

ÖSSZEFOGLALÓ

Az „adat” kifejezés rendkívüli népszerűségnek örvend a mai tudományos és „köznap” diskurzusban egyaránt és ez alól nem jelent kivételt a munka világa sem. Az adat fogalom azonban ritkán kerül kellő komplexitással kifejtésre, aminek következtében egy – a szükségesnél nagyobb – értelmezési nehézség alakul ki rögtön a megismerési folyamat legelején. Az értelmezési nehézség feloldására feltétlenül szükség van a helyes adatkezelési – és egyébként az esetleges adathasznosítási – gyakorlatok kialakításához egyaránt.

Az absztrakt adatfogalom és a DIKW-piramis újragondolása

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 ún. DIKW¹ piramis vagy hierarchia modell² szolgál, amelyet azonban olyan módon alakítottam át,³ hogy az jobban illeszkedjen az adatok hasznosításának témaköréhez (az eredeti DIKW modell terminológiáján⁴ eszközölt legnagyobb változtatások az információ, mint kategória kivezetése, valamint két további elem, a tényyszerűség és az ismeret bevezetése, illetve az ismeret és tudás egymástól való elhatárolása).

1. A folyamat első lépéseként szükség van arra, hogy a körülöttünk lévő valóságban jelenségek létezzenek (mint természeti körülmények, emberi cselekedetek), 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 létezésére. Ennek a szintnek a kiemelése az első eltérés az eredeti modellhez képest, ezt pedig összefoglalóan **tényszerűségeknek** nevezhetjük.⁵ Fontos tulajdonságuk, hogy állandóan és szemléltőlől függetlenül vagy legfeljebb azzal kölcsönhatásban (az ún. megfigyelési effektus

* A tanulmány a 2020-1.1.2-PIACI-KFI-2021-00306 Okos Révkalauz HUB projekt keretében készült.

¹ DIKW akronim jelentése: Data – Information – Knowledge – Wisdom

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

⁴ 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*, Volume 18 Number1, 2013.

⁵ A magyar nyelv értelmező szótára (www.mek.oszk.hu)

- által [Observer Effect]⁶⁾ léteznek. Ez alapján tehát a tényszerűséggel ok-okozati összefüggésben bármi állhat, de a folyamat lezárását követően azon változtatni nem lehetséges, a valóságunk elválaszthatatlan részévé válik.
2. A következő szint maga a kutatott fogalom, az **adat**, amely a tényszerűséget leíró alapvető paramétereket foglalja magába. Így tehát az adat valójában egy átmeneti állapotot, egy potenciált jelent, amely – Schrödinger macskájához hasonlóan⁷ – vagy manifesztálódik a szemlélő jelenléte által, vagy nem. Az minden esetben objektív, azonban önmagában csak egy nyers értéket hordoz.⁸ A modellben a soron következő elem az **információ** lenne, ami az értelmezhető, hasznosítható (értéket hordozó) adatot jelenti.⁹ Az információ hasznos és értékes kvalitásainak kiemelése helyénvaló, de a hierarchiában történő szerepeltetéssel kapcsolatban az álláspontom az, hogy az algoritmizált folyamatok térnyerése által, a fogalmak egyre inkább összemosódnak, ezáltal kiüresítve az adat és az információ ilyen kategorikus elválasztásának szükségességét. 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, a két fogalom nem zárja ki egymást, mindössze az információ többlet jelentéssel (és értékkel) bír az adathoz képest. Tehát jelen értelmezés szerint minden információ adat, de nem minden adat információ.
 3. Az adatok rögzülését követően, azok alapul szolgálhatnak **ismeretek** megszerzésére, amik valamilyen fajta, megismerő tevékenység eredményeként, az ennek segítségével kialakított, a valóságra, valamint annak valamely területére vonatkozó fogalmak összességéként foghatunk fel.¹⁰ Fogalmazhatunk tehát úgy, hogy az adat rögzítettsége által képessé teszi a szemlélőt tapasztalatok megszerzésére, vizsgálatok lefolytatására. Ez által újabb tényszerűségek fedezhetőek fel, amelyek további adatok rögzítését teszik lehetővé, ezért az ismeret az adat feldolgozása során létrejövő értelemmel is azonosítható. Munkaügyi példán értelmezve ilyen ismeret lehet a teljesítmény-értékelés eredménye és azok további adatokkal való összekapcsolása.¹¹
 4. Az előző szintre alapozva, valamilyen ismeret birtokát megszerezni, rendszerbe szedni, összekapcsolni más ismeretekkel és ezekből következtetéseket levonni magasabb szintű **tudást** fog eredményezni a

