

Big Data – kihívások és kockázatok

Manuela Krauß¹ ■ Tóth Tamás¹ ■ Heinrich Hanika dr.¹

Kozlovsky Miklós dr.² ■ Dinya Elek dr.¹

¹Semmelweis Egyetem, Egészségügyi Közzszolgálati Kar, Digitális Egészségtudományi Intézet, Budapest

²Óbudai Egyetem, Neumann János Informatikai Kar, Biomatika Intézet, Budapest

A „Big Data” az utóbbi időben gyakran használt kifejezés, amely arra utal, hogy egyre nagyobb digitális információ-tömeg keletkezik. Ezen információk összekapcsolásával, feldolgozásával és elemzésével újfajta következtetések vonhatók le, és új szolgáltatások alapulhatnak rájuk. Ez kihatással van az élet minden területére, többek közt az egészségügyre is. A szerzők áttekintik a Big Data alkalmazási lehetőségeit az orvostudományból és más területekről vett példák segítségével. A lehetőségek jobb kihasználásához szükség van megfelelő infrastruktúrára, a szükséges szabályozási környezet kialakítására, különös hangsúllyal az adatvédelmi és adatbiztonsági kérdésekre. E problémákat és a megoldásukra tett intézkedéseket is bemutatják a szerzők. *Orv. Hetil.*, 2015, 156(49), 1979–1986.

Kulcsszavak: Big Data, digitális menetrend, személyre szabott orvoslás, telemedicina, adatvédelem

Big Data – challenges and risks

The term “Big Data” is commonly used to describe the growing mass of information being created recently. New conclusions can be drawn and new services can be developed by the connection, processing and analysis of these information. This affects all aspects of life, including health and medicine. The authors review the application areas of Big Data, and present examples from health and other areas. However, there are several preconditions of the effective use of the opportunities: proper infrastructure, well defined regulatory environment with particular emphasis on data protection and privacy. These issues and the current actions for solution are also presented.

Keywords: Big Data, digital agenda, personalized medicine, telemedicine, data protection

Krauß, M., Tóth, T., Hanika, H., Kozlovsky, M., Dinya, E. [Big Data – challenges and risks]. *Orv. Hetil.*, 2015, 156(49), 1979–1986.

(Beérkezett: 2015. június 18.; elfogadva: 2015. szeptember 27.)

A Big Data forradalmasítja az életünket

Big Data – a jelen digitális korunkban kevés dolog megy át olyan nagy változásokon, mint az interaktív kommunikációval összefüggő adatok gyűjtése, elemzése és kiértékelése. A digitális adatok mennyisége évente megduplázódik, és a becslések szerint 2020-ra eléri a 44 zettabyte-ot (44 billió gigabyte-ot) [1]. Ez a mennyiség különféle médiumok között oszlik meg – beletartozik a fényképek és videók megosztása a közösségi oldalakon, a mobil eszközökkel történő különféle interakciók, a GPS-adatok, a pénzügyi tranzakciók, az időjárási adatok éppúgy, mint a személyre szabott betegadatok [2].

A Big Data kifejezés tehát hatalmas mennyiségű, komplex adatra utal, amelyeket a napi életünk különféle

területeiről gyűjtenek össze, és amelyek gyors tárolásához, feldolgozásához és kiértékeléséhez speciális alkalmazások szükségesek: Az információk kinyeréséhez az adatrekord számos mezőjét kell rövid választ idővel feldolgozni, nagyszámú lekérdezést kell párhuzamosan végrehajtani és esetenként más, importált adatokkal kombinálni [3]. Ezek az egyéni és vállalati információk számos lehetőséget kínálnak a mindennapi életünk javítására, megváltoztatására. Az ajánlórendszerek feldolgozzák és a felhasználóhoz rendelik a személyes adatokat, a kedvenc termékeket és szolgáltatásokat, a vásárlások árkatégoriáit és még sok más, majd ezek alapján egyénre szabott, aktuális ajánlatokat kínálnak.

Ezek a speciális ajánlatok a felhasználói profilon és más forrásokból származó adatokon alapulnak, amelyek to-

vábbi feldolgozás céljából a szolgáltató rendelkezésére állnak (legtöbbször a felhasználó szabad akaratából). Ezekből a vállalatok fontos információkhoz juthatnak a vásárlási és szolgáltatás-igénybevételi szokásokról, és nagyon gyorsan reagálhatnak a piaci igényekre a szolgáltatások vagy termékek megváltoztatásával, illetve adaptálásával. A piac közvetlen megfigyeléséből nemcsak a célcsoport viselkedésére vagy az érintett piacra vonatkozóan lehet következtetéseket levonni, hanem akár a jövőben tervezett projektekre nézve is. Az empirikus összefüggések gyors, automatizált kinyeréséhez speciális módszerek és algoritmusok használatára van szükség, amelyeket *adattárolás* néven ismerünk. Ezek az optimalizált és gyors elemzési eredmények legtöbbször indikátorként szolgálnak a menedzsment számára azzal a céllal, hogy hatékonyabbá tegyék a projektcontrollingot [4].

Az adatok alapvető feldolgozása szempontjából a forrásuk lényegtelen; származhatnak az internetről, a közösségi médiából, mobileszközökről, hitel- és vásárlói kártyákról, az energiagazdaságból (intelligens mérőeszközöktől), megfigyelőkameráktól, a közlekedésből vagy az egészségügyből [4].

Az egészségügy és az orvostudomány számára a számítási felhő kiemelkedő szerepet játszik. Lehetővé teszi a szükséges személyes adatok tárolását, és vészhelyzetben azok elérését a világ bármely részéről. A betegek számára ezzel a célzott kezeléshez szükséges információátvitel ideje minimalizálható. Az orvos azonnal hozzáférhet az allergiák, a korábbi kezelések és azok eredményének adataihoz, majd ezek figyelembevételével határozhatja meg a szükséges kezelést. Ez különösen fontos lehet akut betegségek és balesetek esetén. A számítási felhő még további jelentős potenciállal bír. A kutatásban a szakemberek közötti azonnali kommunikáció lehetősége sokszor döntő szerepet játszik. Így a járványok határok közé szoríthatók, a kutatási eredmények átadhatók és a terápiában alkalmazhatók, a klimatikus veszélyek előfordulása észrevehető és a szükséges óvintézkedések megtehetőek.

