



A kutatási adatok és a minőség*

KOLTAY Tibor

Bevezető

Az adatokhoz való nyílt hozzáférés számos területen, így az oktatásban, a munkáltatók, a kormányzatok és a nonprofit szervezetek adatai tekintetében is, nagy jelentőségre tett szert.¹ Azt is tudjuk, hogy a tudományos kutatás negyedik paradigmája az adatokat állította középpontba,² ami abból a szempontból is fontos, hogy – amint a tudományban a dokumentum-központú szemléletet (fokozatosan) felváltja vagy legalábbis kiegészíti az adatközpontú gondolkodás – a könyvtáraknak is „lépést kell váltaniuk”.³ Ez az írás éppen ezért a tudomány művelése során létrejövő adatokra koncentrál. Ezeket az adatokat olyan heterogén és kontextualizált objektumokként határozhatjuk meg, amelyek a különböző tudományterületeken végzett, szisztematikus vizsgálódás nyomán keletkeznek.⁴

A kutatási adatok minőségéről szólva ugyanakkor elkerülhetetlen, hogy az üzleti adatokról is szóljunk. Ezeket többnyire vagyontárgyként kezelik, és velük szemben elvárás, hogy hasznot hozzanak tulajdonosuknak.⁵ Ráadásul az üzleti adatokat egészen az 1990-

es évekig nem értékes üzleti forrásnak, hanem olyan mellékterméknek tekintették, amely a tranzakciók befejeztével elveszíti értékét.⁶ Ma már viszont azt mondhatjuk, hogy a minőség fontos a tudományos kutatásban és az üzleti életben is. E két adattípus között tehát nemcsak különbségek vannak, hanem az is látható, hogy az üzleti adatokkal szemben támasztott minőségi követelmények jelentős mértékben egybeesnek a tudományos adatokéval. Mielőtt azonban a kutatási adatok minőségi jellemzőit megvizsgálánánk, figyelmet kell szentelnünk azoknak a változásoknak, amelyek az adatok és az információ természetére vonatkozó nézeteink átmentek.

Bár az adatokkal kapcsolatos minőségbiztosítás terén számos kihívással kell szembenézniük,⁷ a legfontosabb érdekeltek a kutatók. A kutatási adatok minőségének téje van a nemzetközi, a nemzeti és az ágazati szinten működő finanszírozó szervezetek számára is.⁸ A tudományos kiadók más módon is tudják segíteni az átláthatóságot és a nyitottságot:

- javítják a lektorálás minőségét és objektivitását;
- segítségükkel könnyebben megérthetjük a kuta-

- tók problémáit és motivációit;
- számos problémára rámutatva segítik a kutatói kultúraváltást;
- a repozitóriumokkal karöltve elősegítik a nyílt tudomány meggyökeresését.⁹

Természetesen nem feledkezhetünk könyvtárak szerepéről, – részvételük fontosságáról és néhány részletkérdéséről már több magyar nyelvű írás is megjelent.¹⁰

Az adatminőség természete

A minőség összetett kérdéseinek megismerése nem lehet teljes az adatok ökológiájának ismerete nélkül. Az ökológiai elvek ugyanis magukba foglalják az adatok újrafelhasználását és az ezt elősegítő szolgáltatásokat, amivel az adatvezérelt, digitális tudomány egy lehetséges keretét adják meg.¹¹

A kutatási adatok menedzselésének céljai között ott van a hozzáférhető és újrafelhasználható, tehát jól dokumentált és jó minőségű adatok szolgáltatása.¹² Az adatok minőségének biztosítása része az adatműveltségnek is, amely a kritikus gondolkodás hangsúlyozásával szorosan kapcsolódik az információs műveltséghez,¹³ és amelyet úgy határozhatunk meg, mint az adatok megkereséséhez, elemzéséhez, szervezéséhez, bemutatásához és értékeléséhez szükséges kompetenciák együttesét.¹⁴

Az adatok és az információ viszonyáról alkotott elképzelések

Az adatok és információk között különbség van. Ez nem strukturális jellegű, hanem funkcionális természetű, tehát az adatok – a hagyományos felfogás szerint legalábbis – nem jelentenek tudást, míg az információ a Mit, a tudás a Hogyan kérdésre ad választ.¹⁵ Némileg más megközelítésben nézve az információ feldolgozott adat, amely nemcsak a Mit, hanem a Ki, a Hol és a Mikor szavakkal kezdődő kérdésekre válaszol, amihez hozzátehetjük, hogy a bölcsesség az okok (miértek) ismeretét is jelenti.¹⁶ Ebben a kontextusban tehát az adatokra úgy szokás tekinteni, mint az elemzést és a tényeket megelőző, az interpretáció és az érvek előtti nyersanyagra. Természetesen az ilyen vagy hasonló alapokon nyugvó elképzelések fő pillére az a megközelítés, hogy az adatból meghatározható az információ, az információból a tudás, még hozzá feltételezve, hogy ezek a folyamatok az ellenkező irányban is működnek.¹⁷ Ezt

a szemléletet az az értékpiramis illusztrálja, amelynek a legalján, az információ szintje alatt helyezik el az adatokat.¹⁸ Mindezt azonban napjainkban többen is megkérdőjelezzik. Az adatok és az információk ugyanis kölcsönhatásban állnak egymással, és értéküket az a cél határozza meg, amelynek elérésére felhasználjuk őket.¹⁹

Az ellentétek fényében nagy súlyt kap az, a digitális bölcsészet szempontjából egyébként is alapvető jelentőségű feltételezés, hogy a szöveget adatokként, az adatokat pedig szövegeként értelmezhetjük,²⁰ különös tekintettel arra, hogy célközönségük meghatározható meghatározható.²¹ Ráadásul azt az állítást is elfogadhatjuk, hogy bármi lehet információ, ami informáló erővel bír, tehát gyakorlatilag lehetetlen eldönteni, hogy valami információ-e vagy sem, különös tekintettel arra, hogy a hibás vagy a soha nem látott információ sem zárható ki az információk köréből.²² Figyelembe véve tehát az interpretálható tárgyak, szimbólumaik és interpretációik közötti összefüggéseket, adatnak tekinthetünk bármit, ami szemantikai és pragmatikai szempontból megfelelő módon rögzíthető adatbázisokban. Amit ugyanis rögzítünk, annak szemantikai szempontból igaz vagy hamis állításnak kell lennie.²³