⁶ Paul Adrien Maurice Dirac: The Principles of Quantum Mechanics. Clarendon Press, 1981-314 o.

⁷ John D. Trimmer, „The present situation in quantum mechanics: a translation of schrödinger's "cat paradox paper"”, Proceedings of the American Philosophical Society, 124, 323-38.

⁸ Todd Groff, Thomas Jones: Introduction to Knowledge Management. Routledge, London, 2003.

⁹ Kenneth C. Laudon, Jane P. Laudon: Management Information Systems: Managing the Digital Firm, Global Edition. Pearson Higher Ed, 2021.

¹⁰ A magyar nyelv értelmező szótára (www.mek.oszk.hu)

¹¹ Ld. Hohmann Balázs: Hol a határ? - A munkáltatói teljesítményértékelési rendszerek adatvédelmi vonatkozásai. In: Pál, Lajos; Petrovics, Zoltán (szerk.) Visegrád 16.0 A XVI. Magyar Munkajogi Konferencia szerkesztett előadásai. Budapest, Wolters Kluwer, 2019. 317-327. o.

tényszerűséggel kapcsolatban.¹² A tudás felhasználásával az értelem elhelyezhető a valóságban.

5. Az utolsó szintje a piramisnak, az értelem – az ismeret és tapasztalat együttesen – valóságnak megfelelő alkalmazása, amit **bölcsességnek** nevezhetünk.¹³

A hierarchia öt pontja alapján látható, hogy a legegyszerűsebbtől, 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 tényszerűségeinek megismerése.

Megállapítható az is, hogy egy tényszerűség egy aspektusa egy adott pillanatban csak egyféle adat leképezését teszi lehetővé, az adatnak a szemlélő 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.

Technológiai adatfoglalom

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 tényszerűségek 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 végezhető”, ezt követően 1954-ben jelentek meg az adatfeldolgozás és 1962-ben az adatbázis szóösszetételek.¹⁴ Az adat technológiai szempontból alapvetően két nagy kategóriái osztható, technológiai és nem technológiai jellegű (organikus) adatokra. Mindkettő a DIKW modell absztrakt megközelítésének megfelelően valamilyen tényszerűséget leíró, információt hordozó paramétert jelent, de előbbi kifejezetten az emberi technológiához köthető módon kezelhető, utóbbi azt feltétlenül nélkülözi. Organikus adatok között említhetők az épített és természetes környezetet, valamint élőlényeket jellemző, észlelhető tulajdonságok (pl. fizikai jellemzők), de ide tartozhatnak (függetlenül adatvédelmi jogi relevanciájuktól) a biológiai mintákban levő genetikai adatok¹⁵ vagy például a tenyerünkön hordozott biometrikus adatok is. A technológiai jellegű adatokat szintén tovább bonthatjuk analóg és digitális adatokra, amely

¹² David Boddy, Albert Boonstra, Graham Kennedy: *Managing Information Systems: Strategy and Organisation*. Pearson Education, 2008.

¹³ Leonard M Jessup, Joseph S Valacich: *Information systems today: managing in the digital world*. Upper Saddle River, N.J.: Pearson Prentice Hall, 2008.

¹⁴ Etymonline, „Online Etymology Dictionary” Harper Family LLC. Elérhető: <https://www.etymonline.com/word/data>, Letöltve: 2024.01.14.