Már önmagában a nagy mennyiségű információ összegyűjtése, a különböző kiértékelési kritériumok szerinti kategorizálása és az információk gazdasági és szociális szektorokhoz való célzott eljuttatása teljesen új szolgáltatásokat és innovatív technológiai megoldásokat igényel [5]. Gyakori, hogy az adatokat redundánsan, a világ bizonyos pontjain elosztva tárolják, hogy esetleges sérülés vagy egyes kommunikációs kapcsolatok kiesése esetén se vesszenek el adatok. Azonban ezeket a redundanciákat gyorsan tudni kell replikálni, hogy azok frissítés esetén rövid idő alatt elérhetőek legyenek [5].

Ehhez elosztott rendszerekre van szükség, amelyek biztosítják a folyamatosan növekvő adatmennyiség horizontális skálázhatóságát, és lehetővé teszik szükség szerint további számítási kapacitások bevonását, illetve rendszerek közötti szétosztását [5]. Ezzel szemben a vertikális skálázhatóság korlátait az elérhető hardverek jelentik. Ha a legújabb és leggyorsabb hardverek imple-

mentálásra kerülnek, a vertikális skálázhatóság lehetőségei kimerülnek [6]. Az elosztott rendszerek skálázhatóságának szükségessége tehát egy fontos konceptuális előfeltétel, amelyet figyelembe kell venni a tervezés, az implementálás és a modellezés során is [5].

Döntő szerepet játszanak az adatbázisok, amelyek technológiái az adattárolás, -elemzés és -feldolgozás igényeihez igazodnak. *Park Kieun* az alábbi megközelítési lehetőségeket javasolja a Big Data adatbázis-technológiájára [7]:

- *Massively parallel processing (MPP)* vagy *párhuzamos DBMS* [8]: Az adatlekérdezések párhuzamos végrehajtása az elosztott adatbázis-menedzsmentrendszer különböző csomópontjain nagy mennyiségű adat egy időben történő feldolgozása céljából.
- *Oszloporientált adatbázis* [9]: Az azonos mezőben lévő adatok megjelenítése oszlopként és nem egyedi adatrekordként.
- *Streaming processing (ESP vagy CEP)* [10]: Adatbázis-konceptió olyan adatok vagy események számára, amelyek tartalma állandóan és folyamatosan változik.
- *Kulcsértéktárolás* [11]: Olyan technológia, amely egy kulcsértéket használ a teljesítmény növelésére egy egyedi rekord kiolvasásakor (a relációs adatmodell egyszerűsítése).

A fentiek csak néhány kiválasztott példát mutatnak a Big Data kezeléséhez szükséges adatbázis-technológiákra. Az adatok akár közel valós idejű gyűjtésével, feldolgozásával és kiértékelésével kapcsolatos további követelmények függvényében optimalizált adatbázis- és szoftvermegoldásokra van szükség.

Az új korszak

A kommunikációs formák az utóbbi években alapvető változáson mennek át. A mobileszközöknek (okostelefonok, tabletek) köszönhetően szinte mindig, mindenhol személyesen elérhetők vagyunk. A világon az okostelefon-felhasználók száma hamarosan átlépi a 2 milliárdot [12], Magyarországon pedig a becslések szerint több mint 3 millió ilyen eszközt használnak [13].

Az első Apple mobileszközt 2007-es bevezetése óta használhatjuk a vizuális és auditív élményeink rögzítésére, hírek olvasására, szórakozásra, privát és szakmai információk átvitelére és számtalan más célra. A mobileszközről a kommunikációs platformok eléréséhez csak egy böngészőre vagy egy speciális, az adott operációs rendszerre fejlesztett alkalmazásra van szükség. Ha a felhasználó telepíti ezeket az appokat az eszközére, számos személyes információhoz, mint például a tartózkodási helyéhez (GPS segítségével) hozzáférést biztosít, és lehetővé teszi a tárolt adatok elérését. Ezek az információk egy adattárházba kerülhetnek, majd onnan egy célzott adatbázisba. Az adatok hozzárendelése az ETL-folyamat (extract, transform, load) során történik, amelyben a különböző forrásokból származó releváns adatok a céladatbázis sémájába és formátumába konvertálódnak, illetve

abba betöltődnek. Innentől kezdve megtörténhet az adatok feldolgozása és különféle információs rendszerekhez vagy döntéstámogató rendszerekhez rendelése (adatbányászat) [14].

Az európai digitális menetrend

Az Európai Bizottság 2014 júliusában bemutatott egy új stratégiát a Big Data kezelésére és az adatközpontú gazdaságra való áttérés támogatására és felgyorsítására. „Az adatközpontú gazdaság ösztönözni fogja az adatokkal kapcsolatos kutatást és az innovációt, miközben Európában új üzleti lehetőségekhez vezet és növeli a tudás és a tőke elérhetőségét” [15].

A bizottság szerint a Big Data új lehetőségeket nyit meg, amelyek segítenek olyan új termékek és szolgáltatások létrehozásában, mint például különféle mobilalkalmazások vagy az üzleti intelligenciatermékek a vállalkozások számára. A Big Data fokozhatja a növekedést és a foglalkoztatottságot, valamint javíthatja az európaiak életminőségét. „A Big Data széles körű használata az egészségügyi szektorban segítheti az orvosokat abban, hogy gyorsabban hozzák meg a helyes döntést, felhasználva a más egészségügyi szolgáltatók által gyűjtött adatokat is. A betegek a gyorsabb és megfelelőbb kezelésekből profitálhatnak, és jobb tájékoztatást kaphatnak az egészségügyi szolgáltatókról. Az adatelemzés növekvő mértékű használata jelentős megtakarításokhoz is vezethet, a szükségtelen eljárások vagy duplikált tesztek pontosabb azonosítása által. A nagy klinikai adatállományok elemzése az új gyógyszerek klinikai és költséghatékonyságának optimalizálásához vezethet” [16].

