adatpiacok meglétéből vagy az adatkultúra sajátosságaiból eredhetnek. Ehhez különböző szoftverek, valamint a meglévő projek-

tek és kezdeményezések eredményei és megállapításai is felhasználhatók. Ezek mellett olyan folyamatokat kell kezelnünk, amelyek magukban foglalják az adat megosztásának, közzétételének és egy adott projekt befejezésén túli, későbbi felhasználásának kezelését is. Ezen a téren meg kell oldanunk az adatok megtervezését, leírását és dokumentálását, majd létre kell hoznunk az ezeket követő adattárolás, gondozás és a hosszú távú hozzáférhetőség megteremtéséhez szükséges eszközöket.

Figyelembe véve, hogy az adatok tervezése az egyszerű információktól az adatok kezelésére vonatkozó ajánlásokon át a konkrét követelmények meghatározásáig terjed, az adatok menedzselésének tervezése még mindig elég következtelen. Első lépése mindenesetre egy-egy adatkezelési terv (DMP, németül *Datamanagement Plan*, angolul *Data Management Plan*) összeállítása. Készítésüket már számos ország finanszírozó testületei és az EU is előírja. Ezeknek a formális dokumentumoknak a lényege a kutatási adatok menedzselésére vonatkozó általános jellegű kérdések megválaszolása, érintve az adatok gyűjtésének, rendszerezésének, (biztonságos) tárolásának, archiválásának és megsemmisítésének módját, figyelembe véve a nyílt hozzáférés diktálta követelményeket is.

Feltételezhető, hogy sokak számára újdonságot jelent az *adattárfelhasználás* fogalma, amely a nyers adatok értelmezhető és újrafelhasználható adatokká való átalakítását jelenti. Ennek keretében gondoskodnunk kell az adatok megtalálhatóságáról és felhasználhatóságáról is.

Az adatok hosszú távú tárolása és fenntartható hozzáférhetősége viszont sokkal régebb óta meglevő és fontos kívánalom a kutatásban. A kísérletek, mérések, szimulációk vagy értékelések ugyanis olyan adatokat szolgáltatnak, amelyekhez optimális tárolási lehetőségeket kell találni. A kutatás ugyanis olyan igényeket támaszt, amelyek túlmutatnak az adatok pusztá tárolásán. Ennek azonban nincs univerzális módszere, hanem minden egyes kutatási projekthez meg kell találnunk a megfelelő megközelítést. A kutatási adatok értékelésétől függően különböző megőrzési stratégiák alkalmazhatók, amelyek viszont eltérő követelményeket támasztanak a fizikai tárolással és a hozzáféréssel szemben. A szelekció és az adatok kurátori gondozása ezért is játszik egyre fontosabb szerepet a kutatási adatok értékelésében.

A *minőségmenedzselés* (minőségirányítás) témájával foglalkozó fejezet egyik szerzője Király Péter, akinek a neve sokak számára ismerősen hangozhat idehaza. Ebben a fejezetben arról is olvashatunk, hogy az utóbbi évtizedekben a múzeumok, levéltárak és könyvtárak számos rendszert dolgoztak ki digitális állományaik nyilvántartására. Ezt természetesen e sorok olvasói is jól tudják. Kérdés azonban, hogy a kutatási adatok esetében ez hogyan valósítható meg. Bár a minőségnek önmagában nincs egységes definíciója, a szakirodalomban többnyire azt olvashatjuk, hogy – minimális követelményként – összhangban kell lennie a céloknak való megfeleléssel.

Amikor (még) nem állnak rendelkezésre a tudományos kutatás megfelelő infrastruktúrái és folyamatai, az adatok menedzselésének mozgatórugói a projektek lehetnek. Repozitóriumok, szakpolitikai eszközök, oktatási anyagok, valamint számos eszköz és módszer született meg projektek keretében. Ideális esetben az ilyen projektek eredményei hozzáadott értéket jelentenek az adott tudományág vagy az egész kutatói közösség számára, és hosszú távon megtalálják a helyüket egy vagy több intézmény szolgáltatási palettáján, valamint egy olyan közösségekben, amely továbbra is fenntartja őket. Ahhoz, hogy a projektek eredményei ne csak egyedi problémákat oldjanak meg, feltétlenül szükséges, hogy a projektpartnerek között ott legyenek az infrastrukturális intézmények, például könyvtárak, számítóközpontok és a kutatás-igazgatás szervezetei. Legalább ilyen fontos projektpartnerek azok a tudományos intézmények, amelyek meg tudják határozni a követelményeket, és biztosítják, hogy az eredmények felhasználhatók és reálisan beilleszthetők legyenek a kutatási folyamatba.

