

Az adatok „szép új világa”

Koltay Tibor

Bevezetés

Írásom címét Aldous Huxley könyvének címéből (Szép új világ, 1982) vettem. Huxley egy olyan világról ír, amelyben kiveszett az emberek életéből a kulturális sokszínűség, hiányoznak a művészetek, az irodalom, a tudomány, a vallás és a filozófia.

Szerencsére ez csak a szerző által elképzelt negatív utópia, azonban Kitchin (2014) arra figyelmeztet bennünket, hogy az adatok, és különösen az úgynevezett nagy adatok (Big Data) szerepét és fontosságát sokan eltúlozzák. Ugyanakkor tudnunk kell, hogy a digitális adatok áramlása gyors, így elárasztanak bennünket, és kezdik betölteni azt a szerepet, amelyet korábban a média játszott (Smith, 2013). Nagy adatnak azokat az adatállományokat nevezzük, amelyeknek a feldolgozása meghaladja a hagyományos számítógépes rendszerek és szoftverek képességeit, mivel gyűjtésük és feldolgozásuk több időt vesz igénybe, mint amit tolerálni tudunk (Gordon-Murnane, 2012). Ezeknek az adatoknak a mérete azonban mozgó cél, ezért a mai értelemben vett nagy adatok esetében arról is szó van, hogy a számítástechnikai eszközök hatékonyságának eredményeként lehetővé vált összekapcsolásuk, elemzésük és összevetésük más adatokkal (Boyd & Crawford, 2012). Mindez lehetővé teszi, hogy emberek százainak, sőt ezreinek elképzelései, véleménye, gondolatai és érzései váljanak tanulmányozhatóvá, elsősorban a közösségi médiából vett adatok felhasználásával (Manovich, 2012).

Sokan gondolhatják azt, hogy az adatok pusztán mennyisége azonos az igazsággal, tehát bármilyen elmélettől és tudományos bizonyítéktól függetlenül képesek arra, hogy önmagukért beszéljenek. Ez a nézet különösen üzleti körökben népszerű, és az adattudomány (amelyről a következőkben még szólni fogunk) egyes területein is gyökeret vert. Az adatok azonban főként azért nem hasznosulnak kellő mértékben, mert kevés a hozzájuk kötődő kontextuális információ (Gebre & Morales, 2020).

Mindeközben az adatvezérelt kutatás ragaszkodik a tudományosság megőrzéséhez. Ez még akkor is így van, ha a kutatók hipotéziseiket nemcsak az elméletből, hanem adatokból is képesek és hajlandók generálni. Nem szabad azonban elfejtenünk azt, hogy adatokat egyes személyek, csoportok és szervezetek az állampolgárok beleegyezése nélkül gyűjtenek. Márpedig számos adat felhasználható titkos megfigyelésekre és a magánélet megsértésére is (Llebot & Rempel, 2021).

Ahogy például Z. Karvalics László figyelmeztet rá, a „bejövő jelek tömegtermelése olyan nagyságrendekbe emelkedett, ahol számítási teljesítménnyel (feldolgozási, átviteli és tárolási kapacitásokkal és mesterséges intelligencia-megoldásokkal) még sikerül követni a mennyiségi kihívást, de valódi elemző és megismerő erőt már nem lehet mellé rendelni”. Hozzáteszi, hogy ehhez még hiányzik a „humán intelligencia”, amely jelentést és kontextust volna képes generálni (Z. Karvalics, 2008, p. 352). Ez a probléma ma is fennáll.

A tudományos (kutatási) adatok bármely megfigyelésen, kísérleten vagy valamely hipotézis megvizsgálásán alapuló, szisztematikus kutatás végtermékei. Más adatokhoz hasonlóan tehát lehetnek numerikusak vagy szövegalapúak, állhatnak hangokból, álló- és mozgóképekből. Forrásaik között ott vannak a természettudományos megfigyelések, kísérletek és modellek. A társadalomtudományok kutatói maguk is létrehozhatnak adatokat, vagy nyilvánosan hozzáférhető iratokból is kinyerhetjük őket. Míg a hagyományos humán tudományi oktatás és kutatás keveset foglalkozik az adatokkal, nagy figyelmet kapnak a digitális bölcsészet területén, amely modern információtechnológiai eszközöket használ az irodalom(tudomány), a nyelvek (nyelvtudomány), a történelem(-tudomány) és a filozófia területén folyó kutatás és az oktatás támogatására.

Amellett, hogy egyre több adat fog létrejönni digitális formában, az is várható, hogy az analóg formájú adatok nagy tömegét fogják digitalizálni (Robinson, 2016). Ennek folyamányaként egyre inkább meghatározó jelentősége van az adatokra alapozott gondolkodásnak, amely összekapcsolódik annak képességével, hogy az adatokat absztrakt fogalmakká alakítsuk át (Davies, Fidler & Gorbis, 2011).