A német út

Az utóbbi években a digitális változások társadalmunkat az élet minden területén befolyásolták. Az innovációk és az új gazdasági területek előmozdítására olyan keretfeltételeket kell szabni, amelyek kijelölik a digitális információk használatának lehetőségeit és határait. 2014 augusztusában a német kormányzat elfogadta a digitális menetrendet azzal a céllal, hogy pontos ütemtervet nyújtson a jövőbeni növekedéshez és foglalkoztatottsághoz, hozzáférést biztosítson a nagy teljesítményű hálózatokhoz és biztosítsa az internet használatához szükséges bizalmat és biztonságot [17]. Ezzel elindult a változásoknak egy olyan nyitott folyamata, amelynek résztvevői teljes körű párbeszéd során alakítják ki a következő évek információs technológiáját.

Digitális menetrend, 2014–2017

A német digitális menetrend életre hívta az Ipar 4.0 (negyedik ipari forradalom) projektet, amely nemcsak a számítógépes gyártástechnológiákat veszi figyelembe, hanem az innovatív fejlesztések, az intelligens gyárak, új

üzleti területek és szolgáltatások, valamint a gazdaság biztonsága iránti egyre növekvő igényeket is. Ezen fejlesztések eredményeként növelhető a lakosság életszínvonal és biztosítható Németország fenntarthatósága.

Egy fontos projektelem a komplex digitális hozzáférés és kommunikációs lehetőségek biztosítása a vidéki területeken. A vezetékes és vezeték nélküli, nagy sebességű hálózatok kiépítése és kiterjesztése nagymértékben megkönnyíti a lakosság megélhetését a nagyvárosokon kívül is, ami hatással lesz a társadalmi és magánélet minden területére. Az orvosi és egészségközpontú ellátás már önmagában új lehetőségeket nyújt.

Új közlekedésbiztonsági rendszerek számára is megnyílik az út. A hangsúly a biztonság automatizált rendszerekkel történő növelésén van, ezért vezetnek be a vészhívó rendszert (eCall) kötelezően minden 2018. március 31. után forgalomba helyezett új személygépjárműben és kishaszonjárműben. Baleset bekövetkezése esetén a beépített eCall-egység néhány ezredmásodpercen belül SMS-ben értesíti az egységes európai vészhívószám (112) keresztül a legközelebbi mentőállomást, amely azonnal küldi a megfelelő egységeket. Az Európai Bizottság eSafety kezdeményezésében 28 ország vesz részt [18]. Ez azt is eredményezi, hogy további jogi keretekre és új üzleti területekre is szükség lesz.

Hatások az egészségügyre és az orvostudományra (eHealth)

A már említett eSafety projekt ötlete garantálja az európai együttműködést és a gyors beavatkozást közúti balesetek esetén – egy fontos lépés, amelyhez az orvosi és technikai eszközök együttműködése szükséges a közös európai platformon.

A széles körű digitalizáció és a kommunikációs hálózatok kiterjesztése fontos előfeltételei a digitális menetrendnek, és egyúttal az egészségügy minden területére kiterjedő kommunikációs hálózat kiépítésének. Ezért 2018-ra egész Németországban legalább 50 Mbit/s sáv szélességű hálózat kiépítése szükséges [17]. Különös figyelmet kell fordítani a vidéki régiókban fennálló hátrányok leküzdésére. Az országos, nagy teljesítményű hálózat kiépülésével a polgárok optimális orvosi ellátásban és gondozásban részesülhetnek a lehető legjobb egészségi állapot eléréséhez, és a kor előrehaladtával a lehető leghosszabb ideig maradhatnak a megszokott személyes környezetükben. Ez elérhető „az eHealth-kezdeményezések bővítése és az egészségügyi vállalkozások innovációs tevékenységének szorosabb együttműködése” [17] által. Ezért két kompetenciaközpontot állítanak fel Berlinben és Drezdában, amelyek feladata a Big Data egészségügyi, tudományos és üzleti kihívásainak és lehetőségeinek vizsgálata. Az innováció támogatásával lehetővé válik a javuló minőségen és hatékonyságon keresztül egy biztonságos és eredményes egészségügyi rendszer létrehozása.

Új kihívások az orvostudományban

Telemedicina

A telemedicina elterjedése jelentős hatással lesz az egészségügyi ellátórendszerre. A német tartományi egészségügyi vezetők tanácsának elnöke, *Sabine Bätzing-Lichtenthäler* szerint például a telemedicina ellensúlyozhatja az orvoshiányt az országban. „Nem ez lesz a csodaszer a demográfiai változások ellen, de segíthet az orvosi ellátás lehetőségei szerinti biztosításában. Ez nem azt jelenti, hogy le kell tennünk a háziorvosok és szakorvosok számának növeléséről, de segítséget nyújthat a jelenlegi helyzet javításában. Ez is egy nyomós érv ahhoz, hogy az eHealth-törvény formát öltön” [19].

A szívelégtelenségben szenvedő betegek számára indult például egy telemonitoring projekt Rajna-vidék-Pfalz tartományban (E.He.R. projekt). Ezentúl a vérnyomásméréshez már nem kell felkeresni az orvost. Ez a példa más betegségekre is alkalmazható, hiszen a tünetek távfelügyelete fontos eleme annak, hogy a beteg a lehető legtovább a megszokott otthoni, családi környezetében maradjon. Ez támogatja a felépülést, és a speciális intézményben történő ellátás költségei is megtakaríthatók [20].

Az ilyen megfigyelőrendszerek különösen szükségesek a vidéki és a központoktól távoli területeken. A közlgő házi- és szakorvoshiány ellen indította a német kormányzat a „Masterplan 2020” programot, amelynek célja, hogy vonzóbbá tegye a nagy kiterjedésű és ritkán lakott területeken elhelyezkedő orvosi praxisokban való munkát [21]. Rajna-vidék-Pfalz tartományban egy további helyi kampány is indult a vidéki háziorvosi praxisok támogatására. Ez a probléma nem csak Németországra jellemző. Az európai nyitással és az egészségügyi szolgáltatások határon átnyúló igénybevételi lehetőségeivel a telemedicina elsődleges fontosságúvá válik. A diagnosztikai és terápiás szolgáltatások térbeli és időbeli távolságtól függetlenül lehetővé válnak. A telekommunikáció segítségével az orvosok egymással vagy akár a betegekkel, gyógyszereszekkel és más szakemberekkel is kommunikálhatnak, és különféle tevékenységeket írhatnak elő vagy végezhetnek el.

