

AZ ADATHASZNOSÍTÁS ADATFOGALMÁNAK MEGHATÁROZÁSI LEHETŐSÉGEI

dr. Kollár Gergő

Tudományos segédmunkatárs, Technológiai Jogi és Energiajogi Tanszék

Pécsi Tudományegyetem, Állam- és Jogtudományi Kar

kollar.gergo@ajk.pte.hu

DOI: [10.47272/KIKPhD.2024.1.2](https://doi.org/10.47272/KIKPhD.2024.1.2)

ABSTRACT

Az adatok hasznosítása egyre inkább előtérbe kerül, mind technológiai, mind jogi kutatási szempontból. Jelen tanulmány az adathasznosítás fogalompárjának első elemének vizsgálatát tűzi ki célul, a jelenleg fennálló hézagos és sokszor félreértésekre alapot adó állapot megszüntetése érdekében. Ennek érdekében a meglévő szakirodalom és fellelhető gyakorlat alapján az adatfogalom több aspektusból kerül megvilágításra, kiemelten a hasznosítással kapcsolatos későbbi kutatások megkönnyítése érdekében.

KEYWORDS

Adat, adathasznosítás, DIKW piramis

I. Bevezetés

Az adathasznosítás területének átlátható szeletekre bontásához és az azon belül található fogalmi egységek értelmezéséhez fontos egy olyan közös nevező létrehozása, amelyre biztos alapok építhetők. Egy ilyen alapvető – jogászok számára kiemelten fontos – terület a definíciós kísérletek, számbavétele és lehetőség szerinti egységesítése. Jelen tanulmány célja az adat fogalmának és típusainak elemi vizsgálata, a komplexebb, nagyobb volumenű kutatások terminológiai szempontú támogatása érdekében. A kutatás eredményeként egy olyan, a gyakorlati és elméleti szempontokat is figyelembe vevő, egységesen alkalmazható adatfogalom-rendszer jön létre amely jobban adaptálható a modern, algoritmizált folyamatokhoz.

II. Az absztrakt adatfogalom és a DIKW-piramis újragondolása

Az „adat” kifejezés rendkívüli népszerűségnek örvend a mai tudományos és „köznap” diskurzusban egyaránt. Az adat fogalom azonban ritkán kerül kellő komplexitással kifejtésre, aminek következtében egy – a szükségesnél nagyobb – értelmezési nehézség alakul ki rögtön a megismerési folyamat legelején. Ez egyébiránt visszatérő eleme azon jelenségeknek, amelyek első ránézésre túlzottan is egyszerűen definiálhatónak tűnnek, hiszen ki ne tudná mit jelent az, hogy adat. Azonban a határterületek, rokon fogalmak és szinonimák számbavétele során a fogalom pontos meghatározása nem is olyan egyszerű feladat. Márpedig abban az esetben, ha az e fogalom alá tartozó elemekkel valamit véghez szeretnénk vinni (pl. hasznosítani kívánjuk őket), tudnunk kell mire gondolunk pontosan, illetve

szükséges az is, hogy ugyanaz a kifejezés alatt mindannyian ugyanarra a jelenségre gondolunk.

Az adatok lényegének megragadása az adat legabsztraktabb rendszertani elhelyezésén keresztül mutatkozik célszerűnek, ami során az említett határterületek is kifejtésre kerülhetnek, ezáltal az adatok esszenciális értelme jobban kiemelhető. Ehhez zsinórmértékül az un. DIKW¹ piramis vagy hierarchia modell² szolgál, amelyet azonban olyan módon alakítottam át, hogy az jobban illeszkedjen az adatok hasznosításának témaköréhez.

1. A folyamat első lépéseként szükség van arra, hogy a körülöttünk lévő valóságban jelenségek létezzenek, illetve, hogy ezekkel kapcsolatban valamilyen változások történhessenek meg, mehessenek végbe. A kiindulópont, tehát relatíve egyszerű, szükség van az általunk megismerhető és dinamikusan változó fizikai valóság (a külvilág tárgyi és az emberi lelkivilág jelenségei³) létezésére. Ennek a szintnek a kiemelése az első eltérés az eredeti modellhez képest, ezt pedig összefoglalóan **tényszerűségeknek**⁴ nevezhetjük. Fontos tulajdonságuk, hogy állandóan és a szemléltől függetlenül vagy legfeljebb azzal kölcsönhatásban (az un. megfigyelési effektus által [Observer Effect]⁵) léteznek. Ez alapján tehát a tényyszerűséggel ok-okazati összefüggésben bármilyen állhat, de a folyamat lezárását követően azon változtatni nem lehetséges, a valóságunk elválaszthatatlan részévé válnak. Bár egy meglehetősen elvont kategóriának tűnhet, ennek a szintnek a vizsgálata a jogirodalom számára sem ismeretlen terület. Tárgyi jogi szempontból releváns tényyszerűségeket jogi tényeknek nevezzük, amelyeket alapvetően három kategória szokás bontani. Ezek az emberi magatartások, a természetes és épített környezeti folyamatok és események, valamint az állami aktusok.⁶ *Tényszerűségnek tűnik a fogalom alapján, hogy a bolygónkon a H₂O molekula szobahőmérsékleten folyékony halmazállapotú, de tényyszerűség az is, hogy én ebben a mondatban ezt állítom.*⁷
2. A következő szint maga a kutatott fogalom, az **adat**, amely a tényyszerűséget leíró alapvető paramétereket foglalja magába. Így tehát az adat valójában egy átmeneti állapotot, egy potenciált jelent, amely – Schrödinger macskájához

¹ DIKW akronim jelentése: Data – Information – Knowledge – Wisdom (magyarul: Adat – Információ – Tudás – Bölcsesség)

² Jennifer Rowley: The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163-180. o. <https://doi.org/10.1177/0165551506070706>

³ Kengyel Miklós: *Magyar Polgári Eljárásjog*. Budapest, Osiris Kiadó, 2014. 287 o.

⁴ A magyar nyelv értelmező szótára (www.mek.oszk.hu): „A valóság folyamatában lezárt szakaszt alkotó esemény, jelenség, mozzanat; olyan, ami a valóságban előfordulhat v. megtörtént.” (2024.04.30.)

⁵ Paul Adrien, Maurice Dirac: *The Principles of Quantum Mechanics*. Oxford, Clarendon Press, 1981. 314 o.

⁶ Visegrády Antal: *Jogi alaptan*. Pécs, Janus Pannonius Tudományegyetem, Állam- és Jogtudományi Kar, 1996. 44. o.

⁷ Igaznak hat a kijelentés, de fontos, hogy a tényyszerűséget nem tudjuk teljes valójában vizsgálni, csak a rendelkezésre álló adatokon keresztül. Még egy ilyen triviális esetben sem zárható ki, hogy a rögzített adatok hiányosak vagy pontatlanok és a kijelentés részben, vagy egészben téves.

hasonlóan⁸ – vagy manifesztálódik a szemlélő jelenléte által, vagy nem. Az minden esetben objektív, azonban önmagában csak egy nyers értéket hordoz.⁹ *Adatnak minősülhet például minden, amit a vízzel kapcsolatban rögzíteni lehetséges, a fizikai és kémiai összetétele, a feltételek, amiknek a fennállása esetén a halmazállapota megváltozik, hatásai a környezetére, az emberi szervezetre stb.*

A modellben a soron következő elem az **információ** lenne, ami az értelmezhető és hasznosítható, egyediesített értéket hordozó adatot jelenti.¹⁰ Az információ hasznos és értékes kvalitásainak kiemelése helyénvaló, de a hierarchiában történő önálló szerepeltetéssel kapcsolatban az álláspontom az, hogy – például az algoritmizált folyamatok térnyerése által – a fogalmak egyre inkább összerosódnak, ezáltal kiüresítve az adat és az információ ilyen kategorikus elválasztásának megfelelőségét. Ennek egyik oka, hogy az „érték”¹¹ adathoz rendelése időben és térben is elválhat magától az alapfolyamattól, mivel annak résztvevői számára az érték nem ítéhető meg megalapozottan az adott pillanatban. A másik kiemelt ok a külön szerepeltetéssel szemben, hogy információelméleti szempontból az adat és az információ a megismerési folyamatban szoros kölcsönhatásban működik, ugyanis utóbbit a folyamat tartalmának, míg előbbit annak formájának tekinthetjük.¹² A tanulmány további részében az információt az adat egyedi esetben kikristályosított *beltartalmának* tekintem, tehát az adat hordozza az információt, de azok nem válnak el élesen egymástól. *Tehát jelen értelmezés szerint minden információ adat, de nem minden adat információ.*

3. Az adatok rögzülését és befogadását követően, azokon – a megismerési folyamat lépéseinek keretében – műveletek végezhetők. Logikailag legelső lépések között kell szerepeljen az adatok eltávolítása és valamilyen szempontok szerinti (pl. adatbázisokba) rendezése. Az ilyen, legalább tárolt és rendezett, adatok alapul szolgálhatnak **ismeretek** megszerzésére, amik valamilyen fajta, megismerő tevékenység eredményeként, az ennek segítségével kialakított, a valóságra, valamint annak valamely területére vonatkozó fogalmak összességeként foghatunk fel.¹³ Fogalmazhatunk tehát úgy, hogy az adat rögzítettsége által képessé teszi a szemlélőt tapasztalatok megszerzésére, vizsgálatok lefolytatására. Ez által újabb ténytérületek fedezhetők fel, amelyek további adatok rögzítését teszik lehetővé, ezért az ismeret az adat feldolgozása során létrejövő értelemmel is azonosítható.