Azt viszont csak néhány éve – főként a tudományos kutatást támogató szervezetek nyomására – tapasztaljuk, hogy ezeknek az adatoknak stratégiai jelentőséget tulajdonít számos intézmény. A tudományos kutatás ugyanis fokozott mértékben támaszkodik a nagy kapacitású számítástechnikai eszközök használatára, ami többek között magával hozta, hogy nemcsak a publikációkhoz való nyílt hozzáférés (Open Access) követelménye jelent meg, hanem a nyílt kutatási adatokat (Open Data) magában foglaló tudomány (Open Science) iránti igény is. A publikációkhoz való nyílt hozzáférés minden ellentmondásával együtt viszonylag széles körben elterjedt már. A kutatási adatok esetében azonban jóval nehezebb ezt megvalósítani, mivel szükségességét nem minden országban ismerték fel a kutatók és intézményeik, ezért elismertsége és fejlettsége is igen változó (Holl, 2018). Ennek ellenére az utóbbi években egyre gyakrabban olvashatunk ilyen irányú új fejleményekről. Ezek részben összefüggésben vannak az információs műveltség és az adatműveltség széles körű elterjesztését célzó erőfeszítések egyre nagyobb számával.

Információs műveltség és adatumveltség

Az adatumveltség gondolatának fontos előzménye, hogy megjelentek az információs műveltségről alkotott elképzelések (ALA, 1989). Egy korszerű, új meghatározása szerint az adatumveltség egy sor készséget és képességet foglal magában, hangsúlyozva, hogy ezek mindenki számára szükségesek az információval kapcsolatos feladatok elvégzéséhez, például ahhoz, hogy képesek legyünk felfedezni, hozzáférni, értelmezni, elemezni, kezelni, létrehozni, kommunikálni, tárolni és megosztani információkat. Fontos eleme, hogy kiemeli az információ felhasználásával kapcsolatos kritikai gondolkodás és tudatosság, valamint az etikai szempontok mozgósításának fontosságát. Ez a meghatározás egyúttal kimondja, hogy az információs műveltség az információ minden formájára vonatkozik, tehát nemcsak a nyomtatott, hanem a digitális tartalmakra, adatokra, képekre és a szóbeli közlésekre is érvényes (CILIP, 2018).

Calzada Prado és Marzal (2013) nemcsak az információs műveltség és az adatumveltség közötti folytonosságra mutatnak rá, ezért aláhúzzák, hogy oktatásuknak már az iskolában el kell kezdődnie, majd a felsőoktatásban kell tökéletesednie és specializálódnia, továbbá az egész életen át tartó tanulás részévé kell válnia. Maybee és Zilinski (2015) arra is rámutat, hogy az adatumveltséggel kapcsolatos megközelítések jellemzően az információs műveltségből táplálkoznak.

Összeségében az adatumveltség egy olyan készségkészlet és tudásbázis, amely képessé tesz bennünket arra, hogy az adatokat információvá és cselekvőképes tudássá alakítsunk át, lehetővé téve számunkra az adatokhoz való hozzáférést, valamint azok értelmezését, értékelését, kezelését és etikus felhasználását (Koltay, 2017).

Mielőtt azonban jellemzőire rátérnénk, említsük meg, hogy a szakirodalom 1975 és 2018 közötti alakulását vizsgálva Onyancha (2020) azt találta, hogy a cikkekben leggyakrabban használt kifejezés az *információs műveltség* (information literacy) volt. Az adatumveltség csupán a 22. helyet foglalta el az írástudások 40 leggyakoribb előfordulása között. Nyilvánvalóan ez a különbség abban az egyszerű tényben gyökerezik, hogy az adatumveltség fogalma sokkal később, a 2010-es években kezdett lendületet venni, amikor a számítógépek képességei elérték az a szintet, amivel nagy mennyiségű adat tárolására és nagy sávszélességű hálózatokon történő továbbítására váltak képessé.

Az adatok megnövekedett fontossága ugyanakkor nemcsak a kutatási adatokat érinti, ezért az adatumveltség (data literacy) iránt számos területen van igény, tehát nemcsak az oktatás, hanem a munkáltatók, a kormányzatok, a nonprofit szervezetek és a kutatók is szükségét érzik, ezért az adatumveltség szakterületeken átívelő, interdiszciplináris és transzverzális fogalommá vált (Corrall, 2019).

Az adatok életciklusának számos fázisa van. Az első fázis az adatszerzés, amelyet az adatok tisztítása, majd az adatmodellezés követ. Az életciklus része az optimalizálás, az elemzés, a vizualizáció, az adatok értékelése, megosztása, archiválása és törlése. Ennek megfelelően az adatműveltség operatív definícióinak középpontjában az elfogadhatóan működő műveltség áll, míg a kritikai adatműveltség egy sor magasabb rendű probléma felvetését is magával hozza, beleértve a kérdésfeltevés és a reflexió képességét, amely jóval több annál, hogy egyszerűen dekódoljuk a szimbólumokat, ahogyan azt egy gép teszi. Az adatokra alkalmazva azonban nagyon nehéz elképzelni, hogy a kritikai dimenzió nélkül mi lenne a működési műveltség megfelelője (Pangrazio & Sef-ton-Green, 2019).