Gondoljunk csak a zsúfolt kórházi és szakrendelőbeli várótermekre – az aktuálisan várakozó betegek száma a telemedicina használatával követhető lenne. Ez megkönnyítené az orvosválasztást, illetve a vizitre alkalmas időpont megválasztását. A betegek számára ez csak egy távoli vízióknak tűnik, azonban az orvosi személyzet részvételével valósággá válhat. A technikai feltételek már régóta adottak hozzá.

Bioinformatika

Orvosbiológiai területen a Big Data egyik legjellemzőbb alkalmazási motivációját a bioinformatika szolgáltatta. Alapvetően (hasonlóan a csillagászhoz vagy a You-

tube/Twitter típusú közösségi médiákhoz) az alábbi négy nagy kihívással küzd napjainkban ez a terület is:

- Az adatok begyűjtésének problémája.
- Az adatok tárolásának problémája.
- Az adatok hozzáférhetővé tétele mások számára.
- Az adatok feldolgozásának problémája.

A bioinformatika interdiszciplináris terület: informatikai (mérnöki) megoldások alkalmazása olyan orvosi problémákra, amelyek jellegükből fakadóan speciális követelményeket támasztanak, s ezek közül is talán a legfontosabb a nagy mennyiségű adatok hatékony kezelésének igénye [22]. Ez speciális algoritmusokat, speciális adatbázis-kezelési módszereket, speciális adatszerkezeteket igényel.

Bár a bioinformatika alkalmazási területei rendkívül széleskörűek (és szinte napról napra bővül a lista), érdemes néhány jellemzőjével közelebbről is megismerkedni [23]. A genomika tudománya (és alapvetően az „omics” világ) az 1980-as évektől indult el hódító útjára [24]. A bioinformatika és a genomika korán egymásra talált, köszönhető ez a hatalmas mennyiségű adathalmaznak, amely kezdetben fehérjék aminosavsorrendjét, később kiegészülve a genomok bázissorrendjét tartalmazta. A különféle szekvenciaadatok feldolgozása, hasonló szekvenciák keresése, illetve szekvenciák illesztése (sequence alignment) alapvetően napjainkban is a „legklasszikusabb” területek, amikkel máig sokan azonosítják a bioinformatikát.

Ezen területek azért bírnak különös jelentőséggel, mert a hasonló szekvenciák hasonló eredetre utalhatnak, így vizsgálható a konzerváltság, filogenetikai fák konstruálhatók stb., továbbá levonhatunk funkcionalitásra vonatkozó következtetéseket is. A szekvenciaillesztés kézi úton nyilván nem járható, ám számítógéppel sem triviális, hiszen a sok összehasonlítandó szekvencia és az egyes szekvenciák hossza miatt a kombinációk száma a direkt végigpróbálgatáshoz reménytelenül sok.

A genomika egyik nagy korai (1990-es évek) kihívása a hatalmas összegeket felemészítő, 15 évesre tervezett Humán genom projekt (HGP) volt. A projekt során az új számítógépes támogatásra épülő módszerek segítségével (shotgun módszer) a tervezett projektet évekkel korábban sikerült befejezni [25]. Azóta a teljes genom szekvenálása sokat változott, a 2001-es évek környéki 100 millió dolláros teljesgenom-szekvenálási árak napjainkra leestek 8–10 ezer dollár alá [26]. A shotgun módszert követő újabb generációs szekvenálási módszerek (NGS) gyorsabbak, pontosabbak és növekvő sebességgel egyre hatalmasabb mennyiségű adatot állítanak elő. A jelenlegi méret/hordozhatósági, valamint pontossági viszonyokat jól jellemzi az Oxford Nanopore Technologies által fejlesztett MinION berendezés [27], amely USB-stick méretben képes a DNS, RNS, fehérje és az apróbb molekulák analizálását végezni valós időben. Ezen adatok a közel exponenciálisan növekedő adattárházakba kerülnek. A 2000-es évektől kezdve leginkább három nagyobb, egymással napi frissítésben álló adattár-

házról (az amerikai NCBI/GenBank, az európai EMBL-Bank [28] és az ázsiai DDBJ), valamint több szétszórta kutatói/intézeti/vállalati adattárházról beszélhettünk, amelyek között referencia-ID-ekkel lehetett eligazodni. A közel 15 év alatt ezen adattárházak tovább, egyre csak tovább híztak, nemcsak méretben, hanem komplexitásukban is [29].

Ezen hatalmas információtengert csak szuperszámítógépekkel, párhuzamos és elosztott működésű infrastruktúrákkal, valamint erős szoftveres támogatással lehet kezelni. Az adatmennyiség megköveteli a skálázható, optimalizált algoritmusokat, amelyek képesek akár párhuzamos fűrtökön (cluster) több száz gépen futva számolni a feladat megoldását. A korábbi években ezeknek a problémáknak a megoldását klasszikusan a dinamikus programozásra (probléma dekompozíciója kisebb részproblémákra, ilyen a *Needleman–Wunsch*- [30] vagy *Smith–Waterman*- [31] algoritmusok) építették, de ma már gyakran mesterséges intelligenciát is alkalmazó eljárásokat használnak. Ezen algoritmusok jellemzően függvénykönyvtárakban érhetőek el mindenki számára. Talán leghíresebb képviselőjük a BLAST algoritmus [32].

A bioinformatika természetesen egyéb területekre is kiterjed. Említésre méltó példa a teljes genomikai és megbetegedési adatok egymással való szembesítése, és ez alapján betegségek genetikai alapjainak keresése (genome-wide association analysis). Ennek klasszikus módszere, hogy adott betegségben szenvedő, illetve hozzájuk a lehető legjobban hasonlító, de a betegségben nem szenvedő alanyok genomját összevetik, keresve, hogy mik azok a genetikai variánsok, amelyek a betegeknel gyakoribbak. Ez nyilván szintén csak megfelelő informatikai támogatással képzelhető el, ma már számos nevezetese alkalmazási példa mutatható erre is [33].