⁸ John D. Trimmer: The present situation in quantum mechanics: a translation of schrödinger's "cat paradox paper. *Proceedings of the American Philosophical Society*, 124(5), 323-338. o.

⁹ Todd Groff, Thomas Jones: *Introduction to Knowledge Management*. London, Routledge, 2003.

¹⁰ Kenneth C. Laudon, Jane P. Laudon: *Management Information Systems: Managing the Digital Firm, Global Edition*. London, Pearson Higher Ed, 2021.

¹¹ Lásd: jelen tanulmány 4. fejezete

¹² Balogh Zsolt György: *Jogi Informatika*. Budapest-Pécs, Dialóg campus Kiadó, 1998. 17. o.

¹³ A magyar nyelv értelmező szótára (www.mek.oszk.hu) (2024.04.30.)

Ismeret lehet, hogy vizet kell inni az életben maradáshoz, illetve ismeret az is, hogy túl sok vízbe bele lehet fulladni.

4. Az előző szintre alapozva valamilyen ismeret birtokát megszerezni, rendszerbe szedni, összekapcsolni más ismeretekkel és ezekből következtetéseket levonni magasabb szintű **tudást** fog eredményezni a ténszerűséggel kapcsolatban.¹⁴ Fogalmazhatunk úgy, hogy az ismeretek építőelemei valami nagyobb, bonyolultabb rendszernek, amelynek a felismerési (felépítése) által az értelem elhelyezhető a valóságban. *Tudást jelenthet, hogy egy anyag vagy felület egy szennyeződéstől vízzel megtisztítható, azonban forró víz használatával még bizonyos kórokozók is elvárolíthatóvá válnak.*
5. Az utolsó szintje a piramisnak, az értelem valóságnak megfelelő alkalmazása, amit **bölcsesség**nek nevezhetünk.¹⁵ *Bölcsesség lehet, ha forró vízzel fertőtleníjük a cumisüveget mielőtt egy újszülöttnak adnánk, mivel az immunrendszere fejletlensége miatt fokozott veszélynek van kitéve.*

A hierarchia öt pontja alapján látható, hogy a legelemibb törvényszerűségektől, a magas szintű gondolkodásig, az értelem működéséig milyen út megtételére van szükség. Talán az is észlelhető, hogy az adat milyen helyet foglal el a folyamatban, hiszen a piramis második szintje, vagyis az adat az, ami összeköti a külső világ történéseit az elme belső folyamataival, ablakot nyitva a világ ténszerűségeinek megismerése. A piramis-szerűen felépülő ábra mellett a folyamat-jelleg is hangsúlyozható, amihez az algoritmizált rendszerekből ismert tulajdonságok felhasználása által tovább pontosítható a megismerési tevékenység. Egy algoritmus – amit egy zárt rendszeren belül lezajló, logikailag összekapcsolt lépések láncolataként foghatunk fel – működése alapvetően három működési fázisra osztható, egy bemenetre, egy számítási teljesítményre és egy kimenetre. Az algoritmizált folyamatnak tehát lényegi eleme, hogy a „bemenetből” létrejöjjön a „kimenet”.¹⁶ A fenti fogalmak által jellemzett megismerési folyamat meglátásom szerint megfelelően igazodik ebbe konstrukcióba. E tekintetben a bemeneti oldalon láthatóak ténszerűségekből leképzett adatok – amelyek bizonyos esetekben információként jelenhetnek meg – míg a kimenetet első lépésként az ismeretek alkotják. Az algoritmusok törvényszerűségeinek megfelelően az előző folyamat kimenete könnyen válhat a következő folyamat bemenetévé, így lehetséges, hogy az ismeretek (adatainak) további feldolgozása (a megismerési tevékenység mélyítése) által eljuthatunk a tudásig, majd később a bölcsességig. Ezek a ciklusok végtelenül kombinálhatók, a piramis egyes fokai között pedig a hierarchikus felépítettségre tekintet nélkül létezik átjárhatóság. Így nem kizárt, hogy egy tudásalapú megismerési

¹⁴ David Boddy, Albert Boonstra, Graham Kennedy: *Managing Information Systems: Strategy and Organisation*. London, Pearson Education, 2008.

¹⁵ Leonard M. Jessup, Joseph S. Valacich: *Information systems today: managing in the digital world*. London, Pearson Prentice Hall, 2008.

¹⁶ Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein: *Introduction to Algorithms. Third Edition*. Cambridge, Massachusetts, London, England, The MIT Press, 2009. 5. o.

folyamat eredményeként eddig nem látott tényyszerűség válnak észlelhetővé, amelyekből leképzett adatok ismereteink bővítésére alkalmasak, azoknak azonban gyakorlati alkalmazhatósága egyéb tudás birtokának hiányában még nem lehetséges.

Megállapítható az is, hogy egy tényyszerűség egy aspektusa egy adott pillanatban csak egyféle adat leképezését teszi lehetővé, az adatnak a szemléltető számára a befogadása pedig lehet könnyebb vagy nehezebb feladat. A rögzítése lehet teljes és pontos, de lehet részleges és tökéletlen is. Az adatok gyűjtésének és rögzítésének módjai, valamint a felhasználás céljai nagyban meghatározzák annak jellemzőit, amelyeket legalább technológiai szempontból, jogi megítélésük szerint és a bennük rejlő érték alapján kategorizálhatjuk.

III. Az adat fogalmának meghatározási és tipizálási kísérletei

3.1. A technológiai adatfogalom

Az adat szó használata és annak térnyerése mindig szorosan együtt mozgott a tudomány és a technológia fejlődésével. Mint főnév – bár latin eredetű, a *dare* (adni) igéből származik – az 1640-es évektől tényyszerűségek azonosítására szolgált, később az 1890-es évektől már kifejezetten a számszerűsített tényeket nevezték adatoknak. A mai értelmét az informatika megjelenésével szerezte meg. 1946-ban „továbbítható és tárolható információként” jelölték az adatot, amellyel „számítógépes műveleteket végezhetők”, ezt követően 1954-ben jelentek meg az adatfeldolgozás és 1962-ben az adatbázis szóösszetételek.¹⁷ Az adat technológiai szempontból alapvetően két nagy kategóriái osztható, digitális és analóg adatokra, amely megkülönböztetés alapja a rögzítéshez felhasznált technológia, valamint az adathordozó típusa.

3.1.1. Az analóg adatok

Analóg adatnak tekinthetünk egyfelől minden formák megjelenítésére alkalmas mesterséges vagy természetes felületen, illetve anyagon, leképzett adatot. Ezen kategória elsődleges képviselői a papír alapú iratanyagok, de ide tartoznak például a műanyagokra nyomott különféle kártyák, a vászonra felvitt festmények, illetve polaroid fotók is. Másfelől léteznek modern analóg eszközök, amelyeknek közös jellemzője, hogy bennük az adatáramlás és tárolás folytonos analóg jelekkel (pl. elektronikus, mechanikus, pneumatikus, hidraulikus) történik (ellentétben a digitálissal, ahol ez diszkrét, számokból álló formában történik).¹⁸ Az ilyen analóg eszközök tovább bonthatók nem-elektromos és elektromos eszközökre. Előbbiekre példaként említhetők a napórák, a higanyos hőmérők vagy a mérlegek. Utóbbiak

¹⁷ Etymonline: *Online Etymology Dictionary*. Harper Family LLC. <https://www.etymonline.com/word/data> (2024.04.30.)

¹⁸ W.K. Jenkins: *Signal Processing, Analog*. In: Robert A. Meyers (Eds.): *Encyclopedia of Physical Science and Technology (Third Edition)*. Cambridge, Academic Press, 2003. 711-735. o. <https://doi.org/10.1016/B0-12-227410-5/00686-4>

között szerepel a legtöbb ma használt digitális technológiai eszköz előképe, mint az analóg számítógép, az analóg televízió és rádió, vagy a teleautográf.¹⁹

3.1.2. A digitális adatok

A digitális adatok – tehát számítógépek számára használható adatok²⁰ – így az informatika működési elveinek megfelelően döntően numerikus értékekkel (jellemzően bináris rendszerben) kerülnek létrehozásra, tárolásra és továbbításra.²¹ Az Információs és Kommunikációs Technológiai (IKT) eszközök teljes működési ciklusuk alatt, folyamatosan valamilyen adatkezelési műveletet végeznek, meglévő adatot gyűjtenek, illetve rögzítenek, új adatot hoznak létre, mindezeket tárolják, összekapcsolják, valamilyen cél érdekében felhasználják vagy továbbítják. Minden rendszer kezdő pontján (input), a működése során és a folyamat zárásakor (output) adatokon végzett műveletek zajlanak. Egy operációs rendszer napló fájljainak generálásától, az okostelefonok térkép applikációjának útvonalterv készítésén keresztül az olyan bonyolult folyamatokig, mint egy fejlett nyelvi modell kommunikációja mind ezen az elven működik. A digitális adatkategóriához tartozik a szöveg, a kép (álló vagy mozgó), a hang, a szám és bármilyen automatikusan rögzített technikai leírás (logok) is, továbbá ide tartoznak a rögzített tartalmakat leíró egyéb információk (metaadatok) is. Ezek az informatika állása és az alkalmazott keretrendszer szerint sokféle fájlformátumban létezhetnek, ezeknek a változatossága pedig szintén egyre bővül.²² Így szöveg lehet a TXT vagy DOCX, kép a JPG, míg audió a FLAC, a videó pedig az AVI.