Az adatműveltség természete

Az adatműveltség magában foglalja a tudást, valamint az annak alkalmazására való hajlandóságot és képességet. Megköveteli, hogy szakszerűen használjuk az adatokat, ami olyan kompetenciákat igényel, amelyek lehetővé teszik az adatok, az információk és a vélemények megkülönböztetését.

Az a gondolat, hogy a kritikai szemlélet az adatműveltségre is kiterjeszhető, már korán megjelent. Carlson, Fosmire, Miller és Sapp Nelson (2011) például kiemelték fontosságát, különös tekintettel arra, hogy a kritikai jelleg az információs műveltségnek is sajátja. Annak megértését foglalja ugyanis magába, hogy mit jelentenek az adatok, ideértve azt is, hogy miként kell megfelelően olvasni a grafikonokat és diagramokat, helyes következtetéseket levonni az adatokból, továbbá felismerni, ha az adatokat félrevezető vagy nem megfelelő módon használják.

Az adatműveltség oktatásának tehát nemcsak a kutatókra és az egyetemi hallgatókra kell kiterjednie. A közoktatásban és a felsőoktatásban tanulókat is fel kell készítenünk az adatokkal való munkára és arra is, hogy életük számtalan területén szükségük van állampolgári tudatosságra és az adatműveltségre (Li, Jiao, Zhang, & Xu, 2019).

A közoktatásban tanuló diákok hatékony felkészítésének feltétele, hogy a szükséges attitűdök és készségek elsajátíttatása figyelembe vegye az adatokkal való munka emberi dimenzióit, ideértve a következők jelentőségét:

- a diákoknak az adatokkal, a méréssel és az adatgyűjtés kontextusával kapcsolatos személyes és közvetlen tapasztalatainak beemelése az oktatásba,
- azoknak a kulturális és társadalmi-technikai infrastruktúráknak és értékeknek a bemutatása, amelyek az adatgyűjtés és -felhasználás során érvé-

nyesülnek (beleértve, de nem kizárólagosan a különböző iskolai, kulturális és diszciplináris közösségekhez kapcsolódó rutinokat, technológiákat és normákat),

- annak figyelembevétele, hogy miként befolyásolják az oktatás céljait és módszereit egyes, tartósan jelen lévő politikai és társadalmi narratívák (Wilkerson & Polman, 2020).

Az adatumveltség területén azonban továbbra is kettősséget látunk. Vannak ugyanis kompetenciaorientált perspektívái, amelyek a hangsúlyt főleg a technikai és fogalmi készségek fejlesztésére helyezik, hogy lehetővé tegyék a főként mennyiségi adatok kezelését, beleértve azok minőségének, hitelességének és relevanciájának meghatározását, elemzését és értelmezését. Létezik azonban egy felhatalmazásorientált perspektívája is, amely része a kompetenciaorientált szemléletnek, viszont túlmutat a technikai készségeken (Gebre, 2018).

A fentebb említett felhatalmazás alapja az, hogy képesnek kell lennünk az adatok és a valóság igazságtartalma közötti eltérések felismerésére. Ennek tudatában érthetjük csak meg az adatvezérelt elemzés korlátait, továbbá leszünk képesek információt keresni egy-egy adott narratíva megalkotásához, kiegészítéséhez vagy egy adathalmazból levezetett állítások ellenőrzéséhez (Davies, Fidler & Gorbis, 2020).

Sok hasonlóságot fedezhetünk fel az információs műveltség, az adatumveltség és a sokak által ismert DigComp állampolgári digitáliskompetencia-keretben lefektetett ismeretek, készségek és attitűdök között is. Nem véletlen, hogy ez a keretrendszer 2.1-es változatában már információs és adatumveltségről beszél a korábbi információs kompetenciaterület megnevezés helyett (Carretero, Vuorikari & Punie, 2017).

Látunk kell viszont, hogy az adatumveltség megítélése egyelőre ellentmondásos, ugyanis – amíg több egyetemen a mesterszakok és doktori képzések szokásos tantárgyává kezd válni – a kutatók többsége csak a munkahelyén szerez adatumveltséget (Ince, Hoadley & Kirschner, 2018).

Aki adatumveltségét hatékonyan kívánja gyakorolni, annak a következő kompetenciákkal kell rendelkeznie:

- felismeri az adatok előállítását és újrafelhasználását magába foglaló életciklusának kontextusait;
- felismeri a nyers adatok fontosságát, típusait és formátumait;
- meghatározza, hogy mikor van szüksége adatokra;
- képes hozzáférni az információigényeinek megfelelő adatok forrásához;
- kritikai szemlélettel fér hozzá az adatokhoz és forrásaikhoz;
- meg tudja határozni és tudja használni a megfelelő irodalomkutatási módszereket;

- képes kezelni és elemezni az adatokat;
- tudja, hogy miként lehet adatokat válogatni, továbbá azokat más forrásokkal és előzetes ismeretekkel kombinálni;
- képes az adatokat tükröző kvantitatív információt (specifikus adatokat, táblázatokat stb.) prezentálni;
- etikusan használja az adatokat;
- képes az eredményeket a tanulás, a döntéshozatal és a problémamegoldás céljaira felhasználni;
- önmagát ellenőrizve képes ezeket a folyamatokat tervezni, szervezni (Calzada Prado & Marzal, 2013).