Személyre szabott versus osztályozott orvoslás

A technológia fejlődésével a különféle új bioinformatikai vizsgálómódszerek egyre gyorsabbak, gazdaságosabbak, megbízhatóbbak lesznek, s elvezetnek a személyre szabott orvosláshoz, amelynek lehetősége jelenleg is a kutatások középpontjában áll. Ez azonban nemcsak technológiai, hanem fontos etikai kérdéseket is felvet [34]. Ezek megtárgyalása kívül esik jelen dolgozat keretein. Az irodalomban gyakran említett alternatíva az „osztályozott orvoslás”, amelyben a betegpopulációkat kisebb alcsoportokba sorolják, amely csoportok mindegyike meghatározott kezelésben részesül. A csoportosítás alapját valamilyen klinikai jellemző, sokszor biomarker jelenti [35].

A személyre szabott orvoslás fogalma használatos arra is, hogy leírja a súlyos betegségekben (daganatos betegség, diabetes, szív-ér rendszeri betegségek stb.) szenvedő betegek megelőzési és egyéni terápiás lehetőségeit. A mikrobiológia és a genetika fejlődése lehetővé teszi az egyéni (mikrobiológiai, szociológiai és akár pszichológiai)

ai) tulajdonságok diagnosztizálását, illetve az egyénre szabott előrejelzések és terápiák kidolgozását [36].

Hans Lebrach, a Max Plank Molekuláris Genetikai Intézet kutatójának víziója egy olyan, anatómiai, fiziológiai, molekuláris biológiai adatokon és teljesgenom-szekvencián alapuló számítógépes modell kialakítása minden egyes betegről, amelynek segítségével az orvosok személyre szabott terápiát nyújthatnak a betegnek [37]. Ha továbbgondoljuk ezt a víziót, nemsokára a súlyos betegségek már kialakulásuk előtt felismerhetővé válnak, és lehetővé válik a célzott, egyénre szabott prevenció vagy a korai fázisban történő kezelés. A kutatásnak ez az állapota a meglévő technikai lehetőségek bevonásával elérhető. A mikrobiológiai eredmények feldolgozása és a szociológiai, pszichológiai ismeretek párhuzamos elemzése lehetővé teszi az orvos számára a további eljárások gyors kiválasztását.

Azoknál a betegeknel, akik egy ritka örökletes variációt hordoznak, a csoportspecifikus kezelési módszerek nem működnek. A diagnózis során az egyes betegeket kell értékelni és speciális kezelési terveket kidolgozni számukra. A határ a betegek és bizonyos genetikai anyagok hordozói között sokszor nagyon keskeny [38].

Gyakran jelennek meg kritikus hangok, amelyek attól óvnak, hogy a gazdasági érdekek előtérbe kerüljenek, illetve a kutatási prioritásokat a betegcsoportok elemzésében (vagy akár szocio-orvosi kutatásokban) látják [38]. Ahhoz, hogy ez a terület a mindennapi élet részévé váljon, még számos jogi, technikai és pénzügyi keretet tisztázni kell. Azonban a demográfiai trendek gyors döntésre sarkallnak.

Big Data az egészségügyben

A Big Data különösen fontos az egészségügy területén, mivel a digitális információk mennyisége exponenciálisan nő. Például egy kórházban naponta 12 millió regénynek megfelelő adatmennyiség keletkezik. Ezek már olyan mennyiségek, amelyeket információk technológiák nélkül lehetetlen kezelni. A Big Data „a XXI. század nyersanyaga”, és – az egészségügyben is – talán nagyobb mértékben változtatja meg a világot, mint azt eddig értjük, kívánjuk vagy elfogadjuk [39].

Cselekvési területek az orvostudományban

Langkafel az alábbi területeket írja le [39]:

- Egészségnevelés és információ: Minden olyan terület, ahol az adatok különböző szereplők számára (újszerű) összeállításra vagy kibővítésre kerülnek (például betegségmenedzsment-programok újratervezése).
- A bemenetek (erőforrások) elemzése lehetővé teszi az erőforrás-menedzsment és ezzel az orvosi minőség és hatékonyság új formáit. Ez lehetővé teszi az egész szervezetre kiterjedő megközelítéseket (pay per performance) vagy új ellátási formák kialakítását.

- Közegészségügyi monitoring: Gyorsabb és célzottabb beavatkozások az egészségügyi trendek alapján, egyéni egészségjavító intézkedések.
- Gyors szubjektív és objektív visszacsatolás a termékfejlesztéseknél – új szolgáltatások a betegségek kezelése terén vagy újfajta orvostechnikai eszközök az AAL (ambient assisted living) területén.
- Betegségprognózis az „egyéni egészséganalitikai” megközelítés által: Az élő környezet adatainak integrálása lehetővé teszi a betegségek jobb megértését.
- Megelőzés: Az adatok integrálása és elemzése lehetővé teszi a gyorsabb és célzottabb prevenciót és az adaptálását kampányok aktuális kihívásaihoz vagy olyan akut veszélyekhez, mint például a járványok.
- A betegségmentesítő-programok egyénileg testre szabhatók.
- Klinikai vizsgálatok: A kiválasztás és az adatok „szimulációja” gyorsan lehetővé teszi új vizsgálati típusok kialakítását, és jobban összekapcsolhatja a klinikai ellátást és a kutatást.
- Új termékek és szolgáltatások a kiterjesztett egészségpiac számára (táplálkozás, fitness, wellness).
- Orvosi teljesítményoptimalizálás: Lehetővé teszi az aktuális irányelvek jobb végrehajtását (az orvos támogatásával) és az átláthatóbb kimenetmérést.
- Betegek közötti információcsere: Az első és legismertebb példa a www.patientslikeme.com – egy internetes platform, ahol a betegek más, „hasonló” tünetekkel és tapasztalatokkal rendelkező betegeket kereshetnek.
- Kommunikációs folyamatok és scenáriók, amelyek segítenek leküzdeni az intézmények közötti korlátokat, és lehetővé teszik újfajta egészségügyi szolgáltatások tervezését és végrehajtását.
- Adherencia: Olyan szituációk, amelyekben egy személy viselkedése, például gyógyszeresedés, diéta és/vagy életmódbeli változások jobban illeszkednek a kezelő szakember ajánlásaihoz.
- Új üzletek és új szereplők: A posta, amely gondozási szolgáltatást nyújt; a szupermarket, amely diétás programban vesz részt; a távközlési szolgáltató, amely telemedicina szolgáltatást nyújt.
- Kezelési hibák: A Big Data alkalmazása lehetővé teszi a műhibák elleni jobb védekezést
- Az ellenőrzések új intézkedéseket integrálhatnak, amelyek hatékonyabb egyéni vagy területi komponenseket tartalmaznak.
- A bizonyítékok gyorsabb implementálása a Big Data alkalmazások egyik kiemelt témája.
- Orvos és beteg közötti információcsere: Telemedicina-szolgáltatások, valós idejű adatelemzés, diabetes-monitoring stb.