21

3.1.3. A Big Data jelenség

A digitális adatokkal kapcsolatban sokszor kerül megemlítésre az un. „Big Data” vagy az adathalmazok jelensége, amely a világon adott pillanatban, potenciálisan rendelkezésre álló (hatalmas és gyorsan növekvő mennyiségű, változatos tartalmú) összes adatot jelöli. A Big Data tehát nem jelent egy önálló adatkategóriát, azonban a téma szempontjából kiemelt jelentőséggel bíró digitális adatok mai leglényegesebb megjelenési formája. Ezek a nagy adathalmazok többnyire nyers adatként léteznek, így azok további feldolgozása szükséges, ahhoz, hogy ténylegesen hasznosíthatóvá váljanak. A Big Data mára már kevésbé számít nóvumnak, hiszen a egyes felfogások

¹⁹ The Writing Telegraph. *Scientific American*, 11(13), 1879. 196. o. <https://doi.org/10.1038/scientificamerican03291879-196a>

²⁰ Juan O. Freuler: Datafication, Identity, and the Reorganization of the Category Individual. *Temple Law Review* 95(4), 2023. 378 o.

²¹ Britannica.com, Encyclopædia Britannica: *Digital computer*. <https://www.britannica.com/technology/digital-computer> (2024.04.30.)

²² Adanma Cecilia Eberendu: Unstructured Data: an overview of the data of Big Data. *International Journal of Emerging Trends & Technology in Computer Science* 38(1), 46-50. <http://dx.doi.org/10.14445/22312803/IJCTT-V38P109>

szerint legalább 2012-óta a Big Data korában élünk,²³ ettől függetlenül az ezzel kapcsolatos tanulmányok a téma szempontjából továbbra is relevánsak. A Big Data mögött tatláható adatok ugyanis kiemelt forrásai a hasznosítható – ezáltal technológiailag és jogi szabályozás szempontjából kezelendő – adatoknak. Ahogy az elnevezés is mutatja a Big Data alapvetően nagy (hatalmas) mennyiségű adat létezését és ezeknek a fejlett társadalmak általi folyamatos generálását takarja.²⁴ A nagy adathalmazokat több szempont szerint lehet leírni, annak jobb megértése érdekében. Ezt a felosztást (angol nyelven) összefoglalóan az 5V attribútumoknak szokás nevezni:²⁵

- mennyiség (volume), ami rendkívül nagy, a 2025-re a becsült Big Data állomány elérheti a 181 zetabyte²⁶ mennyiséget,²⁷
- sebesség (velocity), ami a gyűjtés sebességét jelenti és szintén értékelendő, hogy sok esetben valós időben történik (nem véletlen tehát, hogy az előbbi mennyiségi tényező exponenciális bővülését tapasztaljuk),
- változatosság (variety), a változatosság jelen esetben a formátumok változatosságára utal, ez a digitális adattípusoknál taglaltakkal megegyező kategória (így pl. TXT, FLAC, AVI, XML stb.),
- hitelesség (veracity) alatt az adatforrás megbízhatóságát (annak bizonytalanságát) értjük,
- értékesség (value), ami az adatból kinyerhető hasznosíthatóságát (tehát az adott cél elérésére való alkalmasságot) takarja.

Az adathalmazoknál általában, de különösen a Big Data jellegű adathalmazok esetén fontos kitérni az adatok strukturáltságának kérdésére, mivel ez nagyban befolyásolja a hasznosíthatóságukat (legalábbis annak bonyolultságát). A strukturált adat séma szerint működő, rendezett, egyértelmű formátumú és méretű adat, amelyet elsődlegesen kereshetősége miatt könnyebb elemezni. Példaként említhető egy relációs adatbázis (SQL), amelyben az adatok rendezetten jelennek meg.²⁸ Részben strukturált adat mutathat strukturáltságot (pl. címkézett), ez azonban nem illeszkedik egy könnyen felismerhető sémába.²⁹ Részben strukturált adatok például az XML, CSV és JSON fájlok. Strukturálatlan adatok nem rendezettek és nem

²³ Steve Lohr: *The Age of Big Data*. <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html> (2024.04.30.)

²⁴ Alexandra Anghel, Elena Novacescu, Madalina Cuc: Big Data and Its Secrets: Types of Big Data. *Romanian Intelligence Studies Review*, 2022/27. 54-73 o.

²⁵ Amy Affelt: Big Data, Big Opportunity. *Australian Law Librarian*, 21(2), 2013. 78-89. o.

²⁶ Petroc Taylor: Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025. STATISTA, 2023. <https://www.statista.com/statistics/871513/worldwide-data-created/>. (2024.04.30.)

²⁷ A mérték érzékeltetése érdekében megjegyzendő, hogy 1 zetabyte egyenlő 1 000 000 000 terabyte-al.

²⁸ Adanma Cecilia Eberendu, 2016., i.m.

²⁹ Alexandra Anghel, et. al., 2022., i.m. 54-73 o.

követnek semmilyen sémát. Strukturátlan adatok például a PDF-ek, az audiovizuális és képi tartalmak.³⁰

A nagy adathalmazok térnyerésének és a mindennapi életünkhöz szervesen kapcsolódó pontjainak megértéséhez további négy, a jelenség alapvetően input oldalát, tehát a bemenő adatokat befolyásoló tényezőről szükséges röviden szót ejteni. Ezek a Dolgok Internete eszközök, az adatosítás társadalmi trendje, valamint a cloud computing és a web 2-es alkalmazások használata.

- A Dolgok Internete (Internet of Things, IoT) hálózatra kötött, egymással jellemzően IP³¹ alapon kommunikáló és szenzorosan adatokat gyűjtő, valamint továbbító eszközöket jelenti.³² Az IoT egyre elterjedtebbé válik számos iparágban, többek között a közlekedésben, a mezőgazdaságban és az egészségügyben is. Az IoT eszközök olyan előnyökkel járnak, mint valós idejű adatgyűjtés és feldolgozás, a pontos és aktuális visszacsatolási hurkok létrejötte, valamint a távoli felügyelet és irányítás képessége, amelyek a köz- és a magánszféra szervezetei számára egyaránt nagyobb működési hatékonyság elérését teszik lehetővé.³³ Az IoT eszközök árnyoldalához tartoznak különösen adatbiztonsági és adatvédelmi aggályok, úgy, mint a túlzó adatgyűjtési tendenciák, adatbázisok jogellenes összekapcsolása vagy a kibertámadásoknak való fokozott kitettség.
- Az internet fennállása alatt több átalakuláson is keresztül ment, az adatok szempontjából legnagyobb horderejű a második korszaka (web 2.0). Az előző állapotához képest, amelynél az adatáramlás leginkább vertikálisan épült fel és egy egyirányú útként képzelhető el az adatot létrehozó és az adatot fogyasztó felhasználó között. A web 2.0 adatáramlása ehhez képest horizontálisan épül fel, az adatot létrehozó és a felhasználó mellérendeltek, sőt a két pozíció jellemzően egymásba is csúszik, tehát az adatot létrehozó egyszerre felhasználó is és fordítva, a felhasználó maga is létrehozza az adatot.³⁴ A web 2.0 állapot lehetőséget biztosított mindenki számára (cégek, államok és magánszemélyek) az adathoz való hozzáférésre és az adat megosztására egyaránt, ami nagyban járul hozzá a nagy adathalmazok mennyiségi duzzasztásához.
- Az adatosítás („datafication”) egy olyan modern jelenség, amely a világ eseményeinek digitális adattá történő átalakításának folyamatát, egyre

³⁰ Christian Hänig, Martin Schierle, Daniel Trabold: *Benefits of Unstructured Data for Industrial Quality Analysis*. In: Sio-long Ao, Mahyar Amouzegar, Burghard Rieger (Eds.): *Intelligent Automation and Systems Engineering. Lecture Notes in Electrical Engineering, vol 103*. New York, Springer. https://doi.org/10.1007/978-1-4614-0373-9_20

³¹ Az internetprotokoll, az interneten keresztüli kommunikáció egyik alapvető szabványa
³² Techopedia: *What Does internet of Things (IoT) Mean?*
<https://www.techopedia.com/definition/28247/internet-of-things-iot> (2024.04.30.)

³³ Larisa-Antonia Capisizu: *Legal Perspectives on the Internet of Things*. In: *Conferinta Internationala de Drept, Studii Europene si Relatii Internationale*. Bucharest, Titu Maiorescu University, 2018. 523. o.