Mindeközben fontos tudnunk, hogy az adatok egy adott valóság pillanatfelvétel jellegű reprezentációi, ezért hajlamosak vagyunk könnyebben elhinni az adatokból levont következtetéseket (Fontichiaro & Oehrli, 2016). Bár megpróbálhatjuk döntéseinkből kizárni az elfogultságot, az adatok létrehozása során hozott döntéseink befolyásolják a végső adathalmazt, mivel magunk választjuk ki, hogy mit számolunk vagy mérünk (Schild, 2010).

A Grillenberger és Romeike (2018) által meghatározott adatumveltségi tudásterületek a következők:

- Adatkezelés (az adatok menedzselése), ideértve az adatok szervezését, manipulációját, az adatkonverziót, metaadatok létrehozását és használatát, az adatok kurátori gondozását, újrafelhasználását és az adatbiztonság kérdéseit.
- Adatértékelés, ideértve az elemzést és értelmezést, az adatok felhasználásával kapcsolatos problémák azonosítását és az adatok vizualizálását.
- Az adatok alkalmazása, amihez hozzátartozik a kritikai gondolkodás, az adatkultúra, az adatokkal kapcsolatos etikai kérdések vizsgálata, az adatokra történő hivatkozás, az adatmegosztás és az adatokon alapuló döntések értékelése.

Mivel az adatokat gépek és emberek egyaránt létrehozhatják, az adatok ereendőően nem objektívek, ezért az adatok értékelése kapcsán többek között a következő kérdéseket kell feltennünk:

- Ki vagy mi hozta létre az adatokat vagy irányította létrehozásukat?
- Hogyan jöttek létre ezek az adatok?
- Milyen emberi vagy gépi alapú elfogultságokat tükröznek?

Ezek a kérdések azért is fontosak, mert az adatumveltség a már említett módon közeli rokonságban áll az információs műveltséggel. A kettő közötti különbség forrása viszont, hogy az információs műveltség az emberek által

létrehozott tudásformákat foglalja magában, az adatok előállításához nem feltétlenül van szükség közvetlen emberi inputra (Abner, 2020).

Az UNESCO által támogatott kutatásoknak köszönhetően beszélhetünk média- és információs műveltségről (Media and Information Literacy, MIL) is. Ennek mindkét alkotóeleme pedagógiai természetű, és célja az, hogy az információk és médiatartalmak használatának, megértésének és előállításának készségeit fejlessze (Lee & So, 2014).

Az akadémiai írástudás és a statisztikai írástudás

Röviden szólnunk kell az akadémiai és a statisztikai írástudásról. Az adatumveltség ugyanis mindkettőhöz kapcsolódik.

Az *akadémiai írástudás* elnevezés első fele a felsőoktatásban betöltött szerepére utal. Esetében az írástudás szó használata különösen indokolt, mivel – bár közel áll az információs műveltséghez – legismertebb jellemzője az írásbeli kommunikáció középpontba állítása. Ettől függetlenül, az akadémiai írástudás oktatásának más, esetenként új információs módokra és formátumokra is figyelnie kell (Norgaard & Sinkinson, 2016). Alapjában véve tehát az akadémiai írástudás azoknak a készségeknek, beállítódásoknak és ismereteknek az összessége, amelyeket egy egyetemi hallgatónak el kell sajátítania annak érdekében, hogy egy adott tudományos közösség elvárásainak megfelelő szövegeket tudjon olvasni, értelmezni és létrehozni (Koltay, 2018). Tudományos írások létrehozása során általában arra törekszünk, hogy a felhasznált forrásokból bizonyítékokon alapuló, helyes és pontos következtetéseket vonjunk le. Állításaink azonban nemcsak a szövegelemzésben gyökerezhetnek, hanem adatokból levont következtetésekből is megtestesülhetnek (Shome, 2019). Általánosabb összefüggésben viszont azt mondhatjuk, hogy az akadémiai írástudás nem korlátozódhat arra, hogy a hallgatónak segítséget nyújtson az egyetemen való túléléshez jól megformált és szabatosan megfogalmazott írásbeli dokumentumok előállításával, hanem adatok elemzésére is ki kell terjednie.

A *statisztikai írástudást* viszont úgy határozhatjuk meg, mint a statisztikák olvasásának és értelmezésének képességét grafikonokban, táblázatokban, kimutatásokban, felmérésekben és tanulmányokban. Ennek megfelelően nem véletlenül említik olyan gyakran az adatumveltséggel összefüggésben, mivel mindkettő hasonló problémák megoldását keresi, méghozzá sok tekintetben hasonló jellegű és szintű megközelítésekkel (Schild, 2010).