A rögzített jellegnek ez a szükségszerűsége megjelenik *Michael Buckland* jól ismert tipológiájában, amelyben kimondja, hogy az információnak három létformája van.²⁴ Az információ lehet átadott tudás, tehát az, amit annak eredményeként, hogy informáltak bennünket, megtanultunk. Ez az első létforma, vagyis az *információ-mint-tudás*. Ugyanakkor nézhetjük azt a folyamatot is, amely akkor valósul meg, amikor informálnak bennünket. Ez a létforma az *információ-mint-folyamat*, amely tudatállapotunkat módosíthatja. Az információ dolog (tárgy) is, tehát bitek, bájtok sorozata, vagy alkothatják hangok, képek és bármi, ami fizikailag jelenéssel bíró entitásként észlelhető, amelyet ki tudunk fejezni, le tudunk írni, reprezentációk formájában tükrözni tudunk, vagy fizikailag (jelként) képviselve van. Ennek a létformának a megtestesülése nem korlátozódik a szöveges médiára, tehát lehet a könyv és a szöveg mellett akár múzeumi tárgy és adat is. Ez a harmadik létforma, amelyet az *információ-mint-dolog* elnevezéssel szoktunk említeni.

Ha ezt a hármas felosztást kiindulási pontnak tekintjük, akkor érdemes elgondolkoznunk azon, hogy az adatokra is kiterjeszthető-e. Az adatok informálási képességét figyelembe véve úgy gondolom, hogy – az információ-mint-tudás mintájára – az adat-

mint-folyamat fogalmáról beszélhetünk. Az adatok tárgyi megvalósulása pedig az *adat-mint-dolog*. Az információ-mint-tudás létformának az adatokra való alkalmazása azonban kérdéses. Ez elsősorban azért van így, mert a kritikai megközelítések jegyében megkérdőjelezhető, ha egyenlőségjelet teszünk az adatok és az információk közé.²⁵ Mindazonáltal hasznosabbnak tűnik, ha a kritikák ellenére elfogadjuk *Bawden* és *Robinson* érvelését, miszerint az információkat és az adatokat célszerű egymáshoz szorosan kapcsolódó fogalmaknak tekintenünk.²⁶ Az adatok és az információ közötti szoros kapcsolat elfogadását az is mutatja, hogy az adatumveltség elméletével és gyakorlatával foglalkozó szakemberek egy része az *adat- és információs műveltség* (data information literacy) terminust használja.²⁷

Azon kívül, hogy érintettük az adatok és az információ közötti hasonlóságokat és különbözőségeket, valamint kitértünk a szövegekhez való viszonyukra, a dokumentum fogalmával is érdemes röviden foglalkoznunk. *Paul Otlet* ugyanis 1934-ben kiadott *Traité de documentation* című könyvében nemcsak egy olyan rendszert írt le, amely a modern digitális eszközök és platformok, mindenekelőtt a hipertext és a World Wide Web elődjének tekinthető, hanem aláhúzta, hogy a dokumentum rögzíti, amit értékeltünk és elménkben megalkottunk, tehát nemcsak a fizikai vagy elvont dolgok reprezentációi, hanem maguk a dolgok is dokumentumoknak tekinthetők. Mivel az intézmények mellett mi magunk is képesek vagyunk dokumentumokat létrehozni magukról és tevékenységükről, lehetséges, hogy ezeknek az adatoknak és metaadatoknak a felhasználásával modellezhetővé váljunk,²⁸ tehát szokásainkat és preferenciáinkat akaratlanul is felfedjük. Az adatoknak az információhoz, a szöveghez és a dokumentumhoz való viszonya kapcsán azonban elég leszögezünk, ebben az esetben is felfedezhető köztük a rokonság.

Részösszegzésként leszögezhetjük, hogy nincs megfelelő képünk azoknak a folyamatoknak a természetéről, amelyek az adatokat, az információt és a tudást egymással összekötik. Fontos viszont hangsúlyoznunk, hogy a kutatási adatok nemcsak empirikus kutatómunka eredményei vagy statisztikai elemzések nyersanyagai, hanem „saját jogon” kutatási tárgyak is,²⁹ ami erősíti a tudomány világának az adatok iránt érzett elkötelezettségét. Mivel a könyvtár- és információtudomány érdeklődési körébe mindig is beletartoztak az információ szervezésének és terjesztésének problémái, az adatok és az információk közötti rokonság okán tényként kezelhetjük, hogy e

tudományterület egyre gyakrabban foglalkozik az adatok kezelésének, gondozásának és szervezésének kérdéseivel.³⁰

Az adatminőség attribútumai

Az adatminőségnek számos attribútuma, azaz a minőségtől elválaszthatatlan tulajdonsága, jellemzője van, amelyeket sokféleképpen csoportosíthatunk. *Wang* és *Strong* 1996-ban például 179 minőségi jellemzőt sorolt fel.³¹

Az adatminőséget olyan külsőleges sajátságok is meghatározhatják, mint az adatok használatra való alkalmassága.³² Ugyanakkor a digitalizált objektumoknak vannak belső lényegükből fakadó, intrinzikus értékei is, amelyek függetlenek attól, hogy használják, vagy valaha használták-e az adatokat.³³ Ezek közé tartozik a pontosság, az objektivitás és a hihetőség, amelyeket a minőség később említendő, szemantikai attribútumai között is megtalálunk. A kontextuális jellemzők közé a relevancia, a hozzáadott érték megléte, az időszerűség, a teljesség és az adatok mennyisége tartozik.³⁴ A kontextus esetükben nem annyira szövegösszefüggést jelent, mint inkább az adatok alkalmazásából és a kutatás teremtette szituációkból eredő tulajdonságokat határozza meg.

