

# NAGY ADAT ÉS DIGITÁLIS TÖRTÉNELEM: EGY IZGALMAS HÁZASSÁG MÚLTJA, JELENE ÉS JÖVŐJE

## BIG DATA AND DIGITAL HISTORY: PAST, PRESENT AND FUTURE OF A THRILLING MARRIAGE

Z. Karvalics László

egyetemi docens, Szegedi Tudományegyetem Bölcsészettudományi Kar Könyvtár- és Humán Információtudományi Tanszék  
zkl@hung.u-szeged.hu

### ÖSSZEFOGLALÁS

A tanulmány fő tézise szembemegy a lelkes kommentárokkal. Önmagában sem a nagy adat (Big Data), sem az az átfogó digitális transzformáció, amelynek a nagyadat-jelenség csak része, nem hoz paradigmaticus változást a történeti megismerés módjában, az új tudás szintetizálásában. Ellenben egyfajta „kreatív rombolással” úgy alakulnak gyökeresen át a történeti kutatást támogató háttérrendszerek, hogy radikálisan feljavítják a történeti munka idő- és erőforrásigényben megtestesülő peremfeltételeit – s ezzel előbb-utóbb mégiscsak soha nem volt lehetőségek sorát nyithatják meg új típusú történeti tudásműveletekhez. A szerző nyolc pontban foglalja össze, hogy a nagy adat, a digitális történelem és a crowdsourcing hol és mit változtat, *s mit nem* a történeti kutatásban.

### ABSTRACT

The main thesis of the paper is turning against the mainstream commentaries: Big Data itself (and as a part of a complex digital transformation too) does not effect paradigmatically the modes of historical cognition and the synthesis of new knowledges. But the background systems of historical research, transformed by creative destruction, are augmenting the boundary conditions of the historians' work concerning the need for time and resources. This support opens up lot of possibilities for new types of historical research production. The author summarizes the arguments in 8 points, where and what was changed and not changed in history research by Big Data and digital history methods.

**Kulcsszavak:** nagy adat, digitális bölcsészet, digitális történelem, professzionális amatőrök, crowdsourcing

**Keywords:** Big Data, digital humanities, digital history, professional amateurs, crowdsourcing

Igaza van-e a Harvard Egyetem adattudósainak, hogy a nagy adat (Big Data) forradalma, legmélyebb lényegét tekintve semmi más, mint az a különös mód, ahogyan az emberiség saját tevékenységéből immár megőrizhető történeti feljegyzést (record) teremt? Hogy ennek legfőbb következménye a saját magunkról alkotott kép átalakulása volna a társadalom természetének megismerését és megértését hatékonyabban lehetővé tévő új nézőpontokon keresztül? S vajon valóban az emberi közösségek és tevékenységek kvantifikálhatóságának új léptéke eredményezi-e a társadalom- és bölcsészettudományok radikális átalakulását (Aiden–Michel, 2013)?

Nem kétséges, hogy egészen különös élménnyel szembesül, aki a Google N-Gram Viewer grafikus felületén a kultúránk egyedi objektumait kifejező két-milliárd szó és szókapcsolat valamelyikének útját végigköveti, első megjelenésétől elterjedésén és variációinak kialakulásán át mostani előfordulásaiig. A forrás, ahonnan az adatok származnak, sem hétköznapi: a Földön ez idáig megjelent mintegy 130 millió könyv 12%-át, több mint 15 millió darabot digitalizált a (most már 30 milliónál járó) Google. Az amerikai kutatók ebből a 15 millióból választották ki tudatosan azt az ötmilliót, amely véleményük szerint a releváns információk kinyeréséhez már elégséges mennyiség (URL1). A jelentéseken keresztül feltárt egyedi kultúrtények (n-gramok) megtalálhatóvá és mintázatokba rendezhetővé tétele Erez Aidenék szerint azt is ígéri, hogy áthidalhatjuk a szakadékot, amely az emberi kreativitás produktumainak szakadatlan termelése és azok későbbi korok számára való megőrzésének képessége között kultúránk kezdetei óta tátong. S mivel ma már eleve túlnyomórészt digitális formában születnek és tárolódnak az n-gramok, mindez a jövő történészei számára jelenthet kincseshányát, amikor jelenünket közelmúltként vizsgálják majd. Régebbi korok n-gramjaihoz is közelebb tudunk azonban kerülni a digitalizálás révén, s így a humán tudományoknak esélye nyílhat arra, hogy a nagy adatnak köszönhetően „nagy tudománnyá” (Big Science) alakítsák magukat, közelebb kerülve a természettudományokhoz.

A Massachusetts Institute of Technology (MIT) ünnepelt adattudósa, Alex Pentland nem véletlenül formulázza új tudományként a társadalomfizikát (social physics). Félretéve azt a historiográfiai fenntartást, hogy számos 19. századi előfutárt követően Pentlandet legalább fél évszázaddal megelőzi az ugyanezt a tudományt hol humán termodinamikának, hol humán atomizmusnak, hol homeokinetikának nevezve „felfedező” Arthur S. Iberall is, annyi újdonság kétségkívül van ebben a megközelítésben, hogy sarkalatos állításaihoz egészen újszerű, csakis a nagy adat révén lehetőségessé váló módszerrel jut (Pentland, 2014, 2015). A társadalmi hálózatokon végbemenő információcsere mintázatait rajzolja meg ugyanis anélkül, hogy a tartalommal, a jelentéssel magával foglalkozna. Pusztán az adattá fosszilizálódott gondolatok és eszmék áramlásának (idea flow) követésével képes feltárni a legkülönfélébb méretű és helyzetű csoportok teljesítménynövekedés-

ben, adaptációban és új intézmények létrehozásában megnyilvánuló kollektív intelligenciáját.