³⁴ Darin K. Fox: *What Is Web 2.0. AALL Spectrum* 13(9), 2009. 2.o.

növekvő trendjét jelenti.³⁵ Mérhető a hatása ebben a folyamatban is az IoT eszközök térnyerésének, de az adatosítás a társadalmi együttélés digitalizációjának egy aspektusaként is azonosítható.³⁶ Az online terek (és az azokat működtető szolgáltatók) egyértelmű törekvése, hogy a felhasználók minél több mindennapi emberi magatartásukat és életük átlagos eseményeit digitális adatként is leképezzék. Mindkét olvasat jelentősége kifejezetten az új, korábban adatként nem létező adattípusok összegyűjtésében, illetve generalálásában rejlik. A különböző, web 2.0-es olvasztótégelyben megalakult, közösségi média platformok élen járnak az adatosítás társadalmi jelenlétének proliferációjában. A felületeik és funkcióik kiosztásával, valamint fejlett algoritmizált eszközeik és hatalmas mennyiségű felhasználói személyes adat birtoklásával képessé váltak alapvető hormonális impulzusok és tevékenységi hurkok létrehozására, amelyekkel maximalizálni képesek az adatmegosztási késztetést.³⁷

Összegezve technológiai szempontból a hasznosítással érintett adatok túlnyomóan a digitális adatkategóriákhoz tartoznak, ezeket egyre inkább valamilyen algoritmizált eszközzel vagy annak részeként hasznosítják.

3.2. A jogi adatfogalmak

Az adatfogalom normatív meghatározottságát a vizsgált jogszabályokban a jogág vagy jogterület szektorális jellege és a joganyag megalkotásának ideje nagyban befolyásolja. A számítástechnika fejlettségi foka és az adatok jelentőségének felismerése 2010-es években érte el az európai jogalkotást, ezért a korábbi aktusok az adatokat kevésbé tematizáltan kezelték. Időbeli meghatározottságtól függetlenül döntéshozók jogot a jogpolitikai célkitűzéseik mentén alkotnak, ott, úgy és akkor, amikor az életviszonyok szabályozásra szorulnak. E két tényező együttesen sporadikus, sokszor csak egy jogszabályi kontextusban izolált adat fogalom meghatározást eredményezett (ez alól az uniós jogalkotás 2020 után egyértelmű kivételt jelent). Szintén az ágazati jogalkotásból következik, hogy bár az adatfogalmak jogterületekhez igazodnak, de azok között létezik átjárás, azáltal, hogy egyes fogalmakat a jogszabályok kereszthivatkozzák, a helyett, hogy újból definiálnák azokat (tipikus példa erre a személyes adat fogalma). A vizsgálatom az Unió és hazai jogszabályok áttekintésére irányult, ezek közül is elsődlegesen azokra, amelyek leginkább kötődnek a tanulmány témájához.

³⁵ Ulises A. Mejias, Nick Couldry: Datafication. *Internet Policy Review Journal On Internet Regulation*, 8(4), 2019. <https://doi.org/10.14763/2019.4.1428>

³⁶ Clare Southerton: Datafication. In: Laurie Schintler, Connie McNeely (Szerk.): *Encyclopedia of Big Data*. New York, Springer, 2020. 1-4 o. https://doi.org/10.1007/978-3-319-32001-4_332-1

³⁷ Trevor Haynes: *Dopamine, Smartphones & You: A battle for your time*. Harvard University, Science in the News. <https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/> (2024.04.30.)

3.2.1 Az adat fogalma a releváns jogforrásokban

Az adat általános érvényű definiálására legmagasabb szinten (és a témával legszorosabb kapcsolatban) viszonylag későn, 2022-ben került sor, az Adatkormányzási rendelet³⁸ által. Ez a fogalom megjelenik több másik, az Unió „digitális jogalkotási hullámához” tartozó joganyagban, így a Digitális Piacokról szóló rendeletben³⁹ és az Adatrendeletben.⁴⁰ A definíció szerint az adat: „aktusok, tények vagy információk bármilyen digitális megjelenítése, vagy az említett aktusok, tények és információk összeállításai, többek között hang-, kép- vagy audiovizuális felvétel formájában is”.⁴¹ A fogalom aktusok, tények és információk megjelenítésére utal, ahol az aktus (tett, cselekedet⁴²) és a tények nagyban megegyeznek a DIKW-piramis alapján általunk meghatározott absztrakt adat-fogalommal. Az aktusok, tények és információk összeállításainak említése az adatok eredeti és származékos jellegére utal helyesen. Ezen jogi definíció egyetlen deformitása az információ fogalomba építése, amit az adat háttérben álló entitásként fog fel. Megjegyzendő, hogy ez egy létező, a gyakorlatban sok területen uralkodó álláspont, azonban a saját meglátásom szerint az információ sokkal inkább az adatok következő létállapota, egyediesített értéke. Tehát az adat nem az információ megjelenítése, hanem annak potenciális hordozója. Mindentől függetlenül a fogalom elméletben és gyakorlatban is alkalmazhatónak mutatkozik, kiemelendő továbbá, hogy formátumra utalást nem tartalmaz, így bármilyen adatkezelési helyzetre adaptálható, bár a kiemelt adattípusokból (pl. audiovizuális tartalmak) a digitális adatok vélelmezett túlsúlyára lehet következtetni.

Elhagyva az Unió jogot, hazai jogunk eredményeként egy saját adat fogalom is említhető, amely az Ibtv-ben⁴³ található. Ez alapján adat: „az információ hordozója, a tények, fogalmak vagy utasítások formalizált ábrázolása, amely az emberek vagy automatikus eszközök számára közzétre, megjelenítésre vagy feldolgozásra alkalmas.”⁴⁴ Ez a fogalom a törvény megalkotásának idejére tekintettel valamivel absztraktabb, ettől függetlenül mind szóhasználatában, mind logikájában fedi az eddigiekben felépített adatfogalmunkat. Ehhez a jogszabályhoz kapcsolódik

³⁸ Az Európai Parlament és a Tanács 2022. május 30-i (EU) 2022/868 rendelete az európai adatkormányzásról és az (EU) 2018/1724 rendelet módosításáról (továbbiakban: Adatkormányzási rendelet vagy DGA) 2. cikk 1. pont

³⁹ Az Európai Parlament és a Tanács 2022. szeptember 14-i (EU) 2022/1925 rendelete a digitális ágazat vonatkozásában a versengő és tisztességes piacokról, valamint az (EU) 2019/1937 és az (EU) 2020/1828 irányelv módosításáról (továbbiakban: Digitális Piacokról szóló rendelet vagy DMA) 2. cikk 24. pont

⁴⁰ Az Európai Parlament és a Tanács (EU) 2023/2854 rendelete (2023. december 13.) a méltányos adathozzáférésre és -felhasználásra vonatkozó harmonizált szabályokról, valamint az (EU) 2017/2394 rendelet és az (EU) 2020/1828 irányelv módosításáról (továbbiakban: Adatrendelet vagy DA) 2. cikk (1) bekezdés

⁴¹ DGA 2. cikk 1. pont; DA 2. cikk (1) bekezdés; DMA 2. cikk 24. pont

⁴² *Aktus*. Idegen Szavak Gyűjteménye. <https://idegen-szavak.hu/aktus> (2024.04.30.)

⁴³ Az állami és önkormányzati szervek elektronikus információbiztonságáról szóló 2013. évi L. törvény (továbbiakban: Ibtv.) 1. § (1) bekezdés 1. pont

⁴⁴ Ibtv. 1. § (1) bekezdés 1. pont

az egyetlen információ fogalom is, ami nem is meglepő a törvény által szabályozott terület (információbiztonság) figyelembevételével. Az Ibtv. szerint információ: „bizonyos tényekről, tárgyokról vagy jelenségekről hozzáférhető formában megadott megfigyelés, tapasztalat vagy ismeret, amely valakinek a tudását, ismeretkészletét, annak rendezettségét megváltoztatja, átalakítja, alapvetően befolyásolja, bizonytalanságát csökkenti vagy megszünteti.”⁴⁵ Bár az információ és az adat relációjának tisztázása elmarad, de elmondható, hogy hasonlóan az adatfogalomhoz az információ meghatározása is megfelel jelen tanulmány által képviselt felfogásnak, abban megjelenik az egyediesített érték hordozása, illetve a ténszerűségek reprezentációs funkciója.

A normatív adatfogalmak vertikális és horizontális irányokban ágaznak szét. A legalapvetőbb, vertikális különbségtétel a személyes adat és a nem-személyes adat megkülönböztetéssel valósul meg, ezt metszik valahol horizontálisan az egyéb adatfogalmak.

3.2.2. Az adatok jogi szempontú csoportosítása

Az egyes, jogrendszerben előforduló adattípusok igen színes képet mutatnak, így a csoportosításuk során egy egyszerű, az adathasznosítás jogi szabályozása szempontjából azonban fontos kiindulópontként szolgáló felosztást ismertettek,⁴⁶ amely az adatokat egyrészt személyes adatra és nem személyes adatra, másrészt, e kategóriákat horizontálisan érintve közadatokra és üzleti adatokra bontja.