Vannak ezeken kívül olyan, az időbeliséggel összefüggésbe hozható jellemzők, mint az érvényesség, az illékonyág (gyors elavulás), amelyek általános jellegüknél fogva számos kontextusban megjelenhetnek, ezért is találjuk itt az időszerűséget, amellyel már a kontextuális jellemzők kapcsán találkoztunk.³⁵ Ezeket a jellemzőket kiegészíthetjük a teljességgel, az összehasonlíthatósággal, a széleskörűséggel és az eredetiséggel.³⁶

A legfontosabb minőségi jellemzők egyike a bizalom, amelyet az adott szakterület jellemzői mellett az adott kutató jó hírneve és az adatokat értékelők előítéletei határoznak meg. Az adatok életciklusának egy másik pontján úgy is megközelíthető, hogy segítségével arra keresünk válaszokat, hogy a kérdéses adatállományt bírálta-e szerzőjén kívül más is.³⁷

A bizalom vonatkozhat magára az adatállományra, vagy kiterjedhet valamely repozitóriumban elhelyezett példányára is.³⁸ Több összetevője is van. Ezek közül az érthetőség annak függvénye, hogy elegendő kontextus áll-e rendelkezésre az adatok eredetéről. A használhatóság azt jelenti, hogy az adatok megtalálhatók, hozzáférhetők és használható formátumban vannak. Az adatoknak emellett a bitek szintjén azonosnak kell lenniük egy korábbi, elfogadott és ellen-

őrzött állapottal.³⁹ A bizalom része, hogy az adatok szolgáltatójának kultúrája garanciát jelentsen arra, hogy biztosítani tudja a felhasználókat az adatok számos jellemzőjének megfelelő voltáról. A felsőoktatási intézmények és a kutatóintézetek esetében a bizalom lényegében automatikus, ha egyébként nyitottságot és átláthatóságot mutatnak.⁴⁰

Azt is tudnunk kell, hogy a bizalom megítélése szubjektív, hiszen mértékét befolyásolja az adott szakterületen kialakult gyakorlat, valamint azok jóhírneve, akik az adatok létrehozásáért felelősek, de hatással vannak rá az adatokat értékelő személyek előítéletei is.⁴¹ Az adatrepozitóriumok esetében a bizalom lényege, hogy úgy gondoljuk, nem lehet szó félrevezetésről, viszont ez nincs összefüggésben avval a bizalommal, amely magukra az adatokra vonatkozik.⁴² Az eredetiség szintén az adatok minőségének alapvető attribútuma, amely azt mutatja, hogy megítélésünk szerint egy adott adatállomány a tudományos kutatás megfelelő módját képviseli-e, ide értve az adatok gyűjtésére használt műszerek megbízhatóságát és a kutatást megalapozó elméletek helyességét, az adatok teljes voltát, pontosságát és érvényességét.⁴³ Az érvényesség egyúttal olyan dimenziókkal függ össze, mint az időszerűség, a teljesség és a biztonságosság.⁴⁴ Hogy eredetiségüket meg tudjuk ítélni, az adatoknak érthetőnek kell lenniük, amit a dokumentáció és a metaadatok megléte folytán létrejövő kontextus tud biztosítani. Ehhez használhatónak is kell lenniük, ami megköveteli a megfelelő formátumot és azt, hogy felfedezhető és hozzáférhető legyenek. Azoknak, akik az adatok minőségét megítélik, szükségük van egy olyan eszközre is, amelyet az integritás (sértetlenség) jellemez.⁴⁵

A szemiotikai megközelítés természetét a mondatok példájával illusztrálhatjuk. Egy mondat minősége jó, ha a megfelelő betűkből, szavakból áll, és követi a nyelvtani szabályokat. Ha viszont nem anyanyelvi beszélőként nem, vagy csak részben értünk egy mondatot, akkor minőségét rossznak érezzük. Mindez arra hívja fel a figyelmünket, hogy az adatok minőségét a kódolásukra, tárolásukra és kommunikációjukra vonatkozó, előre rögzített specifikációk alapján kell megítélnünk. Amikor tehát a minőség attribútumait szemiotikai szempontból csoportosíthatjuk, a tapasztalat (empíria) szintjén ott találjuk az időszerűséget, a biztonságos jelleget, a nyomon követhetőséget és

a hozzáférhetőséget. A szintaktikát tekintve szükség van a pontosságra, a lényegre törő reprezentációra, a konzisztens jellegre, a könnyű kezelhetőségre, míg a szemantikai attribútumok közé tartozik az egyértelműség, a hihetőség és az interpretálhatóság.

Az adatok minősége és az újrafelhasználás

Az adatok újrafelhasználása kapcsán újra felmerül az adatminőség számos, már említett attribútuma. Van-e esetek, amikor nem csak azok a kutatók akarják használni az adatokat, akik eredetileg összegyűjtötték azokat. Ahhoz azonban, hogy mások is használhassák őket, jó minőségű és gondosan kezelt adatállományokra van szükség. Ezzel biztosítható ugyanis ezeknek az adatoknak az integritása és eredetisége. Ha a tudományos szigor és a jó minőségű tudományos kutatás összekapcsolása eredményeként létrejövő intrinzikus jellemzőket nézzük, akkor azok hagyományosan nem tartoznak az adatok kurátorok által történő gondozásának körébe. Ugyanakkor viszont érdemes elgondolkoznunk azon, hogy ezek a jellemzők miként integrálhatók az adatgondozás gyakorlatába, hiszen az amúgy is számos, eltérő végzettségű és munkakörű szakember együttműködésében valósul meg.