¹⁵ Worku Gedefa Urgessa, „The feasibility of applying eu data protection law to biological materials: challenging data as exclusively informational” *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 7(2), (2016). 96-109 o.

megkülönböztetés alapja a rögzítéshez felhasznált technológia, valamint az adathordozó típusa. Így analóg adatnak tekinthetjük a papír alapú iratanyagot, míg digitális adatnak az informatikai eszközön tárolt adatot. Tekintve, hogy a ma dominánsan jelen lévő mód algoritmizált adathasznosítás és ez kifejezetten a digitális adatokkal mutat szoros kapcsolatot, továbbá a szabályozási kérdések fókuszában is ez áll,¹⁶ ezzel az adattípussal foglalkozok részletesebben.

A digitális adatok alkategóriája

A digitális adatok főszabály szerint megegyeznek fogalmilag az adat általános fogalmával, azzal az eltéréssel, hogy valamilyen elektronikus módszer vagy elv szerint épített vagy működő eszközhöz, azzal kapcsolatban vagy azt használva léteznek.¹⁷ A digitális adatok – tehát számítógépek számára használható adatok¹⁸ – így az informatika működési elveinek megfelelően döntően numerikus értékekkel (jellemzően bináris rendszerben) kerülnek létrehozásra, tárolásra és továbbításra.¹⁹ Az adatok technikai jellemzőik és az alkalmazott programnyelv szerint tovább tipizálhatók, attól függően milyen elemekből épülnek fel (így pl. az „Integer” egész számokat jelöl, a „Character” az ábécé betűit jelöli, a „Date” egy naptári dátumot tárol más programozási információkkal együtt, a „Floating Point” pedig tört számokat foglal magába).²⁰ 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.

Digitális adat alatt értünk tehát minden digitális formában létező (rögzített és tárolt) adatot, ami körkörös érvelésnek tűnhet, de mégis a legpontosabb definíciót adja. 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, valamint a mindezekből levont következtetések. Ezek az informatika állása és az alkalmazott keretrendszer szerint sokféle fájlformátumban létezhetnek, ezeknek a változatossága pedig szintén

¹⁶ Bankó, Zoltán; Szőke, Gergely László: Issues of Digital Workplace: The Situation in Hungary. Pécs, JURINFO (2016), 12 o.

¹⁷ Merriam-Webster.com, Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/digital>. Letöltve: 2024.01.14.

¹⁸ Juan Ortiz Freuler, "Datafication, Identity, and the Reorganization of the Category Individual", Temple Law Review, vol. 95, no. 4. 2023. 378 o.

¹⁹ Britannica.com, Encyclopædia Britannica, <https://www.britannica.com/technology/digital-computer> Letöltve: 2024.01.14.

²⁰ Luca Cardelli, Peter Wegner, „On Understanding Types, Data Abstraction, and Polymorphism”, Computing Surveys, Vol 17 n. 4. 1985. 471-522 o.

egyre bővül.²¹ Így szöveg lehet a TXT vagy DOCX, kép a JPG, míg audio a FLAC, a videó pedig az AVI.

A Big Data jelenség

A digitális adatokkal kapcsolatban sokszor kerül megemlítésre az ún. „Big Data” vagy nagy adathalmazok jelensége, amely a világon adott pillanatban, potenciálisan rendelkezésre álló (hatalmas és gyorsan növekvő mennyiségű, változatos tartalmú) összes adatot jelöli. Ezek a nagy adathalmazok többnyire nyers adatként léteznek, tehát azok további feldolgozása szükséges, ahhoz, hogy ténylegesen hasznosíthatóvá váljanak.²² A Big Data mára már kevésbé számít novumnak, hiszen egyes felfogások szerint legalább 2012-óta a Big Data korában élünk,²³ ettől függetlenül az ezzel kapcsolatos tanulságok 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 (verification) alatt az adatforrás megbízhatóságát (annak bizonytalanságát) értjük,
- értékesség (value), ami az adatból kinyerhető hasznosíthatóságát (tehát az adott cél elérésére való alkalmasságot) takarja.