A digitális bölcsészeti és az adattudományi háttér

Az adatvezérelt világban és ilyen módon az adatműveltség területén is fontos helyet foglal el két olyan diszciplína, amely önmagában is újdonság. Ezek a digitális bölcsészet és az adattudomány.

A digitális bölcsészet néhány sajátossága

A digitális bölcsészet egyaránt felvonultat humán tudományi és informatikai ismereteket, tehát természete inter- és multidiszciplináris jellegű, továbbá együttműködésen alapul. Nemcsak igényli a kutatási adatok kezelését, hanem a digitális anyagok gondozása és a vele közeli rokonságban álló – szintén kurátori – adatgondozás, valamint a kutatási adatok menedzselése is nélkülözhetetlen ezen a területen (Poole, 2017).

A digitális bölcsészet kapcsán figyelmet érdemel két, a hétköznapitól eltérő olvasási típus. Amikor ugyanis a szövegek összetevőinek komplex kapcsolataira és többértelműségeinek részletes elemzésére koncentrálnunk, a megszokottnál közelebbről vizsgáljuk az adott szöveget. Ez a *close reading*, amely a hétköznapi olvasásnál lassabb és alaposabb folyamatot jelent. Ez az olvasás lehetővé teszi, hogy magára a szövegre visszavezethető, ellenőrizhető állításokat tegyünk irodalmi művekről úgy, hogy feltárjuk a többletjelentéseket és kétértelműségeket, ami aztán segíti értelmezésüket, továbbá megmutatja kontextusait is. Ezért is használják az irodalomtanításban.

A digitális eszközök segítségével viszont nagyszámú szöveget dolgozhatunk fel egyszerre, statisztikai módszerekkel mintázatokat keresve, majd vizuális formában megjelenítve azokat. Ez a módszertan a *distant reading*, amely egyedi művek helyett általánosabb tendenciákat vizsgál, hiszen a számítógépek képesek nagy mennyiségű adat együttes kezelésére. Ahogyan az ilyen esetekben történni szokott, e két megközelítés kiegészíti egymást (Szemes, 2018).

Adattudomány és adattudósok

Az adattudomány célja aktuális jelenségek elemzése és megértése adatok segítségével. Ez azt jelenti, hogy az adatok jellemzőit vagy rejtett struktúráit próbálja feltárni összetett természeti, emberi és társadalmi jelenségek modellezésével, a hagyományos szemlélettől eltérő nézőpontokból történő felhasználásával. Ez többdimenziós szemléletet igényel, dinamikus és rugalmas gondolkodásmódot tételez fel (Hayashi, 1998).

Képességeik és munkájuk alapján az adattudósok besorolhatók a fejlesztők, kutatók, sőt az üzletemberek közé is (Davenport & Patil, 2012). Ebben a tekintetben éppen a humán tudományokhoz kötődő digitális bölcsészet ellentéte az adattudomány, bár ritkán szoktuk őket egymással szembeállítani.

Az adattudósoknak megfelelő ismeretekkel kell rendelkezniük az elemzési módszerek, a statisztika és a programozás területén, viszont nem kell programozónak vagy statisztikusnak lenniük. Elvárható tőlük, hogy felismerjék az adatgyűjtési folyamatok fontosságát, továbbá jártasnak kell lenniük az adatok elemzése, értelmezése és vizualizálása terén annak érdekében, hogy az adatokból jelentést tudjanak kinyerni, és adattermékeket hozzanak létre, nemcsak az adattudomány, hanem a matematika, a statisztika és az adatelemzés eszközeinek segítségével. Mindez rendszerszemléletű gondolkodást igényel, amelynek kreatív megközelítéssel kell párosulnia a felmerülő problémák megoldásához (van der Aalst, 2014). Mindezek mellett az adattudósoknak ismerniük kell az elemzési munkájukhoz kapcsolódó szoftvereket, legyen szó statisztikai vagy mesterségesintelligencia-modellek készítéséről, vizuális elemzések generálásáról, döntéshozatali szabályok meghatározásáról vagy szimulációk létrehozásáról (Davenport & Patil, 2012).

Az adattudomány a legkevésbé sem egységes terület. Nem tudhatjuk viszont, hogy ez előnyére vagy hátrányává válik-e majd a jövőben. Az adattudósok inkább multidiszciplináris csapatokban tudnak hatékonyan dolgozni. Ezekben ott vannak az adatelemzők, akik sokkal inkább „valódi” mérnökök, mintsem üzleti szakemberek. A statisztikusok hídát képeznek az adatelemzők és az informatikusok között azáltal, hogy formát adnak a szakterületi szakértőktől érkező kérdéseknek és ötleteknek, finomítják őket, releváns adatokat kérve az adatelemzőktől, továbbá irányítják az adatelemzéseket végző informatikusokat. Az informatikusoknak jártasnak kell lenniük a szoftvereszközök és technológiák, valamint a releváns programnyelvek használatában. Egy-egy ilyen csapat kommunikátorának a szervezeti változásokért kell felelnie azáltal, hogy hatékonyan kommunikál az érintett döntéshozókkal, többek között a releváns problémákról és lehetőségekről, valamint az esetleges eredmények kommunikálásáról. Ezeknek a csapatoknak olyan vezetőkre van szükségük, akik tisztában vannak kollégáik szerepével, és kezelni tudják az erőforrásokkal való gazdálkodást és a feladatok végrehajtását (Baškarada & Koronios, 2017).