Az újrafelhasználással kapcsolatos minőségi döntéseket befolyásolja az is, hogy a kutatók milyen fogalmak, személyes tudás és készségek alapján választanak ki adatokat, és mit tekintenek hibának.⁴⁶ Az olyan minőségi jellemzők, mint a hozzáférhetőség, a teljesség, a hitelesség, valamint az adatokhoz tartozó dokumentáció is befolyásolják az adatállományok felhasználását. Az újrafelhasználást az elnevezésekre vonatkozó megegyezések, a felhasznált mértékegységek, valamint a létrehozás és az aktualizálás dátumai is ösztönözhetik, de akadályozhatják is.⁴⁷ Ebben a körben fontos még minden olyan információ, amely a karbantartás aktuális voltára, a kutatás finanszírozójának kilétére, valamint a korábbi felhasználásokra és aktualizálásokra vonatkozik.

Szintén létfontosságú, hogy az alábbi három kérdésre választ kapjunk:

- Ki felelős a minőség ellenőrzéséért?
- Milyen folyamatok zajlanak le ennek az az ellenőrzésnek a folyamán?
- Hogyan oldhatók meg a hiányzó adatok okozta problémák?⁴⁸

A technikai és a tudományos minőség

Az újrafelhasználás szempontjából lényeges, hogy megfelelő legyen a technikai és tudományos adatminőség, amelyeket azért is kell figyelnünk, mert ez az első lépés ahhoz, hogy megteremtsük a szükséges bizalmat az adatok potenciális felhasználóinak körében.⁴⁹

A technikai minőség annak mértéke, hogy mennyire tartja be egy-egy kutató vagy kutatócsoport a szabványok előírásait és a megfelelő szoftvereket használja-e. Attribútumai közé tartozik még az adatállományok teljessége és konzisztens volta is.

A publikációk minőségét a tudományos közösség előzetes és utólagos lektorálás (szakértői véleményezés) útján állapítja meg. Ez a megközelítés közismerten azon a feltevésen alapul, hogy a lektorok többnyire nem anyagi előnyök fejében, hanem szakmai tekintélyük növelése és az adott tudományos közösséggel szemben érzett lojalitás által vezérelve dolgoznak.⁵⁰ Tudnunk kell viszont, hogy – a tudományos könyvekkel és cikkekkel ellentétben – az adatok esetében nincs lektorálás, pedig az jelentős mértékben garantálná a megfelelő minőséget.⁵¹

A tudományos közösségnek ezért is van szüksége arra, hogy az adatok megosztásához és újrafelhasználására megfelelő mechanizmusok és fórumok álljanak rendelkezésre. Ennek megteremtésében segítenek a kiadók az úgynevezett adatfolyóiratokat megjelenítésével. Ezek többnyire nyílt hozzáférésűek és a folyóiratcikkekhez hasonló formában tesznek közzé metaadat-dokumentumokat, amelyek online hozzáférhető adatkészleteket vagy adatkészlet-csoportokat írnak le. Ez azt jelenti, hogy ezek a kereshető cikkek sok tekintetben megfelelnek a szokásos tudományos közlés gyakorlatának.⁵² Esetükben arról van szó, hogy a kutatók nem magukra az adatkészletekre, hanem az adatfolyóiratokban megjelenő cikkekre hivatkoznak, ami – bár nem jár egységes formai követelményekkel – egyszerűbbé teszi a hivatkozás aktusát. Mivel azonban nincs teljes egyetértés a lektorálás és az adatok minősége közötti összefüggés természetéről, különösen fontos, hogy figyeljük a megbízható és minősített adatrepozitóriumok adatfolyóiratokban megjelenő ajánlólistáit.⁵³

Egyéb minőségi tényezők

Ha az adatok ökoszisztémáját vizsgálva arra is odafigyelünk, hogy az adatok gyűjtése, kezelése és felhasználása etikai szempontból is megfelelő módon

történik-e, akkor jól szolgáljuk az adatokba vetett (már többször említett) bizalom megerősítését. A korábban szintén már említett metaadatok is nélkülözhetetlenek az adatminőség szempontjából, mivel azok számos minőségi paraméterről, például az adatok eredetéről és eredetiségéről informálnak.⁵⁴

Az adatszabványoknak az adatok minőségére gyakorolt hatása az adatok típusától függően más és más. Ez részben attól is függ, milyen a minőséggel kapcsolatos szemléletünk. Koncentrálnunk a specifikációknak való megfelelésre, vagy az adatok használatra való alkalmasságát tekinthetjük igazán fontosnak. Bár ezek nem feltétlenül és nem gyökeresen különböznek egymástól, sokszor nem ugyanazt jelentik. Mindazonáltal, figyelmet érdemel az ISO/IEC 25012 szabvány, amely a minőségi jellemzőket (tulajdonságokat) három kategóriába sorolja:

- az adatok lényegéből fakadó (inherens),
- rendszerfüggő,
- a kettőt egyesítő (inherens és rendszerfüggő) tulajdonságok.

Az inherens tulajdonságok elsősorban az adatoknak az adott szakterületen nekik tulajdonított értékét, használatuk lehetséges korlátait, valamint az értékekhez és a metaadatokhoz való viszonyait tükrözik. A rendszerfüggő tulajdonságok természetét magától értetődőnek tekinthetjük, mivel az adott szakterülettől és az alkalmazott technológiáktól függ, hogy mely adatok kerülnek felhasználásra. Az is meghatározó jelentőségű kérdés, hogy miként érhető el és tartható fenn a minőség megfelelő szintje az adott számítógépes rendszerben. Számolnunk kell a fenti két jellemzőt ötvöző (hibrid) kategória meglétével is. Az inherens minőségi jellemzők között a szintaktikai és a szemantikai pontosságot találjuk. A teljesség jelzi, hogy az adatok minden egyes előfordulása az attribútum elvárt értékét mutatja-e fel. A konzisztencia annak a jele, hogy az adatok más adatokkal koherensek, valamint mentesek az ellentmondástól. A hihetőség azt jelenti, hogy az adatokat valódinak és hihetőnek tekintjük: ennek része az eredetiség, vagyis az adatok forrásainak hitelessége.⁵⁵ A minőségnek ezt a típusát *Strong, Lee* és *Wang* 1997-ben úgy írta le, mint a pontosság, az objektivitás, a hihetőség, valamint a jó hírnév együttesét.⁵⁶ Az utóbbiak közül a pontosság több esetben előfordul az adatminőséggel kapcsolatos írásokban, még ha több, egymástól eltérő keretbe illesztik is be őket.