²¹ Adanma Cecilia Eberendu, „Unstructured Data: an overview of the data of Big Data”, International Journal of Emerging Trends & Technology in Computer Science. 38. 2016. DOI: 10.14445/22312803/IJCTT-V38P109

²² Pataki, Gábor; Szőke, Gergely László: Az online személyiségprofilok jelentősége In: Polyák, Gábor (szerk.) Algoritmusok, keresők, közösségi oldalak és a jog: A forgalomirányító szolgáltatások szabályozása. Budapest, Magyarország: HVG-ORAC (2020) 77 o.

²³ Steve Lohr, The Age of Big Data, N.Y. TIMES (2012) Elérhető: <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>, Letöltve: 2024.01.16.

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

²⁵ Amy Affelt, "Big Data, Big Opportunity" Australian Law Librarian, vol. 21, no. 2, 2013. 78-89 o.

²⁶ Petroc Taylor, „Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025” STATISTA, 2023. Elérhető: <https://www.statista.com/statistics/871513/worldwide-data-created/>. Letöltve: 2024.01.16.

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

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 követnek semmilyen sémát. Strukturálatlan 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 között. A web 2.0 adatáramlása ehhez képest horizontálisan épül fel, az adatot létrehozó és a fogyasztó mellérendeltek, sőt a két pozíció jellemzően

²⁸ Adanma Cecilia Eberendu, „Unstructured Data: an overview of the data of Big Data”, *International Journal of Emerging Trends & Technology in Computer Science*. 38. 2016. DOI: 10.14445/22312803/IJCTT-V38P109

²⁹ Alexandra Anghel, Elena Novacescu & Madalina Cuc, "Big Data and Its Secrets: Types of Big Data" *Romanian Intelligence Studies Review* 27. 2022. 54-73 o.

³⁰ Christian Hänig, Martin Schierle, Daniel Trabold, (2011). Benefits of Unstructured Data for Industrial Quality Analysis. In: Sio-long Ao, Mahyar Amouzegar, Burghard Rieger, (szerk.) *Intelligent Automation and Systems Engineering. Lecture Notes in Electrical Engineering*, vol 103. Springer, New York, NY. 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> (2022.08.02.)

³³ Larisa-Antonia Capisizu, "Legal Perspectives on the Internet of Things," *Conferinta Internationala de Drept, Studii Europene si Relatii Internationale* 2018. 523 o.

egymásba is csúszik, tehát az adatot létrehozó egyszerre fogyasztó is és fordítva, a fogyasztó 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 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 generálásában 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 hurok létrehozására, amelyekkel maximalizálni képesek az adatmegosztási készletét.³⁷
- Szintén megemlítendő az ún. cloud computing,³⁸ ami a szükséges hardver és szoftver erőforrás kiszervezését és azokhoz az interneten történő kapcsolódást jelenti.³⁹ A szolgáltatás a gyakorlatban tipikusan külső tárhelyszolgáltatás és/vagy valamilyen szoftver licenz és/vagy adatbázis rendelkezésre bocsátása formájában valósul meg. Attól függően mi a szolgáltatás súlypontja beszélhetünk Infrastruktúra, mint Szolgáltatás (IaaS),⁴⁰ Platform, mint Szolgáltatás (PaaS),⁴¹ Szoftver, mint szolgáltatás (SaaS)⁴² és Funkció, mint Szolgáltatás (FaaS).⁴³ A Cloud Computing egyik nagy eredménye, hogy tőkeigényes beruházások nélkül biztosít magas szintű

³⁴ Darin K. Fox, "What Is Web 2.0", AALL Spectrum, vol. 13, no. 9. 2009.

³⁵ Ulises A. Mejias, Nick Couldry, „Datafication”, Internet Policy Review Journal On Internet Regulation Volume 8 | Issue 4. 2019. DOI: 10.14763/2019.4.1428

³⁶ Clare Southerton: Datafication. Laurie Schintler, Connie McNeely, (Szerk.): Encyclopedia of Big Data. (Springer, 2020) 1-4 o.

³⁷ Trevor Haynes, „Dopamine, Smartphones & You: A battle for your time” Harvard University, Science in the News. Elérhető: <https://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/> Letöltve: 2024.01.16.