Az adattudomány természete megértésének és megértetésének fontos részét képezik azok a gazdagon rétegzett kapcsolatrendszerek, amelyek nemcsak az adatokról és adatokkal való tanulás eszközei, hanem az adatok társadalmi szövegként való navigálását is lehetővé teszik. Ezért is állítja Wilkerson és Polman (2020), hogy ideális esetben az adattudományi oktatás nemcsak jól fizető állásokra készíti fel a hallgatókat, hanem hozzájárulhat a tudományok fejlődéséhez,

és új eszközöket biztosíthat a társadalom számára. Ennek egyik fontos szelete az, hogy megpróbáljuk megtalálni azt a különbséget, amely az adattudomány oktatását megkülönbözteti a matematika, az informatika vagy a statisztika tanításától.

Mindenesetre a világ több mint 200 egyetemén 2017-ben 530 oktatási program szerepelt az adattudomány, az analitika és a kapcsolódó területek listáján. Ezek túlnyomó többsége mesterképzés volt. A PhD-programok száma viszonylag alacsony volt, az alapképzésben részt vevők száma azonban gyorsan növekedett.

Az alapképzésben oktatóndó kulcskompetenciák egyik lehetséges listája a következőket tartalmazza:

- informatikai gondolkodás (computational thinking) és statisztikai gondolkodás,
- matematikai alapok,
- modellépítés és -értékelés,
- algoritmusok és szoftveralapok,
- adatok kurátori gondozása (data curation),
- tudásátadás és tudáskommunikáció (De Veaux et al., 2017).

Ahogy az már a korábbiakból is leszűrhető volt, az adattudomány természetét nagyban meghatározza interdiszciplináris jellege. Ezen a szakterületen folyamatosan új eszközöket és statisztikai technikákat fejlesztenek ki, amelyek több tudományág módszereit is magukban foglalják, majd megújítva az adatok vizualizálása, tárolása, rendszerezése, feldolgozása és értelmezése céljából használják fel őket. Az adattudományi oktatás célja pedig az, hogy a hallgatók egyaránt megtanulják, miként kell adatok felhasználásával és az adatokról gondolkodniuk. Vannak, akik arról sem feledkeznek meg, hogy ennek a gondolkodásnak része az is, hogy a már említett állampolgári és információs műveltség birtokában el tudjunk igazodni a világban.

A fentebb említett interdiszciplináris jelleg fontosságát az is mutatja, hogy az informatikai gondolkodás nemcsak az adattudományi képzés része, hanem beépülhet a tudományos kutatásról való általános gondolkodásba. Ezt a gondolkodást Wing (2006) olyan problémamegoldásként határozza meg, amely egyaránt figyel a rendszerek tervezésére és az emberi viselkedés megértésére. Egy olyan, általánosan alkalmazható hozzáállást és készségkészletet képvisel, amelyet mindenki, tehát nemcsak informatikusok, hanem a laikusok is szívesen megtanulnának és használnának (Song & Zhu, 2017). Része a több absztrakciós szinten való eligazodás és az adatmodellezés is (Jacob & Warschauer, 2018).

Az adattudósoknak képesnek kell lenniük arra, hogy megértsék a nagy adatokhoz és a feldolgozásukhoz szükséges technológiák szerepét. Ők azok, akik az adattudományi életciklus minden lépésénél ott vannak, hogy feltárják és

megoldják a problémákat (Song & Zhu, 2017). Egy kevésbé ambiciózus meghatározás szerint az adattudomány a számítógépes programozás és a szoftvermérnöki kompetenciák, a statisztika és egy adott szakterület szakértelmeinek elegye. A legtöbben azonban az algoritmusokkal, gépi tanulással és statisztikai technikákkal való munka fontosságát hangsúlyozzák (Burton, Lyon, Erdmann & Tijerina, 2018).

Az adattudomány magában egyesíti az informatika és a statisztika technikáit. A műveléséhez szükséges ismeretek ezért a matematikához és a programozáshoz kötődnek. Fő célja az adatelemzés, valamint új termékek és szolgáltatások kifejlesztése. Számos technikát kombinál, ideértve az adatok szűrését, a gépi tanulást, a rendszertervezést, az elosztott számítástechnikai rendszereket, a statisztikát, a mérnöki ismereteket és a vizualizációt.