	Személyes adat	Nem személyes adat
Közzsféra	Adatvédelmi szabályozás > GDPR: (EU) 2016/679 Rendelet + tagállami jog + EU/tagállami szektorális adatvédelmi szabályok Minősített adatok > Tagállami szintű szabályozás Horizontális szabályozás: (EU) 2022/868 rendelet (Adatkormány + további horizontális szabályok (korlátok): titokvédelem, szel	Információszabadság > Tagállami szabályozás – közérdekű adatok megismerése és terjesztése Közadat-újrahasznosítás > (EU) 2019/1024, irányelv (nyílt adat irányelv) + tagállami jog (Közadat tv.) Minősített adatok > Tagállami szintű szabályozás rendelet) ulajdonjog
	Adatvédelmi szabályozás > GDPR: (EU) 2016/679 Rendelet + tagállami jog + EU/tagállami szektorális adatvédelmi szabályok Személyes adatok megosztására irányuló szabályozás > Adatmegosztási rendelet-tervezet, Data Act, COM(2022) 68 fina Üzleti titok > (EU) 2016/943 irányelv (üzleti titok irányelv) + tagállami szabályozás	Információszabadság > Tagállami szabályozás – közérdekű adatok megismerése és terjesztése Közadat-újrahasznosítás > (EU) 2019/1024, irányelv (nyílt adat irányelv) + tagállami jog (Közadat tv.) Minősített adatok > Tagállami szintű szabályozás rendelet) ulajdonjog Nem személyes adatok megosztására irányuló szabályozás > (EU) 2018/1807 Rendelet a nem személyes adatok szabad áramlásáról > Adatmegosztási rendelet tervezete (Data Act, COM(2022) 68 final. Üzleti titok > (EU) 2016/943 irányelv (üzleti titok irányelv) + tagállami szabályozás

⁴⁵ Ibtv. 1. § (1) bekezdés 25. pont

⁴⁶ A felosztás és az alábbi ábra Szőke Gergely Lászlótól származik. Szőke Gergely László: Mikor tűnik el az ember a személyes adatok mögül? - Avagy az anonimizálás kihívásai. Konferencia-előadás „Az ember a legújabb technológiák között” c. konferencián, Budapest, Nemzeti Közszoigálati Egyetem, 2023.05.31.

3.2.2.1. *A nem személyes adatok*

Jogi szempontból különös jelentősége van annak, hogy egy adat egy természetes személyre vonatkozik-e vagy sem. Amennyiben igen, úgy a kezelésre vonatkozó előírások rendkívül kiterjedtek és részletesek, az alapvető szabályozási hozzáállás pedig az adatok védelmét célozza. Ezzel szemben a nem-személyes adatokra vonatkozó rendelkezések lényegesen heterogénebbek, bizonyos esetekben az adatok védelmét írják elő, míg más esetekben azok kifejezett nyilvánosságra hozatalát.

3.2.2.2. *A személyes adatok*

A jelenleg hatályos jog szerint a személyes adatok: „azonosított vagy azonosítható természetes személyre vonatkozó bármely információ.”⁴⁷ Az azonosíthatósággal kapcsolatban ugyanitt további útmutatás található, e szerint „azonosítható az a természetes személy, aki közvetlen vagy közvetett módon, különösen valamely azonosítója (...) alapján azonosítható”. A fogalom láthatóan meglehetősen tág, az azonosíthatóság első ránézésre úgy tűnik, hogy a direkt azonosító adatok⁴⁸ alapján való azonosíthatóságot preferálja, ez azonban az újfent tágan meghúzott „közvetett” azonosíthatóság lehetőségével felülírásra került. Az azonosíthatóság kapcsán a GDPR preambuluma szerint „minden olyan módszert figyelembe kell venni [...], amelyről észszerűen feltételezhető, hogy az adatkezelő vagy más személy a természetes személy közvetlen vagy közvetett azonosítására felhasználhatja.”⁴⁹ Ez a mérce tehát sok esetben csak a körülmények pontos ismeretében teszi eldönthetővé, hogy valamely adat személyes adatnak vagy anonim adatnak minősül-e. E kérdéskör jelentősége, „tétje” az adathasznosítás iránti igény növekedésével együtt az elmúlt években jelentősen megnőtt, és döntő szerepe van abban, hogy mely jogi norma alkalmazandó.⁵⁰ A 2018-óta eltelt gyakorlat ismeretében elmondható, hogy a jelenlegi személyes adat fogalom tartalmi szempontból meglehetősen inkluzív, olyannyira, hogy egy frappáns közmondás szerint „ha valamivel kapcsolatban felmerül, hogy személyes adat, akkor valószínűleg az is”. Rendszertani megítélésük szempontjából hangsúlyozandó, hogy ezek az adatok kivétel nélkül természetes

⁴⁷ Az Európai Parlament és a Tanács 2016. április 27-i (EU) 2016/679 rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (továbbiakban Általános adatvédelmi rendelet vagy GDPR) 4. cikk 1. pont; Tartalmilag nagyon hasonló az Az információs önrendelkezési jogról és az információszabadságról szóló 2011. évi CXII. törvény (továbbiakban: Info.tv.) 3. § 2. pontjában meghatározott személyes adat fogalom, amely szerint ez az érintettre (értsd: természetes személyre) vonatkozó bármely információ

⁴⁸ Például név, szám, helymeghatározó adat, online azonosító vagy a természetes személy testi, fiziológiai, genetikai, szellemi, gazdasági, kulturális vagy szociális azonosságára vonatkozó, egy vagy több tényező

⁴⁹ GDPR (26) preambulumbek.

⁵⁰ Czapári Dóra – Szóke Gergely László: Az adatvédelem és az adathasznosítás egyik kulcskérdése: a személyes adatok anonimizálása, JURA, 2022/4. 26. o. Az anonimizálás komplex értelmezését ld. ugyanitt.

személyekre vonatkoznak, és mint ilyenek a védelmüket az Unió és hazai jog alapvető emberi jogként nevesíti. Ellentétben tehát, pl. közadatokkal ezek az adatok explicite, a jog erejénél fogva védettek. A személyes adatok terminológiai meghatározottsága kevésbé fragmentált, elsődlegesen – Uniói rendelet jellegéből adódó alkalmazási elsőbbsége révén – a GDPR határozza meg, míg az Info.tv. ezt kiegészítheti, pontosíthatja. Az egyéb (frissebben megalkotott) jogszabályok ezeket konzervens hivatkozzák, így a hazai és az Uniói jogrendszerben nem tapasztalható zavar a személyes adatok fogalmát illetően (megjegyzendő, hogy tág jellege miatt a fogalom alkalmazásával kapcsolatban viszont tapasztalható).

A személyes adatok belső halmazát alkotja a különleges adatok kategóriája, amelyek érzékenyséjük miatt megkülönböztetett védelemben részesülnek. Különleges adatnak minősülnek a faji vagy etnikai származásra, politikai véleményre, vallási vagy világnézeti meggyőződésre vagy szakszervezeti tagságra utaló személyes adatok, valamint a genetikai adatok, a természetes személyek egyedi azonosítását célzó biometrikus adatok, az egészségügyi adatok és a természetes személyek szexuális életére vagy szexuális irányultságára vonatkozó személyes adatok.⁵¹ Fontos továbbá, hogy szintén speciális megítélésűek a bünyügyi személyes adatok, amelyek a büntetőeljárás során vagy azt megelőzően a bűncselekménnyel vagy a büntetőeljárással összefüggésben, a büntetőeljárás lefolytatására, illetve a bűncselekmények felderítésére jogosult szerveknél, továbbá a büntetés-végrehajtás szervezeténél keletkezett, az érintettel kapcsolatba hozható, valamint a büntetett előéletre vonatkozó személyes adatok.⁵²

A különleges adatok fent felsorolt elemei közül a genetikai, a biometrikus és az egészségügyi adatok külön definiálásra kerültek. Az egészségügyi adat egy természetes személy testi vagy pszichikai egészségi állapotára vonatkozó személyes adat, ideértve a természetes személy számára nyújtott egészségügyi szolgáltatásokra vonatkozó olyan adatot is, amely információt hordoz a természetes személy egészségi állapotáról,⁵³ biometrikus adat⁵⁴ egy természetes személy testi, fiziológiai vagy viselkedési jellemzőire vonatkozó minden olyan sajátos technikai eljárásokkal nyert személyes adat, amely lehetővé teszi vagy megerősíti a természetes személy egyedi azonosítását,⁵⁵ valamint genetikai adat⁵⁶ egy természetes személy örökölt vagy szerzett genetikai jellemzőire vonatkozó minden olyan személyes adat, amely az

⁵¹ GDPR 9. cikk (1) bekezdés; Info.tv. 3. § 3. pont

⁵² Info.tv. 3. § 4. pont

⁵³ GDPR 4. cikk 15. pont; Info.tv. 3. § 3c. pont

⁵⁴ Ilyen például az arckép vagy a daktiloszkópiái adat

⁵⁵ GDPR 4. cikk 14. pont; Info.tv. 3. § 3b. pont; vö. az Európai Parlament és a Tanács 2021/206 rendelete a mesterséges intelligenciára vonatkozó szabályok (továbbiakban: MI rendelet) megállapításáról és egyes uniós jogalkotási aktusok módosításáról 3. cikk 33. pont

⁵⁶ A humángenetikai adatok védelméről, a humángenetikai vizsgálatok és kutatások, valamint a biobankok működésének szabályairól szóló 2008. évi XXI. törvény 3. § (1) bekezdés 4-6 pontjai a kódolt, pszeudonimizált és anonimizált genetikai adat fogalmakat vezetik be, a nélkül, hogy a genetikai adatot külön definiálnák

adott személy fiziológiájára vagy egészségi állapotára vonatkozó egyedi információt hordoz, és amely elsősorban az említett természetes személyből vett biológiai minta elemzéséből ered.⁵⁷ A személyes adatok fogalmával kapcsolatban az látható, hogy a jogalkotó az általános meghatározásokon túl az egyedi, a kor kihívásaira reagáló speciális fogalmakat is törekszik megalkotni.