³⁸ Magyarul használatosak a „számítási felhő” és a „felhőtechnológia” kifejezések

³⁹ Techopedia: Cloud Computing, <https://www.techopedia.com/definition/2/cloud-computing> (2022.08.02.)

⁴⁰ Informatikai erőforrások külső szolgáltató általi biztosítása

⁴¹ Infrastruktúrán túl, közbenső szoftverek (pl. adatbázis-kezelő rendszerek) külső szolgáltató általi biztosítása

⁴² Kliens-alapú webes alkalmazások igénybevétele külső szolgáltatótól (pl. ERP, CRM rendszerek)

⁴³ Webes fejlesztési környezet biztosítása, külső szolgáltató által

számítási kapacitást és egyéb erőforrásokat gyakorlatilag bárkinek.⁴⁴ A hátulütője egyfelől, hogy a kapacitás elérhetővé válása képessé tett mindenkit összetett és kockázatos adatkezelési műveletek elvégzésére, másfelől a magas adatkoncentrációból eredően adatbiztonsági aggályokat is felvet az igényérvényesítési akadályok mellett, amelyek a jellemzően országhatárokon átnyúló adatmegosztási szisztémából erednek.⁴⁵

Összegezve technológiai szempontból a munkaügyi adatkezelések adattípusai a szervezet fejlettségétől függően megjelenhetnek analóg is digitális adatként egyaránt. A digitalizáció előrehaladtával természetesen egyértelmű eltolódás várható utóbbi felé.

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 fokának 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 adatfogalom 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 Uniós és hazai jogszabályok áttekintésére irányult, ezek közül is elsődlegesen azokra, amelyek leginkább kötődnek a tanulmány témájához.

Az adat főfogalma

Az adat általános érvényű definiálására legmagasabb szinten (és a témával legszorosabb kapcsolatban) viszonylag későn, 2022-ben került sor, az Adatkormányzási rendelet⁴⁶ által. Ez a fogalom megjelenik több másik, az Unió „digitális jogalkotási hullámához” tartozó joganyagban, így a Digitális Piacokról szóló rendeletben⁴⁷ és az

⁴⁴ Google Cloud, What are the different types of cloud computing? Elérhető: <https://cloud.google.com/discover/types-of-cloud-computing> Letöltve: 2024.01.16.

⁴⁵ Bankó Zoltán – Szőke Gergely László: Az információtechnológia hatása a munkavégzésre, (Utilitates Bt., Pécs, 2015) 22 o.

⁴⁶ 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

⁴⁷ 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

Adatrendeletben.⁴⁸ 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”.⁴⁹ A fogalom aktusok, tények és információk megjelenítésére utal, ahol az aktus (tett, cselekedet⁵⁰) 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. Mindettől 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-ben⁵¹ található. Ez alapján adat: „az információ hordozója, a tények, fogalmak vagy utasítások formalizált ábrázolása, amely az emberek vagy automatikus eszközök számára közlésre, megjelenítésre vagy feldolgozásra alkalmas.”⁵² 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 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.”⁵³ 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 tényszerűségek reprezentációs funkciója.

A normatív adatfogalmak első szétágazása a személyes adat és a közadat megkülönböztetéssel valósul meg, mindkettő további alkategóriákra oszlik. Mindkét adatkategória kiemelten fontos, mivel az adathasznosítás középpontjában állnak.

⁴⁸ 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

⁴⁹ DGA 2. cikk 1. pont; DA 2. cikk (1) bekezdés; DMA 2. cikk 24. pont

⁵⁰ Idegen Szavak Gyűjteménye, Elérhető: <https://idegen-szavak.hu/aktus>, Letöltve: 2024.01.18.