Big Data stratégiák az egészségbiztosítók szemszögéből

Brunner az alábbi stratégiákat írja le és elemzi az egészségbiztosító társaságok szemszögéből [40]:

- Prevenció: Milyen valószínűséggel betegszik meg egy biztosított egy adott betegségben, célzott prevenciók és terápiás programok, költségcsökkentés.
 - Szolgáltatások: Online szolgáltatások kiépítése; „elektronikus számla”, aminek segítségével a biztosított online tájékozódhat arról, mennyit költött a biztosító a legutóbbi kórházi kezelésre vagy vizitre; regisztrációs lehetőség gondozási programokra; chat-ügyfélszolgálat; komplex ajánlatok egyénre szabott szolgáltatásokra.
 - Marketing/sales: Imázsépítés közösségi hálózatokon; online hirdetési kampányok; konkrét alkalmazások megvalósítása.
 - Folyamatautomatizálás: Dokumentumok gyűjtésének automatizálása; adatok kezelése „jobb” üzleti alkalmazásokhoz.
 - Nemkívánatos viselkedés megelőzése: Csalások felismerése, a különböző résztvevők interakcióinak vizsgálata (gyógyszerész, orvos, biztosított), közösségi hálózatok hivatkozásainak figyelése.
- Összefoglalva, a Big Data segítségével elérhető a stratégiai versenylőny, a teljesítmény- és hatékonyságnövelés, a prevenciók és kezelési programok optimalizálása, több szolgáltatás és nagyobb átláthatóság a biztosítottak számára és a gyorsabb innováció.

Big Data stratégiák kórházak számára

Wehmeier és Baumann az alábbi stratégiákat ajánlja [41]:

- A jobb adatbázis jobb diagnózisokhoz vezet, és a digitális, személyre szabott hatásvizsgálatok jobb terápiamentesítőket tesznek lehetővé.
- Már ma is léteznek olyan adatbázisok, amelyek interfésszel rendelkeznek a kórházi, laboratóriumi, radiológiai vagy patológiai rendszerek felé, és lehetővé teszik előre elkészített adatelemzések (riportok) elkészítését. Lehetőség van a Big Data új alkalmazásaira is, mint például a vRad radiológiai szolgáltatás, amely az adatok alapján képes optimalizálni a radiológián belüli logisztikát.
- Genomszekvenálás: Milyen valószínűséggel betegszik meg a beteg egy adott betegségben? Melyik terápiás út nyújtja a legígéretesebb eredményt a hasonló esetek alapján?
- A kórházi informatika biztonságának növelése.
- Analitikai szolgáltatások, mint például a Deutsche Telekom Cyber Sec. Adv. vagy Cyber Defense SOC szolgáltatása, amelyek más iparágakban már bizonyították, hogy képesek a veszélyforrások felismerésére és kezelésére.

A jövőt illetően az látszik, hogy a Big Data alkalmazásának alapja az önkéntes és informált hozzájárulás kultúrája kell, hogy legyen. Ehhez meg kell mutatni a felhasználónak az előnyöket, és biztosítani kell annak átláthatóságát, hogy mi történik az adatokkal. Továbbá az interoperabilitás és a világos jogi környezet is fontos előfeltételei a sikernek.

Nemzeti és nemzetközi adatvédelmi szabályozások

A Big Data egészségügyi alkalmazása magas adatvédelmi követelményekkel jár. Az adatforrások és az elemzések célja előre definiálendő. Az esetleges nem kívánt hatások kezelésére kockázatelemzést kell végezni és a szükséges védelmi intézkedésekről gondoskodni kell. Az anonimizálás és pszeuduonimizálás, valamint a jogilag biztosított technikai, szervezeti és eljárási követelmények biztosítják, hogy a jogszabályban meghatározott, mégis orvosi-lag hasznos eredmény szülessen. Ily módon lehetővé válik a Big Data egészségügyi alkalmazása, a törvényi szabályozással és alapvető személyiségi jogokkal összhangban [42].

Számos szabályozási jogforrás létezik Európában és Németországban is, mint például: az Európai Alapjogi Charta, az adatvédelmi irányelv (95/46/EK), az alaptörvény, a szövetségi adatvédelmi törvény és a tartományok adatvédelmi törvényei, a társadalombiztosítási szabályozások, a kórháztörvény, az egészségügyi szolgáltatókról szóló törvény, a genetikai diagnosztikát szabályozó törvény, a fertőzések elleni védekezésről szóló törvény, a gyógyszerjogi törvény és más ágazati jogszabályok.

Átláthatóság nélkül az érzékeny adatokat kezelő Big Data nem valósítható meg. Az átláthatóságnak az adatfeldolgozás minden lényeges szempontjára ki kell terjednie. Ez magában foglalja a jogi alapokat, az azokból levezetett szabályozásokat, a szervezeti és eljárási szabályokat, a műszaki dokumentációkat, az adatok, az adatbiztonság és az adatvédelem menedzselését. Szükség van etikai bizottságok minősített vizsgálati és jóváhagyási folyamatainak kidolgozására, adatvédelmi felelősök kijelölésére, adatvédelmi hatóságra és ügynökségekre, amelyek felelősek a folyamatok vagy termékek minőségbiztosítási és auditálási szabályainak kidolgozásáért.