A már említett ISO/IEC 25012 szabvány első rendszerfüggő tulajdonságként a rendelkezésre állást említi, ami azt garantálja, hogy csak a megfelelő joga-

sultással rendelkező felhasználók és/vagy alkalmazások férhessenek hozzá az adott adatállományhoz. A hordozhatóság lehetővé teszi, hogy az adatok egyik rendszerből a másikba mozgathatók, ott installálhatók és felülírhatók legyenek úgy, hogy közben meglévő minőségüket megőrizzük. A visszaállíthatóság megléte esetében az adatok minőségének meghatározott szintje még a rendszerek meghibásodás esetében is megmarad.

Az inherens és rendszerfüggő tulajdonságokat magukba foglaló tulajdonságok közül a – már fentebb említett – hozzáférhetőség lehetővé teszi az adatok változó kontextusokban történő felhasználását. A megfelelés azt mutatja meg, hogy az adatok mennyire követik a szabványok, megegyezések és egyéb szabályozások előírásait. A biztonság – a titkossággal párosítva – lehetővé teszi, hogy csak az arra feljogosított személyek férhessenek hozzá az adatokhoz. A hatékonyság azt jelenti, hogy az adatok az elvárható teljesítményszintnek megfelelően dolgozhatók fel. Talán mondanunk sem kell, hogy a pontosságnak ebben a kontextusban az attribútumok tekintetében kell érvényesülnie. A nyomonkövethetőség azon múlik, hogy az adatok attribútumai mennyire teszik lehetővé a hozzáféréshez kapcsolódó ellenőrzés vizskövethetőségét.

Az érthetőség viszont lehetővé teszi, hogy az adatokat a felhasználók olvasni és értelmezni tudják, amihez általában hozzájárul a metaadatok megléte.⁵⁷

Bár csak közvetve hathat az adatok minőségre, érdemes odafigyelni az adatkezelési tervek problémáira. Ezek elkészítését számos, a kutatásokat finanszírozó szervezet előírja. Sok kutatónak azonban nincsenek meg a készítésükhöz szükséges ismeretei.⁵⁸ Jó hír, hogy világszerte és Európában is számos könyvtár kínál olyan szakszerű szolgáltatást, ami megoldásokat kínál erre a problémára.⁵⁹

A nagy adatok minősége

A nagy adatok (big data) megközelítése sok tekintetben megegyezik azzal, ahogy általában is viszonyulunk az adatokhoz, tehát az emberi elme mintákat keres és eszközöket talál arra, hogy kezelni tudja a tárgyak komplexitását.⁶⁰ Ugyanakkor hiányos jellegük és gyakran bizonytalan természetük negatívan befolyásolja minőségüket, ami egyre fokozódó veszéllyé válhat. Mindazonáltal jövőbeli kutatásoknak

kell majd feltárniuk, hogy a minőségnek mely attribútumai alapján határozhatjuk meg azt, hogy mennyire tekinthetők adekvátnak bizonyos adatállományok, továbbá hogy ezeknek a paramétereknek milyen mértékben kell pontosnak lenniük.⁶¹

Ettől függetlenül a minőségi jellemzők többségének érvényesnek kell lennie a nagy adatokra is, azonban esetükben nem mindig egyértelmű, mit jelentenek ezek az adatok, és milyen kontextusban gyűjtötték őket, ezért a döntéshozattal kapcsolatos érvényességük megkérdőjelezhető.⁶² Ez egyúttal arra is rámutat, hogy az adatok és az információ közötti különbséggel kapcsolatos viták mellett további elméleti kérdések is megválaszolásra várnak.

A nagy adatokat mindenesetre három tulajdonság meglétével szoktuk jellemezni. Ezek a mennyiség, a sebesség és a változatosság. Újabban az értéket is idesoroljuk, amihez mára már az érvényesség is csatlakozott. Az utóbbi két aspektus az adatok eredetétől, gyűjtésük és feldolgozásuk módjától függő statisztikai megbízhatóság elérése kapcsán is fontos. A megbízhatóságnak az adatok életciklusának kezdetétől a végéig, vagyis származásuk és tárolásuk tekintetében is meg kell lennie. Fontos további tényezők még:

- az adatok integritása,
- a tároló számítógépek és tárolási platformok megbízhatósága,
- a rendelkezésre állás és az időszerűség,
- a felelősség és a jó hírnév.⁶³

Végül, de nem utolsósorban (és főként a nagy adatok kapcsán) nem feledkezhetünk meg arról, hogy ma már nemcsak információs túlterhelésről beszélhetünk, hanem az adattúlterhelés meglétét is tudomásul kell vennünk.⁶⁴

Összegzés

A tudományos kutatás és az adatok minősége közötti kapcsolat lényegét *Christine Borgman* fogalmazta meg a legfrappánsabb formában. Szerinte ugyanis a kutatók nem több adatot igényelnek, hanem a megfelelő adatokra van szükségük.⁶⁵ Ezt erősíti meg *Frické*, amikor úgy fogalmaz, hogy a tudomány eredményes működésének nem több adaton, hanem több elméleten kellene alapulnia.⁶⁶ Ez azt is jelenti, hogy ha a minőségre figyelünk, az sokkal hasznosabb, mint ha az adatok mennyiségét növelnénk.⁶⁷