Összegzés

Az adatok „szép új világa” akkor jön el, ha a tudomány és hétköznapi életünk terén is tudomásul vesszük az adatok megnövekedett fontosságát. Ez nem történhet meg egyik napról a másikra, ezért az adatumveltségnek különösen fontos szerepet kell tulajdonítanunk. Nem kevesebbről van tehát szó, minthogy gondosan azonosítsuk, melyek azok a különböző szerepek, amelyek betöltését célozzuk. Ezek ismeretében, az életkori sajátosságokat és a változó célokat figyelembe véve kell megterveznünk az oktatás céljait és eszközeit.

Irodalomjegyzék

- Abner, K. (2020). Data Literacy as Digital Humanities Literacy: Exploration of Threshold Concepts. In M. Brooks, M. Hubbard, J. Perkins, & J. Russell (Eds.), *Literacies in a Digital Humanities Context: A dh+ lib Special Issue* (pp. 19-22), dh+ lib.
- ALA (1989). ALA Presidential Committee on Information Literacy, Final report. American Library Association.
- Başkarada, S., & Koronios S. (2017). Unicorn data scientist: the rarest of breeds. *Program: electronic library and information systems*, 51(1), 65-74.
<https://doi.org/10.1108/PROG-07-2016-0053>
- Boyd, D., & Crawford, K. (2012). Az adatrengeteg kínos kérdései: Vitaindító egy kulturális, műszaki és tudományos jelenségről. *Információs Társadalom*, 12(2), 7-23.
<https://doi.org/10.22503/inftars.XII.2012.2.1>

- Burton, M., Lyon, L., Erdmann, C., & Tijerina, B. (2018). *Shifting to Data Savvy: The Future of Data Science in Libraries*. University of Pittsburgh.
- Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123-134.
<https://doi.org/10.1515/libri-2013-0010>
- Carlson, J., Fosmire, M., Miller, C. C., & Sapp Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy*, 11(2), 629-657.
<https://doi.org/10.1353/pla.2011.0022>
- Carretero, S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*, EUR 28558 EN, Publications Office of the European Union, Luxembourg. EUR 28558 EN, <https://doi.org/10.2760/38842>
- CILIP (2018). *CILIP Definition of Information Literacy 2018*. <https://infolit.org.uk/ILdefinitionCILIP2018.pdf>
- Corrall, S. (2019). *Repositioning Data Literacy as a Mission-Critical Competence*. In *ACRL 2019: Recasting the Narrative*, April 10-13, 2019, Cleveland, OH. <http://d-scholarship.pitt.edu/36975/>
- Davenport, Th. H., & Patil, D. J. (2012). *Data Scientist: The Sexiest Job of the 21st Century*. *Harvard Business Review*, 90(5), 70-76.
- Davies, A., Fidler, D., & Gorbis, M. (2011). *Future work skills 2020*. Palo Alto, CA: Institute for the Future, 2011. <http://www.iff.org/our-work/global-landscape/work/future-work-skills-2020/>
- De Veaux, R., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M. Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M. (...) Ye, P. (2017). *Curriculum guidelines for undergraduate programs in data science*. *Annual Review of Statistics and Its Application*, 4, 15-30.
<https://doi.org/10.1146/annurev-statistics-060116-053930>
- Fontichiaro, K., & Oehrli, J. A. (2016). *Why Data Literacy Matters*. *Knowledge Quest*, 44(5), 21-27.
- Gebre, E. H., & Morales, E. (2020). *How 'accessible' is open data? Analysis of context-related information and users' comments in open datasets*. *Information and Learning Sciences*, 121(1-2), 19-36.
<https://doi.org/10.1108/ILS-08-2019-0086>
- Gordon-Murnane, L. (2012). *Big Data: A big opportunity for librarians*. *Online*, 36(5), 30-34.

- Grillenberger, A., & Romeike, R. (2018). Developing a theoretically founded data literacy competency model. In WiPSCE ,18 Proceedings of the 13th Workshop in Primary and Secondary Computing Education (pp. 1-10). Association for Computing Machinery.
<https://doi.org/10.1145/3265757.3265766>
- Hayashi, C. (1998). What is data science? Fundamental concepts and a heuristic example. In C. Hayashi, K. Yajima, H-h. Bock, N. Ohsumi, Y. Tanaka, & Y. Baba (Eds.), *Data science, classification, and related methods* (pp. 40-51) Springer.
https://doi.org/10.1007/978-4-431-65950-1_3
- Holl, A. (2018). A nyílt kutatási adatok kezelésének három oldala. In NETWORKSHOP 2018. (pp. 59-63) HUNGARNET Egyesület.
<https://doi.org/10.31915/NWS.2018.8>
- Hunt, K. (2005). The challenges of integrating data literacy into the curriculum in an undergraduate institution. *IASSIST Quarterly*, 28(2-3), 12-16.
<https://doi.org/10.29173/iq791>
- Ince, S., Hoadley, Ch., & Kirschner, P. (2018). The role of libraries in teaching doctoral students to become information-literate researchers: A review of existing practices and recommendations for the future. *Information and Learning Science*, 120 (3-4), 158-172.
<https://doi.org/10.1108/ILS-07-2018-0058>
- Jacob, S. R., & Warschauer, M. (2018). Computational thinking and literacy. *Journal of Computer Science Integration*, 1(1), 1-19.
<https://doi.org/10.26716/jcsi.2018.01.1.1>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1)
<https://doi.org/10.1177/2053951714528481>
- Koltay T. (2018). Az „írástudó” egyetemi hallgató. *Anyanyelv-pedagógia* 11(1), 5-14.
<https://doi.org/10.21030/anyp.2018.1.1>
- Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3-14.
<https://doi.org/10.1177/0961000615616450>
- Li, S., Jiao, F., Zhang, Y., & Xu, X. (2019). Problems and Changes in Digital Libraries in the Age of Big Data from the Perspective of User Services. *Journal of Academic Librarianship*, 45(1), 22-30.
<https://doi.org/10.1016/j.acalib.2018.11.012>
- Llebot, C., & Rempel, H. G. (2021). Why Won't They Just Adopt Good Research Data Management Practices? An Exploration of Research Teams and