⁵¹ Az állami és önkormányzati szervek elektronikus információbiztonságáról szóló 2013. évi L. törvény (Továbbiakban: Ibtv.) 1. § (1) bekezdés 1. pont

⁵² Ibtv. 1. § (1) bekezdés 1. pont

⁵³ Ibtv. 1. § (1) bekezdés 25. pont

A közadatok fogalmát *expressis verbis* a magyar jog adja meg, de a DGA (24) preambulumbekzdése is rögzíti a „nem személyes közadatok” létezését. Saját jogunk szerint⁵⁴ a közadatokhoz tartoznak a 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 – fogalmi nagyrészt a főfogalmat megalkotó Közadattv-ben olvashatók.⁵⁵ 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.⁵⁶ Utóbbi a digitális formátumú, közfinanszírozású tudományos munkásság eredményeként létrejött vagy az ilyen eredmény bizonyítékául nyilvánosságra hozott – tudományos publikációnak nem minősülő – adatokat foglalja magába.⁵⁷ A közadatok másik altípusai, az információszabadság⁵⁸ jegyében az Info.tv-ben⁵⁹ 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 közérdekű adatok, amelyeknek a nyilvánosságra hozatalát törvény ettől függetlenül megköveteli.⁶⁰ A fogalmak fenti rövid ismertetése alapján látható, hogy a közadatok az állami és önkormányzati igazgatás működése során vagy azzal kapcsolatban létrejövő, egyébként is Alaptörvényi szinten⁶¹ 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, addig a Közadattv. az ilyen adatok⁶² további hasznosításának és újrahasznosításának a lehetőségét teremtette meg.⁶³ A két szabályanyag együttesen a teljes – korlátozott körhöz nem tartozó⁶⁴ – nemzeti

⁵⁴ A közadatok újrahasznosításáról szóló 2012. évi LXIII. törvény (Továbbiakban: Közadattv.) 4. § 4. pont

⁵⁵ A Közadattv. szabályozási rezsímje és fogalomhasználata a nyílt hozzáférésű adatokról és a közzféra információinak további felhasználásáról szóló 2019/1024/EU irányelv előírásaira épít

⁵⁶ Közadattv. 4. § 6. pont

⁵⁷ Közadattv. 4. § 13. pont; 2019/1024/EU irányelv 2. cikk 9. pont

⁵⁸ A közügyek átláthatóságához, a közérdekű és a közérdekből nyilvános adatok megismeréséhez és terjesztéséhez fűződő kommunikációs alapjog, hozzátéve, hogy az átláthatóságnak jogágakon átívelő értelmezése lehet, lásd:

Hohmann Balázs: Interpretation the Concept of Transparency in the Strategic and Legislative Documents of Major Intergovernmental Organizations. Közigazgatási és Infokommunikációs Jogi Phd Tanulmányok 2(1), 2021, 48-51. o. és Hohmann Balázs: The interpretation of transparency from the legal point of view. In: Haffner Tamás (Szerk.): IV. Fialok Európában Konferencia -TANULMÁNYKOTET. Pécs, Sopianae Kulturális Egyesület, 2018. 155-163. o.

⁵⁹ Az információs önrendelkezési jogról és az információszabadságról szóló 2011. évi CXII. törvény (Továbbiakban: Info.tv.)

⁶⁰ Info.tv. 3. § 5. és 6. pontok

⁶¹ Magyarország Alaptörvénye VI. cikk (3) bekezdés

⁶² Például meteorológiai adatok, különböző állami nyilvántartások adatai

⁶³ A 2012. évi LXIII. törvény indokolása a közadatok újrahasznosításáról, Általános indokolás

⁶⁴ Így nem minősülhet az adat személyes adatnak, minősített adatnak, üzleti titoknak vagy döntés előkészítő adatnak

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.⁶⁵ A közadatok a közszféra foglalkoztatói esetén kétségtelenül jelentős kategóriát fednek le, hiszen vélelmezhetően minden dolgozó a közfeladat ellátásában vesz részt, ezáltal a velük összefüggésben keletkező/kezelt adatok közadatok (pl. béradatok, létszám, munkakörök stb.).