Az adatok vizsgálata horizontális spektrumon azt jelenti, hogy azok valahol metszik függőleges tengely által elválasztott személyes és nem-személyes adatok kategóriáit. Így az alábbi adattípusok között egyedileg eltérő valószínűséggel, de potenciálisan bárhol találhatunk személyes és nem személyes adatokat is.

3.2.2.3. A közadatok

A közadatok jelentős része jellemzően a nem-személyes adatok kategóriájába tartozik, ez azonban (pl. közérdekből nyilvános adatok esetén) nem mindig igaz. Ezek fogalmát *expressis verbis* a magyar jog adja meg, de a DGA (24) preambulumbekzdése is rögzíti a „nem személyes közadatok” létezését. Saját jogunk szerint⁵⁸ a közadatokhoz tartoznak a digitális formában létező közérdekű adatok és közérdekből nyilvános adatok, valamint a kulturális közadat és a kutatási adatok. A közadatokhoz tartozó egyes típusok – a kulturális közadat és a kutatási adat – fogalmai a főfogalmat megalkotó az idézett törvényben olvashatók.⁵⁹ Előbbi érezhetően a muzeális és közművelődési intézmények, valamint könyvtárak, illetve köz- és magánlevéltárak továbbá az előadó-művészettel, mozgóképekkel kapcsolatos közfeladatot ellátó szervek kezelésében lévő, kulturális javakra vonatkozó digitális adatokat foglalja magába.⁶⁰ Utóbbi a digitális formátumú, közfinanszírozású tudományos munkásság eredményeként létrejött vagy az ilyen eredmény bizonyítékául nyilvánosságra hozott – tudományos publikációnak nem minősülő – adatokat foglalja magába.⁶¹ A közadatok másik altípusai, az információszabadság⁶² jegyében az Info.tv-ben találhatóak. Ezek szerint a közérdekű adatok az állami vagy helyi önkormányzati feladatot, valamint jogszabályban meghatározott egyéb közfeladatot ellátó szerv vagy személy kezelésében lévő és tevékenységére vonatkozó vagy közfeladatának ellátásával összefüggésben keletkezett adatok, míg a közérdekből nyilvános adatok azok a nem

⁵⁷ GDPR 4. cikk 13. pont; Info.tv. 3. § 3a. pont

⁵⁸ A nemzeti adatvagyon hasznosításának rendszeréről és az egyes szolgáltatásokról 2023. évi CI. törvény (továbbiakban: új közadathasznosítási törvény) 2. § 15. pont

⁵⁹ Az új közadathasznosítási törvény szabályozási rezsimje és fogalomhasználata továbbra is a nyílt hozzáférésű adatokról és a közzféra információinak további felhasználásáról szóló 2019/1024/EU irányelv előírásaira épít

⁶⁰ Új közadathasznosítási törvény 2. § 19. pont

⁶¹ Új közadathasznosítási törvény 2. § 20. pont; 2019/1024/EU irányelv 2. cikk 9. pont

⁶² A közügyek átláthatóságához, a közérdekű és a közérdekből nyilvános adatok megismeréséhez és terjesztéséhez fűződő kommunikációs alapjog

közérdekű adatok, amelyeknek a nyilvánosságra hozatalát törvény ettől függetlenül megköveteli.⁶³

A közadatok kiemelt jelentősége az adathasznosítás lehetőségében rejlő potenciál – ahogy az új közadathasznosítási törvény közvetlen elődjének 2012-es megalkotási ideje is mutatja⁶⁴ – korai felismerésében és jogalkotói deklarációjában rejlik. A fogalmak fenti rövid ismertetése alapján látható, hogy a közadatok az állami és önkormányzati igazgatás működése során vagy azzal kapcsolatban létrejövő, egyébként is Alaptörvényi szinten⁶⁵ nyilvánosnak minősített adatok. A két idézett jogszabály a közadatokhoz való hozzáállás két állapotát is megjelöli, míg az Info.tv. klasszikusan az információszabadság általi megismerhetőséget és közéleti átláthatóságot várja el,⁶⁶ addig az új közadathasznosítási törvény az ilyen adatok⁶⁷ további hasznosításának és újrahasznosításának a lehetőségét teremtette meg.⁶⁸ A két szabályanyag együttesen a teljes nemzeti adatvagyon hasznosítását teszi lehetővé. A folyamat támogatására Magyarországon a Nemzeti Adatvagyon Ügynökség (NAVÜ) hivatott, amely a Kormány által rendeletben kijelölt, a nemzeti adatvagyon hasznosításával kapcsolatos állami feladatokat közfeladatként ellátó szervezet.⁶⁹

3.2.2.4. *Az üzleti adatok*

A közadatok alapvetően a közzsféra által kezelt adatokat fedi le, de jelentős potenciál rejtőzhet az üzleti (vállalati és ipari) szféra által kezelt adatokban is. Ez egy igen heterogén adatkört ölel fel, amelyet egyes esetekben „magánadatnak” is hívnak,⁷⁰ az Európai Unió adatstratégiája „magánkézben lévő adatoknak” nevezi őket⁷¹. Ezek egy része természetesen személyes adat, és ennek megfelelő kezelése indokolt, egy másik része azonban nem személyes adat, akár azért, mert anonimizált adat, akár azért, mert eleve nem személyes adatként jött létre. Ezek hasznosításában igen

⁶³ Info.tv. 3. § 5. és 6. pontok

⁶⁴ Lásd: a közadatok újrahasznosításáról szóló 2012. évi LXIII. törvény. Megjegyzendő továbbá, hogy már 2003-ban létezett az Unióban a közzsféra információinak további felhasználásáról szóló jogszabály (az Európai Parlament és a Tanács 2003/98/EK irányelve)

⁶⁵ Magyarország Alaptörvénye VI. cikk (3) bekezdés

⁶⁶ Ld. Hohmann, Balázs: Interpretation of the Concept of Transparency in the Strategic and Legislative Documents of Major Intergovernmental Organizations. Közigazgatási és Infokommunikációs Jogi PhD Tanulmányok 2(1), 2021, 48-56. o.

⁶⁷ Például meteorológiai adatok, különböző állami nyilvántartások adatai

⁶⁸ A 2012. évi LXIII. törvény indokolása a közadatok újrahasznosításáról, Általános indokolás

⁶⁹ A nemzeti adatvagyonról szóló 2021. évi XCI. törvény 3. §. A törvény 2. § a) pontja szerint a nemzeti adatvagyonhoz tartoznak a közfeladatot ellátó szervek által kezelt közérdekű adatok, közérdekből nyilvános adatok és kutatási adatok, továbbá a személyes adatok összessége.

⁷⁰ Fehér Könyv a nemzeti adatpolitikáról, Fehér Könyv a nemzeti adatpolitikáról, Budapest, Nemzeti Hírközlési és Informatikai Tanács Szakértői Tanácsadó Testülete, Budapest, 2016, https://www.magyar.hu/wp-content/uploads/2019/09/Adatpolitikai_feher_konyv_201608.pdf (2024.04.30), 12. o.

⁷¹ A Bizottság Közleménye az Európai Parlamentnek, a Tanácsnak, az Európai Gazdasági és Szociális Bizottságnak és a Régiók Bizottságának. Európai adatstratégia (COM/2020/66 final), 8. o.

jelentős potenciált lát az Európai Unió,⁷² még akkor is, ha ezen adatok egy része üzleti titokként⁷³ védett.

A jogi adatfogalommal kapcsolatban természetesen szükséges megemlíteni, hogy a jog ezeken felül számos adat- és titoktípust nevesít, amelyek potenciálisan szinte bármelyik fenti kategóriát érinthetik,⁷⁴ és létezhetnek olyan adatok is, amelyek a jog által egyáltalán nem szabályozottak.

IV. Az adat értékessége

Egy adat értéke alapvetően valamilyen szükséglet vagy igény kielégítési képességében rejlik.⁷⁵ Ezek az igények/szükségletek lehetnek egyéni és kollektív jellegűek egyaránt. A korábbi Közadattv.⁷⁶ 4. § 14. pontjának fogalom meghatározása – bár kifejezetten a közadatokra vonatkozik – továbbra is segítségül hívható az adat értékesség mibenlétének megragadásához. E szerint nagy értékű az az adat, amely újrahasznosítása fontos előnyökkel jár a társadalomra, a környezetre és a gazdaságra nézve, különös figyelemmel arra, hogy lehetővé teszi hozzáadott értéket képviselő szolgáltatások, alkalmazások, továbbá új, magasabb hozzáadott értéket képviselő munkahelyek létrehozását, valamint figyelemmel az adaton alapuló, hozzáadott értéket képviselő szolgáltatások és alkalmazások lehetséges felhasználóinak számára. A fogalom tehát egy kollektív jellegű értékességet, kvázi társadalmi hasznosságot definiál, azonban látható, hogy mik azok az elemek, amiktől az adat maga értékessé válik: az adaton végzett műveletek által ezen alapuló új szolgáltatások, alkalmazások és tevékenységek hozhatók létre, amelyek hozzáadott értéket képviselnek valamely szereplő a szereplői érdekköre (pl. ügyfelek) számára.