A nemzetközi szabályozásokhoz a következőket kell hozzáfűzni: Az adatvédelem elfogadott „safe harbor” („biztonságos kikötő”) elve, amely a személyes adatoknak az Európai Unióból az Amerikai Egyesült Államokba továbbítását a magánélet védelmét tiszteletben tartó feldolgozáshoz köti, nem megfelelő. A transzatlanti kereskedelmi és beruházási tárgyalások során az adatvédelmi és biztonsági kérdések az európai vállalatok prioritása kellene, hogy legyen. Az Európai Unió és az Egyesült Államok között jelenleg is tárgyalások vannak folyamatban az eHealth, a titkosítás, a jogorvoslati lehetőségek és az elektronikus azonosítás témakörében. Az Európai Uniónak hamarosan le kell zárnia az új adatvédelmi szabályozással kapcsolatos vitát, és el kell köteleznie magát mindennemű adat átfogó védelme és a magas adatvédelmi szint iránt, amely biztosítja a meglévő adatvédelmi szabályozások betartását. Az Európai Unió vállalatának a megfelelő adatbiztonsággal kell hozzájárulnia az adatvédelemhez, védve ezzel a polgárokat a kémkedés ellen.

A fennálló megállapodások, mint például a légi utasok adatainak átadása (PNR) vagy az Amerikai Egyesült Álla-

mok hozzáférése bizonyos banki adatokhoz (SWIFT) az európai polgárok adatainak igen kiterjedt köréhez biztosít hozzáférést az Amerikai Egyesült Államok számára. Az Európai Uniónak újra kellene gondolnia ezeket. Globális szabályokra van szükségünk a kémkedés visszaszorítására. Előnyös lenne egy ENSZ-megállapodás, amely egy kiegészítést fűzne az ENSZ Polgári és Politikai Jogok Nemzetközi Egyezségokmánya 17. cikkelyéhez. Egy ilyen protokoll a nemzetközi jog alapján kötelező érvényű lenne.

Következtetések

A digitalizáció gyors és tartós változást okoz az élet számos területén. Mindannyian a digitális forradalom közepén élünk. A digitális média által kínált lehetőségeket az egészségügyben azonban még gondatlanul, kevésbé használjuk ki. Ugyanakkor van egy startup robbanás, amely egészségügyi appokkal, hordható eszközökkel, telemedicina-szolgáltatásokkal és online közösségekkel szólítja meg közvetlenül a betegeket, polgárokat és vállalatokat. Ezek az innovációk alkalmasak arra, hogy jobbá tegyék a betegek ellátását. Az (egészség)gazdaság, a betegek és a biztosítottak szerepe alapvetően változik a digitális egészségügyi ajánlatok hatására.

Az Európai Bizottság elismerte, hogy sürgős cselekvésre van szükség, és az aktuális, egységes digitális piacról szóló stratégiai dokumentuma is támogatja az Amerikai Egyesült Államok, Dél-Korea, India, Kína és Japán elleni internet-gazdasági visszavágást [43]. Különösen az online vállalkozásoknak kell megkönnyíteni a határokon átnyúló hozzáférést, erősíteni a felhasználói jogokat az interneten, és a digitális (egészség)gazdaságot erőssé tenni. Ehhez biztosítani kell a modern, digitális európai szerzői jogot és az „európai adatbiztonsági kultúrát”.

Anyagi támogatás: A közlemény megírása, illetve a kapcsolódó kutatómunka anyagi támogatásban nem részesült.

Szerzői munkamegosztás: M. K., H. H.: Irodalomkutatás, a kézirat első változatának megszövegezése. T. T.: Irodalomkutatás, a kézirat végső megszövegezése, szerkesztése. K. M.: A Bioinformatika című fejezet megírása. D. E.: A kutatás koordinálása, a kézirat szakmai véleményezése.

Érdekltségek: A szerzőknek nincsenek érdekltségeik.

Irodalom

- [1] Turner, V., Gantz, J. F., Reinsel, D., et al.: The digital universe of opportunities: rich data and the increasing value of the internet of things. IDC, 2014. <http://idcdocserv.com/1678>
- [2] IBM: Big Data. <http://www.ibm.com/big-data/us/en/>
- [3] Ward, J. S., Barker, A.: Undefined by data: A survey of Big Data definitions. arXiv:1309.5821, 2013. <http://arxiv.org/abs/1309.5821>