- Librarians' Role in Facilitating RDM Adoption. *Journal of Librarianship and Scholarly Communication*, 9(1), P2321.
<https://doi.org/10.7710/2162-3309.2321>
- Manovich, L. (2012). Trending: The Promises and the Challenges of Big Social Data. In M. K. Gold (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press.
<https://doi.org/10.5749/minnesota/9780816677948.003.0047>
- Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (pp. 108-111). American Society for Information Science.
- Norgaard, R., & Sinkinson, C. (2016). Writing information literacy: A retrospective and a look ahead. In B. J. D'Angelo, S. Jamieson, B. Maid, & J. R. Walker (Eds.), *Information literacy: Research and collaboration across disciplines* (pp. 15-35.) WAC Clearinghouse and University Press of Colorado.
<https://doi.org/10.37514/PER-B.2016.0834.2.01>
- Onyancha, O. B. (2020). Knowledge visualization and mapping of information literacy, 1975-2018. *IFLA Journal*, 46(2), 107-123.
<https://doi.org/10.1177/0340035220906536>
- Pangrazio, L., & Sefton-Green, J. (2020). The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2), 208-220.
<https://doi.org/10.1080/17439884.2020.1707223>
- Poole, A. H. (2017). The conceptual ecology of digital humanities. *Journal of Documentation*, 73(1), 91-122.
<https://doi.org/10.1108/JD-05-2016-0065>
- Robinson, L. (2016). Between the deluge and the dark age: Perspectives on data curation. *Alexandria*, 26(2), 73-76. <https://doi.org/10.1177/0955749016661067>
<https://doi.org/10.1177/0955749016661067>
- Schild, M. (2010). Assessing statistical literacy: Take CARE. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 133-152). Wiley.
<https://doi.org/10.1002/9780470710470.ch11>
- Smith, S. (2013). Is Data the New Media? *EContent*, (36)2, 14-19.
- Shome, A. (2019). Closing the Gap: A Practical Guide to Science in the Writing Center. *Praxis: A Writing Center Journal*, 17(1), 1-9.
- Song, I. Y., & Zhu, Y. (2017). Big data and data science: Opportunities and challenges of iSchools. *Journal of Data and Information Science*, 2(3), 1-18.
<https://doi.org/10.1515/jdis-2017-0011>

- Szemes B. (2018). A szoroson túl. A close reading módszere és lehetséges alternatívái. *Prae: irodalmi folyóirat*, 16(3), 73-90.
- Van der Aalst, W. M. P. (2014). Data scientist: The engineer of the future. In K. Mertins, F. Bénaben, R. Poler, & J.-P. Bourrières. *Enterprise interoperability VI*. (pp. 13-26). Springer.
https://doi.org/10.1007/978-3-319-04948-9_2
- Wilkerson, M. H., & Polman, J. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, 29(1), 1-10.
<https://doi.org/10.1080/10508406.2019.1705664>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, (49)3, 33-35.
<https://doi.org/10.1145/1118178.1118215>
- Z. Karvalics László (2008). Az adatsilóktól a tudomány kontrollforradalmáig. *Magyar Tudomány*, 169(3), 352-363.

Absztrakt

Az „adatvezérelt világ” kifejezés ma már nem egy utópiát takar, hanem egyre gyakrabban az üzleti élet, a média és a tudományos kutatás számos területének valósága. Oktatása nemcsak a kutatókra és az egyetemi hallgatókra terjed ki, hanem igény van arra is, hogy a közoktatásban tanuló diákok elsajátíthassák azokat a készségeket, képességeket és jártasságokat, amelyek szükségesek az adatokkal való munkához. Ezentúl a különféle adatok kezeléséhez szükséges szemlélet és kompetenciák egyre inkább az állampolgári és az információs műveltség részévé válnak. Ez azt is megköveteli, hogy az információs műveltség mellé felzárkózzon az adاتمűveltség, továbbá számos, más kompetencia is szerepet kapjon.