Az adatok értékességének változására kézzel fogható példa lehet egy elkészült termelési statisztika egy takarmánykeverő üzem havi teljesítményéről. Ez értékes biztosan a szervezet – és konkurenciája – számára az erőforrásai helyes allokációjának megvalósításához. Azonban általánosságban kijelenthető, hogy az érték fokozatosan csökken ahogy távolodunk az agráragazattól, így még hasznos lehet a kereskedelmi szektor számára, de egy kórháznak nagy valószínűséggel, ebben a formában nem. Az adat értéke tehát nagy mértékben szubjektív alapokon nyugszik, azt ezért mindig az egyedi eset körülményeire tekintettel kell megvizsgálni. Ráadásul – ahogy fentebb említettem – az értékítélet meghozatala elválhat az adat gazdájától, illetve az adatot előállító, azon elsődleges műveleteket végzőktől.⁷⁷ Az előbbi kijelentést árnyalja az adat mennyiségének hatása a teljes adatállomány

⁷² Az Európai Unió adatstratégiájának egyik pillére ezen adatkör megosztását igyekszik bátorítani, és ennek jegyében indult el a ma adatrendelet néven ismert jogszabály [Az Európai Parlament és a Tanács (EU) 2023/2854 Rendelete]. Ld. Európai adatstratégia, 15-16. o.

⁷³ Az üzleti titok védelméről szóló 2018. évi LIV. törvény 1. § (1) bekezdés

⁷⁴ A titoktípusok kategorizálásáról ld. Szőke Gergely László: Gondolatok a hazai titokvédelmi szabályozás rendszeréről, JURA, 2018/2., 241-257. o.

⁷⁵ Corwin D. Edward: The Meaning Of Quality. *Quality Progress*, 1(10), 1968. 36-39 o.

⁷⁶ A közadatok újrahasznosításáról szóló 2012. évi LXIII. törvény

⁷⁷ Anya Bernstein: What Counts as Data?. *Brooklyn Law Review*, 86(2), 2021. 435-460. o.

értékére. Ha nem egy takarmánykeverő üzem, hanem Európa összes üzemének kimutatásai kerülnek összesítésre és rendelkezünk a feldolgozáshoz szükséges informatikai számítási képességgel akkor az adat értéke mennyiségével és kiterjedtségével potenciális növekszik.⁷⁸ Ez igaz lehet függetlenül attól, hogy a benne levő „zaj” is egyre nagyobb. Így bár a szó szerinti értelemben vett adatminőség romlik, magának az adatállománynak az értéke nő, mert több szükséglet kielégítésére lehet alkalmas.⁷⁹

Az adatok értékük szerinti tipizálása szinte lehetetlen, hiszen ahány szereplő annyi kielégíthető igény. Mégis egy rövid, példalózó felsorolással járulnék hozzá a terület kontextusba helyezésének megkönnyítéséhez. Fontos megjegyezni, hogy az adat értékét erkölcsi szempontok figyelembe vétele nélkül, objektív alapokon célszerű vizsgálni. Tehát attól például, hogy nagy mennyiségű adat hasznosításával országos választások kimenetelének mesterséges megváltoztatására is lehetőség van, nem csökkenti az adat értékét.⁸⁰

1. táblázat – Az adat által képviselt érték és mérésének lehetőségei.

Forrás: saját szerk.

Adat által képviselt érték	Az érték mérésének lehetősége
Politika: az államvezetést ellátó politikai vezetés számára egyértelműen hasznos a választópolgárok jobb ismerete, döntéseik előrejelzése, befolyásolásának lehetősége ⁸¹	Az érték a támogatottságot vizsgáló közvéleménykutatások, állami statisztikák, illetve egy-egy választás eredményének számbevételével mérhető.
Közbiztonság: különböző bűncselekmények elkövetési adatainak rögzítése és elemzése hozzájárulhat jobb preventációs eszközök és módszerek bevezetéséhez. ⁸²	Az érték a bűnözési adatok változásával mérhető.

⁷⁸ Maryam Farboodi, Laura Veldkamp: A Model of the Data Economy. *NBER Working Paper No. 28427*, 2021. <https://doi.org/10.3386/w28427>

⁷⁹ Jingran Wang, et. al.: Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *Journal of the Knowledge Economy*, 2023. <https://doi.org/10.1007/s13132-022-01096-6>

⁸⁰ Julia Carrie Wong: The Cambridge Analytica scandal changed the world – but it didn't change Facebook. *The Guardian*, 2019. <https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook> (2024.04.30.)

⁸¹ Zódi Zsolt: *Platformok, robotok és a jog*. Budapest, Gondolat Kiadó, 2018. 76. o.

⁸² Algorithm Watch: *Automating Society Report* 2020. <https://automatingsociety.algorithmwatch.org/report2020/switzerland/switzerland-story/> (2024.04.30.)

Adat által képviselt érték	Az érték mérésének lehetősége
<p>Egészségügy: betegadatok hasznosításával az gyógykezelési képességek gyorsabb tempóban fejlődhetnek, új eszközök, módszerek és eljárások alakíthatók ki, új típusú gyógyszerek hozhatók létre.⁸³</p>	<p>A korábban nehezen vagy egyáltalán nem gyógyítható betegség túlélési esélyeinek javításával, akut vagy krónikus betegségeknek a betegre gyakorolt negatív hatásainak csökkentésével egyértelműen kimutatható a felhasznált adatok értéke.</p>
<p>Üzemi/vállalati/szolgáltatási hatékonyság és produktivitás: Mind termelő, mind a szolgáltató szektorban alapvető fontosságú, hogy egységnyi idő alatt a lehető legnagyobb eredményt lehessen elérni. Ez a kijelentés a 2020 óta eltelt években különösen igaznak bizonyul, a sorozatos válsághelyzetek miatt.⁸⁴ A saját és más szervezetek működési adatainak, valamint a fogyasztók vagy ügyfelek elégedettségének, szokásainak elemzése segítheti a köz- és a magánszféra szervezeteinek hatékony működését.</p>	<p>A racionalizált folyamat-megoldások – megfelelően hasznosítva – csökkenthetik a környezetterhelést, magasabb gazdasági hozzáadott-értéket termelhetnek, miközben a munkameneteket biztonságosabbá és költséghatékonyabbá tehetik. A magasabb szintű, ügyfélbarát szemlélet a hazai és a világ közigazgatási rendszereit is áthatja, így a közpénzek átgondolt költsége mellett magasabb közszolgáltatási színvonal is elérhető lehet. Ezek az eredmények igazolnák az ilyen típusú adatok értékét.</p>
<p>Piaci pozíció: a piac megszerzése és megtartása folyamatos és komoly kihívás a gazdasági szereplőknek a szabadversenyre épülő rendszerekben. A marketing és PR törekvések fogyasztói és felhasználói adatokra épülő megtervezése stabilabb és kiszámíthatóbb működést tehet lehetővé.⁸⁵</p>	<p>Mind a marketing, mind a PR terület már jelenleg is hasznélvezője az adathasznosításnak, a megbízható analitikai adatok értéke e tekintetben igazoltnak tekinthető.</p>

⁸³ Sastry Chilukuri: *The role of big data in medicine*. McKinsey & Company, 2015. [https://www.mckinsey.com/industries/life-sciences/our-insights/the-role-of-big-data-in-medicine#/ \(2024.04.30.\)](https://www.mckinsey.com/industries/life-sciences/our-insights/the-role-of-big-data-in-medicine#/ (2024.04.30.))

⁸⁴ SARTORIUS: How Manufacturers Are Using Big Data Analytics to Improve Processes. 2019. [https://www.sartorius.com/en/knowledge/science-snippets/how-manufacturers-are-using-big-data-analytics-to-improve-processes-507146 \(2024.04.30.\)](https://www.sartorius.com/en/knowledge/science-snippets/how-manufacturers-are-using-big-data-analytics-to-improve-processes-507146 (2024.04.30.))

⁸⁵ Chase Bibby, Jonathan Gordon, Gustavo Schuler, Eli Stein: The big reset: Data-driven marketing in the next normal”, McKinsey & Company, 2021. [https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-big-reset-data-driven-marketing-in-the-next-normal \(2024.04.30.\)](https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-big-reset-data-driven-marketing-in-the-next-normal (2024.04.30.))