- [4] Wu, X., Zhu, X., Wu, G. Q., et al.: Data mining with big data. *IEEE TKDE*, 2014, 26(1), 97–107.
- [5] Falkenberg, G., Kisker, H., Urbanski, J. (eds.): Big Data Technologies – Knowledge for Decision-Makers. [Big-Data-Technologien – Wissen für Entscheider.] BITKOM, Berlin, 2014. [German]
- [6] Michael, M., Moreira, J. E., Shiloach, D., et al.: Scale-up x scale-out: A case study using Nutch/Lucene. *Proc. IPDPS, IEEE International*, 2007.
- [7] Kieun, P.: Database technology for large scale data, 2011. <http://www.cubrid.org/blog/web-2-0/database-technology-for-large-scale-data>
- [8] Potter, J. L.: The massively parallel processor. The MIT Press, Cambridge, 1985.
- [9] Abadi, D. J., Boncz, P. A., Harizopoulos, S.: Column-oriented database systems. *Proc. Int. Conf. Very Large Data Bases*, 2009, 2(2), 1664–1665.
- [10] Bass, T.: Mythbusters: event stream processing versus complex event processing. *ACM DEBS*, New York, 2007.
- [11] Moniruzzaman, A. B., Hossain, S. A.: NoSQL database: New era of databases for Big Data analytics – classification, characteristics and comparison. *Int. J. Database Theory Appl.*, 2013, 6(4), 1–13.
- [12] Statista.com: Number of smartphone users worldwide from 2012 to 2018. <http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>
- [13] Napidroid.hu: More than 3 million smartphones on the Hungarian market. [Három milliónál is több okostelefon lehet a hazai piacon.] <http://napidroid.hu/harom-millional-is-tobb-okostelefon-lehet-a-hazai-piacon/> [Hungarian]
- [14] Albrecht, A., Naumann, F.: Schema decryption for large extract-transform-load systems. In: Atzeni, P., Cheung, D., Ram, S. (eds.): *Conceptual modeling*. Springer, Berlin, Heidelberg, 2012.
- [15] European Commission: Towards a thriving data-driven economy. COM/2014/0442, 2014. <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1404888011738&uri=CELEX:52014DC0442>
- [16] European Commission: What can big data do for you? <https://ec.europa.eu/digital-agenda/en/what-big-data-can-do-you>
- [17] The Federal Government of Germany: Digital Agenda 2014–2017. Federal Ministry for Economic Affairs and Energy, Federal Ministry of the Interior, Federal Ministry of Transport and Digital Infrastructure, 2014. http://www.digitale-agenda.de/Content/DE/_Anlagen/2014/08/2014-08-20-digitale-agenda-engl.pdf
- [18] European Commission: eCall: Time saved = lives saved. <http://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved>
- [19] Riegen, O.: Interview with Sabine Bätzing-Lichtenthäler. 2015. [Interview mit Sabine Bätzing-Lichtenthäler. 2015.] <http://www.heise.de/newsticker/meldung/Gesundheitsministerin-Telemedizin-kann-gegen-Aerztmangel-auf-dem-Land-helfen-2526942.html> [German]
- [20] EHeR Projekt: Telemonitoring. [EHeR-versorgt: Telemonitoring.] <http://www.eher-telemedizin.de/projekt/telemonitoring> [German]
- [21] Federal Ministry of Health, Germany: Masterplan Medizinstudium 2020. [Bundesministerium für Gesundheit: Masterplan Medizinstudium 2020.] <http://www.bundesgesundheitsministerium.de/ministerium/meldungen/2015/masterplan-medizinstudium-2020.html> [German]
- [22] Jin, X.: *Essential bioinformatics*. Cambridge University Press, Cambridge, 2006.
- [23] Edwards, D., Stajich, J., Hansen, D. (eds.): *Bioinformatics: Tools and applications*. Springer Verlag, New York, 2009.
- [24] Yadav, S. P.: The wholeness in suffix -omics, -omes, and the word om. *J. Biomol. Tech.*, 2007, 18(5), 277.
- [25] Celera: A unique approach to genome sequencing. *Biocomputing*, 2006. <https://www.ocf.berkeley.edu/~edy/genome/celera.html>
- [26] Davidson College: Sequencing whole genomes: Hierarchical shotgun sequencing v. shotgun sequencing. Department of Biology, Davidson College, 2002. <http://bio.davidson.edu>
- [27] Wetterstrand, K. A.: DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). <http://www.genome.gov/sequencingcosts>
- [28] Stoeser, G., Baker, W., van den Broek, A., et al.: The EMBL nucleotide sequence database. *Nucleic Acids Res.*, 2002, 30(1), 21–26.
- [29] Stephens, Z. D., Lee, S. Y., Faghri, F., et al.: Big Data: Astronomical or genomic? *PLoS Biol.*, 2015, 13(7), e1002195.
- [30] Needleman, S. B., Wunsch, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 1970, 48(3), 443–453.
- [31] Smith, T. F., Waterman, M. S.: Identification of common molecular subsequences. *J. Mol. Biol.*, 1981, 147(1), 195–197.
- [32] Altschul, S. F., Gish, W., Miller, W., et al.: Basic local alignment search tool. *J. Mol. Biol.*, 1990, 215(3), 403–410.
- [33] Burton, P. R., Clayton, D. G., Cardon, L. R., et al.: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007, 447(7145), 661–678.
- [34] Weston, A. D., Hood, L.: Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.*, 2004, 3(2), 179–196.
- [35] Trusheim, M. R., Burgess, B., Hu, S. X., et al.: Quantifying factors for the success of stratified medicine. *Nat. Rev. Drug Discov.*, 2011, 10(11), 817–833.
- [36] Simmons, L. A., Dinan, M. A., Robinson, T. J., et al.: Personalized medicine is more than genomic medicine: confusion over terminology impedes progress towards personalized healthcare. *Pers. Med.*, 2012, 9(1), 85–91.
- [37] Kühn, A., Lebrach, H.: The “Virtual Patient” system: modeling cancer using deep sequencing technologies for personalized cancer treatment. *J. Verbr. Lebensm.*, 2012, 7(1), 55–62.
- [38] Kroemer, H. K.: Personalized medicine: the state of research. [Individualisierte Medizin: Stand der Forschung.] *Jahrestagung des Deutschen Ethikrates*, 2012. <http://www.ethikrat.org/dateien/pdf/jahrestagung-24-05-2012-kroemer.pdf> [German]
- [39] Langkafel, P.: Intro Big Data for healthcare? In: Langkafel, P. (ed.): *Big Data in medicine and health economics*. [Intro Big Data for healthcare? In: Langkafel, P. (Hrsg.): *Big Data in Medizin und Gesundheitswirtschaft*. Medhochzwei Verlag, Heidelberg, 2014. [German]
- [40] Brunner, T.: Big Data from the point of view of a health insurer. In: Langkafel, P. (ed.): *Big Data in medicine and health economics*. [Big Data aus der Sicht einer Krankenkasse. In: Langkafel, P. (Hrsg.): *Big Data in Medizin und Gesundheitswirtschaft*.] Medhochzwei Verlag, Heidelberg, 2014. [German]
- [41] Wehmeier, A., Baumann, T. T.: Big Data – more risks than benefits for helath care? In: Langkafel, P. (ed.): *Big data in medicine and health economics*. [Big Data – mehr Risiken als Nutzen für die Gesundheitsversorgung? In: Langkafel, P. (Hrsg.): *Big Data in Medizin und Gesundheitswirtschaft*.] Medhochzwei Verlag, Heidelberg, 2014. [German]
- [42] Weichert, T.: Medical Big Data and data protection. In: Langkafel, P. (ed.): *Big Data in medicine and health economics*. [Medizinisches Big Data und Datenschutz. In: Langkafel, P. (Hrsg.): *Big Data in Medizin und Gesundheitswirtschaft*.] Medhochzwei Verlag, Heidelberg, 2014. [German]
- [43] European Commission: Digital Single Market Strategy for Europe. COM/2015/192, 2015. http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf

(Dinya Elek dr.,
Budapest, Ferenc tér 15., 1094
e-mail: dinya.elek@public.semmelweis-univ.hu)