Az adatok visszakereshetőségével kapcsolatban is számos kihívással kell szembenéznünk. Ahhoz, hogy a keresési funkciók megfelelő módon működjenek, a keresés alapjául szolgáló információs igények megértésétől kezdve, az adathalmazok tartalomalapú rangsorolásának modelljein át, a keresési keretrendszeréig sok mindenre szükség van. Ide tartozik az is, hogy különbséget tegyünk a visszakeresés szempontjából értékelhető adatok különböző típusai, nevezetesen a metaadatok és a kutatási adatok között.

Az adatokkal végzett, együttműködésen alapuló munka eredményeként számos kutatási adat gyűjthető össze, majd válhat integrálhatóvá, elemezhetővé és hozzáférhetővé a tudományos közösség számára. Ahhoz, hogy ezek megtalálásához a keresési funkciókat megfelelően lehessen használni, az adatok visszakeresésének több szempontját kell szem előtt tartanunk. Ilyen szempont lehet például a keresés alapjául szolgáló információs igények megértése, vagy az adathalmazok tartalomalapú rangsorolásának modelljei.

Az adatok vizuális megjelenítése fontos kérdésként szerepel a kötetben. Az adatvizualizáció magában foglalja a digitális adatok bármilyen vizuális ábrázolását a számítógép-alapú adatok rendszerében. Még ha az infografikák megjelenítése nem is mindig számítógépes vetítéssel történik, akkor is adatvizualizációról beszélünk. Az adatvizualizáció tehát a vizuális megjelenítés különböző típusainak gyűjtőfogalma, lényege pedig a heterogén adatok kétdimenziós felületre, például képernyőre történő vizuális kivetítése.

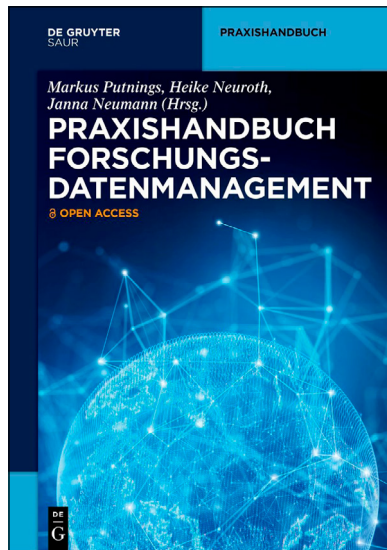
A digitális kutatási adatok publikálása és idézése kapcsán tudjuk, hogy az adatok önálló információs objektumként adatrepozitóriumokban tehetőek közzé, ami kiegészülhet azzal, hogy adatfolyóiratokban közölt, lektorált cikkek formájában is megjelenhetnek, és ez lehetővé teszi, hogy ne csak magukat a kutatási adatokat írjuk le, hanem eredetiségüket, minőségüket, valamint hozzáférhetőségüket és ismételt felhasználhatóságukat is biztosítsuk.

Bár a kötet zárótanulmányának címe – *A kutatási adatok menedzselésének jelenlegi fejleményei és kihívásai Németországban* – magáért beszél, elmondhatjuk, hogy rámutat azokra a kihívásokra, amelyekkel a nemzeti, a tartományi, a helyi, valamint a nemzetközi szintű együttműködésekben részt vevő tudományos közösségeknek szembe kell nézniük.

A Praxishandbuch Forschungsdatenmanagement című nyomtatásban és nyílt hozzáféréssel, elektronikusan is elérhető kézikönyvet a közmondásos német alaposág szép példájának tekinthetjük. Ez a több mint ötven szakember munkájával létrehozott mű úgy szól a gyakorlatról, hogy közben bőségesen tárgyalja az elméleti kérdések egész sorát. Mindemellett számos téma (például a FAIR alapelvek) búvópatakkén hol az egyik, hol a másik fejezetben kerül elő.

Recenzióknak csupán rövid és erősen válogatott betekintést ad a kötet írásainak témáiba, amelyek azonban gazdagságuk és sokrétűségük folytán bőven kínálnak olvasnivalót. Úgy gondoljuk, hogy üdvös volna, ha egyes részei magyar nyelvre lefordítva az érdeklődő hazai szakemberek szélesebb köréhez is eljutna.

Praxishandbuch Forschungsdatenmanagement / herausgegeben von Markus Putnings, Heike Neuroth und Janna Neumann. - Berlin : De Gruyter Saur, 2021. - 587 p. : ill. ; 25 cm



Elektronikus hozzáférés: <https://doi.org/10.1515/9783110657807>