A személyes adatok alkategóriája

Az adat-fogalompáros másik nagy képviselője a személyes adat. A jelenleg hatályos jog szerint a személyes adatok: „azonosított vagy azonosítható természetes személyre vonatkozó bármely információ.”⁶⁶ 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 adatok⁶⁷ alapján való azonosíthatóságot preferálja, ez azonban az újfent tágra meghúzott „közvetett” azonosíthatóság lehetőségével felülírásra került. 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 személyekre vonatkoznak, és mint ilyenek a védelmüket az Uniós és hazai jog alapvető emberi jogként nevesíti. Ellentétben tehát a közadatokkal ezek az adatok explicite, a jog erejénél fogva védettek. A személyes adatok terminológiai meghatározottsága kevésbé szegmentált, mint a közadatoké, elsődlegesen – Uniós rendelet jellegéből adódó alkalmazási elsőbbsége révén – a GDPR határozza meg, míg az Info.tv. ezt kiegészítheti, pontosíthatja. Az egyéb (frissebben megalkotott) jogszabályok ezeket konzekvensen hivatkozzák, így a hazai és az Uniós 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 érzékenységük 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

⁶⁵ 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.

⁶⁶ 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 Info.tv. 3. § 2. pontjában meghatározott személyes adat fogalom, amely szerint ez az érintettre (értsd: természetes személyre) vonatkozó bármely információ

⁶⁷ 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

természetes személyek egyedi azonosítását célzó biometrikus adatok, az egészségügyi adatok és a természetes személyek szexuális életére vagy szexuális irányultságára vonatkozó személyes adatok.⁶⁸ Fontos továbbá, hogy szintén speciális megítélésűek a bűnügyi 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.⁶⁹ 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,⁷⁰ biometrikus adat⁷¹ 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,⁷² valamint genetikai adat⁷³ egy természetes személy örökölt vagy szerzett genetikai jellemzőire vonatkozó minden olyan személyes adat, amely az adott személy fiziológiájára vagy egészségi állapotára vonatkozó egyedi információt hordoz, és amely elsősorban az említett természetes személyből vett biológiai minta elemzéséből ered.⁷⁴ A személyes adatok fogalmával kapcsolatban az látható, hogy a jogalkotó az általános meghatározásokon túl az egyedi, a kor kihívásaira reagáló egyedi fogalmakat is törekszik megalkotni. Munkaügyi folyamatok végrehajtása adatkezelés nélkül elképzelhetetlenek, így minden foglalkoztató rendkívül nagymennyiségű személyes adat kezelését végzi a mindennapokban. Ezek láthatóan de iure védettek és a kezelésükre kiterjedt adminisztratív előírások vonatkoznak.

A jogi adatfogalommal kapcsolatban természetesen szükséges megemlíteni, hogy léteznek a jog által nem szabályozott, vagy nem adatként kezelt, de mégis adatként létező állományok (sőt minden bizonnyal ezek vannak többségben). Ezek a fenti adatfogalmak valamelyikével korrelálva léteznek, azonban azok se nem személyes, se nem közadatok, nem vagy. Így lehetnek üzleti titkok, minősített adatok, technikai lírások, az informatikai működés során automatikusan generált adatok vagy metaadatok, statisztikák, illetve anonimizált adatok, esetlegesen állatokra, növényekre, épületekre vonatkozó adatok. Adathasznosítás szempontjából nem különben fontosak, azonban e helyütt az elemzésükre – sokféleségük miatt – nem kerülhet sor.

⁶⁸ GDPR 9. cikk (1) bekezdés; Info.tv. 3. § 3. pont

⁶⁹ Info.tv. 3. § 4. pont

⁷⁰ GDPR 4. cikk 15. pont; Info.tv. 3. § 3c. pont

⁷¹ Ilyen például az arckép vagy a daktiloszkópiai adat

⁷² GDPR 4. cikk 14. pont; Info.tv. 3. § 3b. pont; Javaslat az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (Továbbiakban: MI rendelet) megállapításáról és egyes uniós jogalkotási aktusok módosításáról 3. cikk 33. pont

⁷³ 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

⁷⁴ GDPR 4. cikk 13. pont; Info.tv. 3. § 3a. pont

Összegzés

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. A munkaügyi folyamatoknak elengedhetetlen része nagy mennyiségű adat kezelése, ezeknek az azonosítására, fogalmi elhatárolására szükség lehet a helyes gyakorlatok kialakításához, illetve a jogszerű működés megvalósításához.