Jegyzetek és irodalmi hivatkozások

- * A cikk megjelenését az EFOP-3.6.1-16-2016-00001 „Kutatási kapacitások és szolgáltatások komplex fejlesztése az Eszterházy Károly Egyetemen” projekt támogatta.
- CORRALL, Sheila. Repositioning data literacy as a mission-critical competence. == ACRL 2019 [online]: Recasting the Narrative, April 10-13, 2019, Cleveland, OH. Hozzáférhető: <http://d-scholarship.pitt.edu/id/eprint/36975> [Megtekintve: 2020. június 15.]
 - HEY, Tony – HEY, Jessie. E-science and its implications for the library community. == Library Hi Tech, 24. (2006) 4., p. 515-528. ISSN 1737-8831
 - CHOUDURY, Sayeed. Data curation: An ecological perspective. == College and Research Library News, 71. (2010) 4., p. 194-196. ISSN 0099-0086
 - SEMELER, Alexandre Ribas – PINTO, Adilson Luiz – ROZADOS, Helen Beatriz Frota. Data science in data librarianship: Core competencies of a data librarian. == Journal of Librarianship and Information Science, 51. (2019) 3., p. 771-780. ISSN 1961-0006
 - Amendments to the International Financial Reporting Standard for Small and Medium-sized Entities (IFRS for SMEs) : commentsto be received by 3 march 2014. - London : International Accounting Standards Board, 2015. – 76. p. – (International financial reporting standard). – ISBN 9781909704770
 - AL-RUITHE, Majid – BENKHELIFA, Elhadj – HAMEED, Khawar. A systematic literature review of data governance and cloud data governance. == Personal and Ubiquitous Computing, 23. (2019) 5-6., p. 839-859. ISSN 1617-4909
 - SCHUMACHER, Jaime – VANDECREEK, Drew. Intellectual capital at risk: data management practices and data loss by faculty members at five American universities. == International Journal of Digital Curation [online], 10. (2015) 2., p. 96-109. ISSN 1746-8256. Hozzáférhető: <https://doi.org/10.2218/ijdc.v10i2.321> [Megtekintve: 2020. június 15.]
 - FOSTER, Jonathan – MCLEOD, Julie – NOLIN, Jan – GREIFENEDER, Elke. Data work in context: value, risks, and governance. == Journal of the Association for Information Science and Technology, 69. (2018) 12., p. 1414-1427. ISSN 1532-2882
 - HRYNASZKIEWICZ, Iain: Publishers’ responsibilities in promoting data quality and reproducibility. == Good research practice in non-clinical pharmacology and biomedicine [online] Berlin, Heidelberg: Springer, 2019. p. 319-348. ISBN 978-3-030-33656-1 (Handbook of Experimental Pharmacology, ISSN 1865-0325, vol. 257.) Hozzáférhető: https://link.springer.com/chapter/10.1007/164_2019_290 [Megtekintve: 2020. június 15.]
 - KOLTAY Tibor. A kutatási adatok és a könyvtár. == Könyvtári Figyelő, Ú.f. 24. = 60. (2014) 2., p. 223-236. ISSN 0023-3773
 - HOLL András. Kutatási adatok kezelésének trendjei. == Tudományos és Műszaki Tájékoztatás, 62. (2015) 5., p. 177-180. ISSN 0041-3917
 - KOLTAY Tibor. Adatkönyvtáros vagy adattudós – Néhány gondolat a különbségekről és a hasonlóságokról. == Tudományos és Műszaki Tájékoztatás, 65. (2015) 10., p. 518-522. ISSN 0041-3917
 - KOLTAY Tibor. Új könyvtári feladatok az adatintenzív kutatás korában. == Könyvtári Figyelő, Ú.f. 29. = 65. (2019) 2., p. 211-217. ISSN 0023-3773
 - STEINEROVÁ, Jela. Information ecology – emerging framework for digital scholarship. == Libraries in the Digital Age (LIDA) Proceedings [online], 12. (2012). Hozzáférhető: <http://ozk.unizd.hr/proceedings/index.php/lida/article/view/66> [Megtekintve: 2020. június 15.]
 - KIM, Jeonghyun. Who is teaching data: meeting the demand for data professionals? == Journal of Education for Library and Information Science, 57. (2016) 2., p. 161-173. ISSN 0748-5786
 - HOBBS, Renee. Multiple visions of multimedia literacy: emerging areas of synthesis. == Micahel C. McKENNA [et al.] eds. International handbook of literacy and technology. Mahwah, N.J.: Lawrence Erlbaum, 2006. p. 15-26. ISBN 0805850880
 - SCHNEIDER René. Research Data Literacy. == KURBANOGLU, Serap, [et al.] eds. Worldwide commonalities and challenges in information literacy research and practice. Cham: Springer International, [2013]. p. 134-140. ISBN 9783319039190
 - ZELENY, Milan. Management support systems: towards integrated knowledge management. == Human Systems Management, 7. (1987) 1., p. 59-70. ISSN 0167-2533
 - ACKOFF, Russell L. From data to wisdom. == Journal of Applied Systems Analysis, 16. (1989) 1., p. 3-9. ISSN 0308-9541
 - WANG, Lin. Twinning data science with digital science in schools of library and digital science. == Journal of Documentation, 74. (2018) 6., p. 1243-1257. ISSN 0022-0418
 - ROWLEY, Jennifer The wisdom hierarchy: representations of the DIKW hierarchy. == Journal of Information Science, 33. (2007) 2., p. 163-180. ISSN 1165-5515
 - MAKANI, Joyline. Knowledge management, research data management, and university scholarship: towards an integrated institutional research data management support-system framework. == VINE, 45. (2015) 3., p. 344-359. ISSN 0305-5728