Adat által képviselt érték	Az érték mérésének lehetősége
<p>Mesterséges intelligencia fejlesztés: a magasan fejlett algoritmizált folyamatok (MI) fejlesztésének elengedhetetlen eleme megfelelő mennyiségű és minőségű adat rendelkezésre állása.</p>	<p>Az egyre kockázatosabbnak tartott,⁸⁶ de mégis vágyott általános mesterséges intelligenciának⁸⁷ a létrehozása hatalmas előrelépést jelentene az augmentációs és automatizációs folyamatokban,⁸⁸ ami a civilizált élet minden területére hatással lenne.</p> <p>Az MI kutatás jelenleg látható végpontja – a még kissé sci-fi hangzatú – un. MI szingularitás elérése, vagyis az emberi mentális képességeket meghaladó mesterséges szuperintelligencia kifejlesztése szintén elképzelhetetlen nagy mennyiségű adat hasznosítása nélkül.⁸⁹</p>
<p>Adat-kommodifikáció: az adatok gazdasági értékkel rendelkeznek, jogügyeletek tárgyai lehetnek. Ilyenkor az adat tényleges, piaci értéke kerül beárazásra, nem feltétlenül a hasznosítása által keletkező közvetett értékkeremtő hatás.</p>	<p>Az érték egyszerűen, pénzben kifejezhető.</p>

A fentiek során az adathasznosításra tekintettel, de az adatok fogalmi rendszere általános értelemben került vizsgálatra a terület biztos alapokra helyezése érdekében. Az adatok absztrakt, technológiai, jogi és érték szempontjából történő vizsgálatával és a szakirodalom felhasználásával a következő fogalmat tudjuk alapul venni az adathasznosítás megragadásához: az adat a világ tényszerűségeinek olyan reprezentációja, amelyen keresztül – értéke függvényében – információ szerezhető a világról, ezáltal az ismeretek bővíthetők, a tudás gyarapítható.

⁸⁶ Jyoti Narayan, Krystal Hu, Martin Coulter, Supantha Mukherjee: Elon Musk and others urge AI pause, citing 'risks to society'. Reuters, 2023. <https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/> (2024.04.30.)

⁸⁷ IBM: What is strong AI?. <https://www.ibm.com/topics/strong-ai> (2024.04.30.)

⁸⁸ Az emberi munka támogatására vagy kiváltására irányuló folyamatok

⁸⁹ Rahul Rao: What happens if AI grows smarter than humans? The answer worries scientists. Popular Science. <https://www.popsci.com/science/ai-singularity/> (2024.04.30.)

Általános értelemben ennek a megismerési tevékenységnek a felgyorsítását és mélyítését nevezhetjük adathasznosításnak.⁹⁰ Modern körülmények között ez a megismerés egyre inkább a digitális térbe helyeződik át, az adatrögzítési, tárolási, feldolgozási (különösen algoritmizációs) és továbbítási kapacitások fejlődése által.

V. Adattípusok az adathasznosítás folyamatában

Az adatfoglalom meghatározása felé tett lépéseket, egyúttal jelen tanulmányt az eddig feltártak szerint az adatoknak a hasznosítási folyamatban betöltött pozíciója alapján történő csoportosítással zárom:

- **Nyers adat:** ez alapvetően valamilyen, az adathasznosításhoz nem kapcsolódó folyamat során létrejövő adathalmazként írható le, amely így a főfolyamat „melléktermékeként” kerülhet hasznosításra.⁹¹ Létezhet analóg és digitális formában, bár minden bizonnyal utóbbi a jellemzőbb, tekintve, hogy a nyers adatoknak kiemelt elemei az informatikai eszközök működése során automatikusan generált adatok, illetve a szenzoros adatok. Ezek mellett ide tartozhatnak a mindennapi élet bármely eseménye/cselekménye során keletkező számtalan adat, függetlenül annak jogi megítéléstől (pl. szerződéskötés, foglalkoztatás, köznevelés stb.). A nyers adatokat – a járulékoságuk miatt – másodlagos hasznosíthatóság jellemzi, tehát valamilyen további feldolgozási művelet nélkül nem hasznosíthatók.
- **Termelt adat:**⁹² az adatok gyűjtése és rögzítése lehet kifejezetten az adat megszerzésére és hasznosítására irányuló folyamat. Ebben az esetben tehát nincs egy alapul szolgáló ügyviteli esemény (pl. autó adásvétel), aminek a velejárója adatok rögzítése, az adatrögzítés azért történik, hogy az adat létrejöjjön. Ide sorolhatók az analóg vagy digitális nyilvántartások, személyiség-profilhoz tartozó adatok és az olyan szenzoros adatok, amelyeknek nincs közvetett céljuk sem (pl. egy okos kazán termosztátjának adatrögzítése járulékos a kazán működéséhez, míg egy okos óra lépésszámlálójának elsődleges és egyetlen célja az adat rögzítése). Az ilyen adatok tehát elsődleges hasznosítás céljára szolgálnak, ennek megfelelően könnyen olvasható és használható formátumban, strukturáltan jönnek létre. Ezek önmagukban is értékesek lehetnek, de további feldolgozással az értékük növelhető.

⁹⁰ Rob Kitchin: *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. New York, SAGE Publications, 2014. 1-26. o. <https://doi.org/10.4135/9781473909472>

⁹¹ Charles I. Jones, Christopher Tonetti: Nonrivalry and the Economics of Data. *American Economic Review* 110(9), 2020. 2819–2858. o. <https://doi.org/10.1257/aer.20191330>

⁹² Kyu Yub Lee, Hyun Park: *Data, Privacy, and Artificial Intelligence*. SSRN Electronic Journal, 2022. DOI: <https://doi.org/10.2139/ssrn.4246578>

- **Finomított adat:** olyan adatok tartoznak ide, amelyekben már a hasznosításhoz szükséges valamely művelet vagy műveletek elvégzésére kerültek, az értékteremtés/értéknövelés elkezdődött/megtörtént. A műveletek teljes mértékben attól függenek, hogy milyen területen, milyen célok szerint, milyen adatokkal kapcsolatban kerülnek végrehajtásra. A „finomítottság foka” természetesen sokféle lehet, az elvárt minőséget a résztvevők jellege határozhatja meg (még egy adatokkal foglalkozó szakember/vállalkozás alacsonyabb feldolgozottsági szintű adatokat is értékesnek találhat, addig egy „végfelhasználó” mindössze eshetőlegesen meglévő szaktudása függvényében lehet rá csak képes).
- **Adattermék:** az adat végállapotát, ahogy fentebb kifejtésre került, a megfelelő mélységű és minőségű feldolgozottsággal és egy adott szükséglet kielégítésére alkalmas egységességgel érhetjük el. Ez a megfelelő adatoknak, a megfelelő helyre/személyhez történő eljuttatását, és szükség szerint az adott helyzetre interpretálását vagy a kérdéses folyamatba integrálását jelenti. Az adattermék tehát egyedi esetben értékes, egy konkrét szereplő számára hasznos adatot jelöl, így az megfelelő működés esetén a végfelhasználó számára információnak minősül, míg az adattermék szolgáltatója szempontjából megőrzi adat jellegét. Fontos azonban, hogy az adattermék létrejöttével az adathasznosítás folyamata véget ér, a következő szakaszokra már eltérő szabályok és gyakorlatok vonatkoznak (pl. információ- és tudásmenedzsment).

VI. Összegzés

Az adathasznosítás vizsgálata számos olyan terület feltárását teszi szükségessé, amelyek hosszú időn keresztül kevésbé kidolgozott fogalmakkal operáltak. A megfelelő értelmezési feltételek megléte azonban elengedhetetlen az adatok hasznosításával kapcsolatos jogalkotási és jogalkalmazási kérdések megválaszolásához.

Jelen tanulmány során bemutatásra került az adatfogalom absztrakt, technológiai és jogi szempontból. Kísérletet tettem továbbá a DIKW piramis újragondolására, annak érdekében, hogy a modern – főként algoritmizált – életviszonyokhoz jobban adaptálhatóvá váljon. Az eredeti modell terminológiáján⁹³ eszközölt legnagyobb változtatások az információ, mint kategória kivezetése (pontosabban az adatba építése), valamint két további elem, a tényyszerűség és az ismeret bevezetése, továbbá az ismeret és tudás egymástól való elhatárolása. Az újragondolt modellel kapcsolatban abban bízom, hogy a mai folyamatok abszolút homlokterében lévő algoritmizáció jellegzetességeire jobban interpretálható lesz. Az

⁹³ Saša Baškarada, Andy Koronios: Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. *Australasian Journal of Information Systems*, 18(1), 2013. 5-24. o. <https://doi.org/10.3127/ajis.v18i1.748>

adathasznosítás, mint tevékenység létjogosultságának alátámasztásához elengedhetetlen értékalapú vizsgálatra – legalábbis annak szempontjainak felvázolására – is sor került. Végezetül a szakirodalom és a terület tapasztalat-alapú vizsgálatának felhasználásával az adathasznosítás szempontjából elsődlegesen alkalmazható adat-fogalomrendszer is felállításra került. Ezen belül a nyers adatokat, a termelt adatokat, a finomított adatokat és végül magát az adatterméket határoltam el. Az újra-gondolt DIKW piramis szemléletéhez hasonlóan itt is azonosítható az algoritmusok működési mechanizmusaiból ismert három fázis, vagyis a bemenet, a számítási teljesítmény és a kimenet. Látható, hogy a nyers és termelt adatok fogják bemenetet biztosítani, míg a feldolgozottság fokától függően a kimenet lehet a finomított adat és a folyamat végállapotában létrejöhet egy adattermék.

Bízom benne, hogy jelen tanulmány során a fogalmak tisztázására tett törekvéseim megkönnyítik a terület további kutatását, illetve a kutatási eredmények hatékonyabb befogadását.