20. KOLTAY Tibor. Gondolatok a digitális bölcsészet, a könyvtár-tudomány és a könyvtárak kapcsolatrendszeréről. == *Digitális Bölcsészet* [online], 2. (2019.), p. 9-19. ISSN 2630-9696. Hozzáférhető: DOI: <https://doi.org/10.31400/dh-hun.2019.2> [Megtekintve: 2020. június 15.]
21. OWENS, Trevor. Defining data for humanists: text, artifact, digital or evidence? == *Journal of Digital Humanities* [online], 1. (2011) 1. ISSN 2165-6673. Hozzáférhető: <http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/> [Megtekintve: 2020. június 15.]
22. LATHAM, Kiersten. Museum object as document: using Buckland's information concepts to understand museum experiences. == *Journal of Documentation*, 68. (2012) 1., p. 45-71. ISSN 0022-0418
23. FRICKÉ, Martin. The knowledge pyramid: a critique of the DIKW hierarchy. == *Journal of Information Science*, 35. (2009) 2., p. 131-142. ISSN 1165-5515
24. BUCKLAND, Michael. Information as thing. == *Journal of the American Society for Information Science*, 42. (1991) 5., p. 351-360. ISSN 0002-8231; BUCKLAND, Michael. What kind of science can information science be? == *Journal of the American Society for Information Science and Technology*, 63. (2012) 1., p. 1-7. ISSN 1532-2882
25. ŠPIRANEC, Sonja – KOS, Denis – GEORGE, Michael. Searching for critical dimensions in data literacy. == *Information Research* [online], 24. (2019) 4. paper colis1922. ISSN 1368-1613. Hozzáférhető: <http://InformationR.net/ir/24-4/colis/colis1922.html> [Megtekintve: 2020. június 15.]
26. BAWDEN, David – ROBINSON, Lyn. The dark side of information: overload, anxiety and other paradoxes and pathologies. == *Journal of Information Science*, 35. (2009) 2., p. 180-191. ISSN 1165-5515
27. CARLSON, Jake R. – FOSMIRE, Michael – MILLER, Chris – SAPP NELSON, Megan R. Determining data information literacy needs: a study of students and research faculty. == *Libraries faculty and staff scholarship and research* [online]. 2011. Paper 23. Hozzáférhető: http://docs.lib.purdue.edu/lib_fsdocs/23 [Megtekintve: 2020. június 15.]
28. LE DEUFF, Olivier – PERRET, Arthur. 'Hyperdocumentation: origin and evolution of a concept.' == *Journal of Documentation*, 75. (2019) 6., p. 1463-1474. ISSN 0022-0418
29. PRYOR, Graham. ed. *Managing research data*. London: Facet Publishing, 2012. XI, 239 p. ISBN 9781856047562
30. GOLUB, Koraljka – HANSSON, Joacim. (Big) Data in library and digital science: a brief overview of some important problem areas. == *Journal of Universal Computer Science*, 23. (2017) 1., p. 1098-1108. ISSN 0948-695X
31. WANG, Richard Y. – STRONG, Diane M. Beyond accuracy: what data quality means to data consumers == *Journal of Management Information Systems*, 12. (1996) 4. p. 5-33. ISSN 0742-1222
32. ALTMAN, Micah [et al.]. Mitigating threats to data quality throughout the curation lifecycle. == *Curating for quality: ensuring data quality to enable new science*. Arlington County, VA: National Science Foundation, 2012. p. 1-119.
33. SPOSITO, Frank Andreas. What do data curators care about? [online]: Data quality, user trust, and the data reuse plan. IFLA 2017 Satellite Meeting, IFLA WLIC 2017 Wrocław. [Wrocław] : IFLA, cp. 2017. 7 p. Hozzáférhető: <http://library.ifla.org/1797/1/S06-2017-sposito-en.pdf> [Megtekintve: 2020. június 15.]
34. STRONG, Diane M. – LEE, Yang W. – WANG, Richard Y. Data quality in context. == *Communications of the ACM*, 40. (1997) 5., p. 103-110. ISSN 0001-0782
35. LARANJEIRO, Nuno – SOYDEMIR, Seyma Nur – BERNARDINO, Jorge. A survey on data quality: classifying poor data. == 2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC) [online]. [S. l.]: IEEE, 2015. p. 179-188. ISBN 9781467393775 Hozzáférhető regisztrációval: <https://ieeexplore.ieee.org/document/7371861> [Megtekintve: 2020. június 15.]
36. DARAIO, Cinzia [et al.]. The advantages of an ontology-based data management approach: openness, interoperability and data quality. == *Scientometrics*, 108. (2016) 1., p. 441-455. ISSN 0138-9130
37. SMITH, MacKenzie. Communicating with data: new roles for scientists, publishers and librarians. == *Learned Publishing*, 24. (2011) 3., p. 203-205. ISSN 1741-4857
38. WOLSKI, Malcolm – HOWARD, Louise – RICHARDSON, Joanna. A trust framework for online research data services. == *Publications* [online] MDPI. Vol. 5. Issue 2. (2017) no. 14. <https://www.mdpi.com/2304-6775/5/2/14/htm> [Megtekintve: 2020. június 15.]
39. GIARLO, Michael J. Academic libraries as data quality hubs. == *Journal of Librarianship & Scholarly Communication* [online], 1. (2013) 3., eP1059. ISSN 2162-3309. Hozzáférhető: <https://jlscc-pub.org/articles/abstract/10.7710/2162-3309.1059/> [Megtekintve: 2020. június 15.]
40. WOLSKI – HOWARD – RICHARDSON, i.m.
41. GIARLO, i.m.
42. YOON, Ayoung. End users' trust in data repositories: Definition and influences on trust development. == *Archival Science*, 14. (2014) 1., p. 17-34. ISSN 1389-0166
43. GIARLO, i.m.

44. MILLER, Holmes. The multiple dimensions of information quality. == Information Systems Management, 13. (1996) 2., p. 79-82. ISSN 1058-0530
45. GIARLO, i.m.
46. BORGMAN, Christine L. Scholarship in the digital age: Information, infrastructure, and the Internet. Cambridge, MA, London: MIT Press, 2010. XXIV, 336 p. ISBN 9780262514903
47. FANIEL, Ixchel M. – KRIESBERG, Adam – YAKEL, Elizabeth. Social scientists' satisfaction with data reuse. = Journal of the Association for Information Science and Technology, 67. (2016) 6., p. 1404-1416. ISSN 1532-2882
48. ZILINSKI, Lisa D. – NELSON, Megan S. Thinking critically about data consumption: creating the data credibility checklist. == Proceedings of the American Society for Information Science and Technology, 51. (2014) 1., p. 1-4. ISSN 1550-8390
49. YOON, Ayoun – LEE, Yoo Young. Factors of trust in data reuse. == Online Information Review, 43. (2019) 7., p. 1245-1262. ISSN 1468-4527
50. COLEPICCOLO, Eliane. Information reliability for academic research: review and recommendations. == New Library World, 116. (2015) 11/12., p. 646-660. ISSN 0307-4803
51. DODDS, Leigh – WELLS, Peter. Data infrastructure. == T. DAVIES [et al.] eds. The state of open data: Histories and horizons. Cape Town: African Minds, Ottawa: International Development Research Centre, cop. 2019. p. 260-273. ISBN 978-1-928331-95-7
52. CHAVAN, Vishwas – PENEV, Lyubomir. The data paper: a mechanism to incentivize data publishing in biodiversity science. = BMC Bioinformatics [online], 12. (2011) 15. ISSN 1471-2105. Hozzáférhető: <https://link.springer.com/article/10.1186/1471-2105-12-S15-S2> [Megtekintve: 2020. június 15.]
53. CANDELA, Leonardo – CASTELLI, Donatella – MANGHI, Paolo – TANI, Alice. Data journals: a survey. == Journal of the Association for Information Science and Technology, 66. (2015) 9., p. 1747-1762. ISSN 1532-2882
54. Data management and use [elektronikus dok.]: governance in the 21st century: a joint report by the British Academy and the Royal Society. London: British Academy, The Royal Society, 2017. 95 p. Hozzáférhető: <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf>. [Megtekintve: 2020. június 15.]
55. ISO/IEC 25012:2008 Software engineering – Software product quality requirements and evaluation (SquaRE) – Data quality model [elektronikus dok.]. Geneva: ISO, 2008. Hozzáférhető: <https://iso25000.com/index.php/en/iso-25000-standards/iso-25012> [Megtekintve: 2020. június 15.]
56. STRONG, Diane M. – LEE, Yang W. – WANG, Richard Y. Data quality in context. == Communications of the ACM, 40. (1997) 5., p. 103-110. ISSN 0001-0782
57. ISO/IEC. i.m.
58. VAN LOON, James E. – AKERS, Katherine G. – HUDSON, Cole – SARKOZY, Alexandra. Quality evaluation of data management plans at a research university. == IFLA Journal, 43. (2017) 1., p. 98-104. ISSN 0340-0352
59. TENOPIR, Carol [et al.]. Research data services in European academic research libraries. == LIBER Quarterly [online], 27. (2017) 1., p. 23-44. ISSN 2213-056X. Hozzáférhető: <https://www.liberquarterly.eu/articles/10.18352/lq.10180/> [Megtekintve: 2020. június 15.]
60. LE DEUFF – PERRET, i.m.
61. ABRAHAM, Rene – SCHNEIDER, Johannes – VOM BROCKE, Jan: Data governance: A conceptual framework, structured review, and research agenda. == International Journal of Information Management, 49. (2019) p. 424-438. ISSN 0268-4012
62. JANSSEN, Marijn – VAN DER VOORT, Haiko – WAHYUDI, Agung. Factors influencing big data decision-making quality. == Journal of Business Research, 70. (2017) p. 338-345. ISSN 0148-2963
63. DEMCHENKO, Yuri – Membrey, Peter – Grosso, Paola – de Laat, Cees. Addressing big data issues in a scientific data infrastructure. == 2013 International Conference on Collaboration Technologies and Systems (CTS). [S. I.]: IEEE, 2013. p. 48-55. ISBN 978-1-4673-6403-4
64. VIRKUS, Sirje – GAROUFALLOU, Emmanouel. Data science from a library and information science perspective. == Data Technologies and Applications, 53. (2019) 4., p. 422-441. ISSN 2514-9288
- KOLTAY Tibor. Egy „örökzöld téma”, az információs túlterhelés. == Információs Társadalom, 17. (2017) 3., p. 39-54. ISSN 1587-8694
65. BORGMAN, i.m.
66. FRICKÉ, Martin. Big data and its epistemology. == Journal of the Association for Information Science and Technology, 66. (2015) 4., p. 651-661. ISSN 1532-2882
67. HALME, Panu – KOMONEN, Atte – HUITU, Otso. Solutions to replace quantity with quality in science. == Trends in Ecology & Evolution, 27. (2012) 11., p. 586. ISSN 0169-5347

Beérkezett: 2020. május 11.

Megújult a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár

Hivatalosan is átadták a Debreceni Egyetem megújult Egyetemi és Nemzeti Könyvtárát.

Az összesen csaknem 400 millió forintos ráfordítással megvalósult beruházásnak köszönhetően 21. századi megújult infrastruktúra várja a látogatókat.

Az Élettudományi Épület és Könyvtár hallgatói terei 59,4 millió, míg a Főépület Egyetemi és Nemzeti Könyvtár terei 170,7 millió forintból újultak meg.

„A Debreceni Egyetem felsőoktatási infrastruktúra fejlesztése a gyakorlati és szakmai képzés megújítása érdekében” EFOP-4.2.1-16-2017-00015 projekt keretein belül az Élettudományi Könyvtár fejlesztésére informatikai- és szemléltető eszközöket is vásároltak 61 millió forint értékben. Az építési beruházásokat kiegészítve, a szükséges bútorok beszerzését az egyetem saját forrásból 81,7 millió forint összegben támogatta. Emellett egy modern, az automatikus azonosításhoz és adatközléshez használt Radio Frequency Identification (RFID) technológia beszerzésével egy további fejlesztés is megvalósult 16,7 millió forint saját forrásból. Az ünnepélyes szalagátvágást követően a program ünnepi üléssel folytatódott a Főépület Aulájában.”

Forrás:

<https://hirek.unideb.hu/hu/node/23677>

ahol a hír tovább olvasható.

Közzétette: Haraszti Katalin, IFLA-HUN moderátor, Katalist,

2020. szeptember 9.