Kétségtől eltekintve izgalmas dolog erről az oldalról is rápillantani arra a bonyolult szövevényre, amelyben a társadalmi hálózatokba kapcsolt cselekvő egyének tevékenységéből és döntéseiből történelmi tények, események, összefüggések, trendek és törvényszerűségek lesznek. Erős túlzás azonban a lelkes méltatók véleménye, hogy ez a fajta társadalomfizika forradalmian új kisablakot nyitna a társadalmi térben érvényesülő hatások és az azokat részben vagy egészben magyarázó mentális univerzumok kapcsolatának misztériumára. Ez ugyanis egy sokkal bonyolultabb, sokkal több változós problémátér, amellyel szembesülve a történészek joggal óvatosak állításaik újdonságértékével vagy megközelítéseik jelentőségével és érvényességi körével kapcsolatban.

Különös módon a nagy adat-keltette felfokozott várakozások hangnemében, de több mint száz évvel ezelőtt, a pozitivista árhullám csúcán ülve fogalmazta meg a történelmi megismerésben a lejegyzett tudásanyag elérhetőségétől remélt ugrást John Dalberg-Acton. „Egyedülálló alkalom kínálkozik arra, hogy a lehető legtöbb ember számára hasznosítható módon jegyezzük le a teljes tudásanyagot, melyet a 19. század örökül hagy. [...] Értelmes munkamegosztással ez elérhető, és ily módon minden ember hozzáférhet a fellelhető dokumentumok és a nemzetközi kutatómunka kiérlelt végkövetkeztetéseihez. Végleges történelemmel ugyan korunk még nem szolgálhat, de a hagyományos történetíráson immár túl tudunk lépni, és most, hogy már minden információ hozzáférhető, és minden problémához kezünkben a kulcs, meg tudjuk mutatni azt is, meddig jutottunk el az utóbbihoz vezető úton.”<sup>1</sup>

Korunk nagy adat-prófétái ugyanazt a hibát követik el, mint a Lord Actonként elhíresült, kiváló férfiú, a brit történetírás viktoriánus nagyja. Figyelmen kívül hagyják, hogy az 'adat' (és az általa reprezentált 'tény') csak az előszobája a történészmunkának, amelyre még számos magas rendű szellemi művelet épül, egy hosszú értéklánc részeként, azoknak a diskurzusoknak a folyamában, amelyek hipotéziseket, modelleket és értelmezési kereteket sodornak egymás mellé, hogy (narratív) konstrukciókká és rekonstrukciókká épülhessenek, későbbi továbbfejlesztésük, részleges meghaladásuk vagy teljes lecserélésük tudatában.

Ebben a tanulmányban azt kívánjuk bizonyítani, hogy önmagában sem a nagy adat, sem az az átfogó digitális transzformáció, amelynek a nagyadat-jelenség csak része<sup>2</sup>, nem hoz paradigmátikus változást a történelmi megismerés módjában.

<sup>1</sup> Idézi Carr, 1993, 7., és nyomában számos hazai szerző számos kontextusban.

<sup>2</sup> Érdeemes megjegyezni, hogy a nagy adat fogalma nem az informatikában, hanem a társadalomtudományban született meg, messze a digitális forradalom előtt. Legkorábbi előfordulása Charles Tilly 1980-as *The Old New Social History and the New Old Social History* című művében található. A fogalomtörténetet részletesen lásd Z. Karvalics, 2015.

*ban, az új tudás szintetizálásában.* Ellenben egyfajta „kreatív rombolással” úgy alakulnak gyökeresen át a történeti kutatást támogató háttérrendszerek, hogy radikálisan feljavítják a történeti munka idő- és erőforrásigényben megtestesülő peremfeltételeit – s evvel előbb-utóbb mégiscsak soha nem volt lehetőségek sorát nyithatják meg új típusú történeti tudásműveletekhez. Mindezzel már jó ideje elméletben és gyakorlatban is foglalkozik a digitális történelem (Digital History), a digitális bölcsészettudományok (digital humanities) részeként. S közben elfogadott és széles körben művelt irányzatként közel húsz éve írja saját tudománytörténetét (Rosenzweig, 2010; Galgano et al., 2012; Weller, 2012; Nawrotzki–Dougherty, 2013), egyúttal attraktív és közérdeklődésre számot tartó projektek százai is születtek szerte a világban. Ma már szinte naponta futunk bele egy-egy izgalmas kutatási eredménybe a hírekben, tudományos és szakosított történelmi ismeretterjesztő oldalakon, amelyek mögött kirajzolódik a nagy adat és a digitális technológia teljes fegyvertára. Néhány nagy port kavaró, látványos és tipikus kutatási eredmény bemutatására két ok miatt van szükségünk: egyrészt fontos, hogy azok, akik nem követik a történettudomány friss fejleményeit, képet kaphassanak jellegzetes törekvésekről, megoldásokról és eredményekről, másrészt e hevenyészett sokaság is szükséges ahhoz, hogy rendszerező állításokat tehessünk a digitális átalakulás természetéről és belső rétegzettségéről.

#### AZ ELEMI ADATMŰVELETEKTŐL A VIZUÁLIS VARÁZSLATOKIG

2002 januárjában lefagytak a brit okmányiroda szerverei. Ekkor jelent meg ugyanis az 1901-es népszámlálás anyaga: 1,5 millió adatlapról 32 millió állampolgárról kinyert és feldolgozott adattal, s olyan mennyiségű keresés érkezett egy óra leforgása alatt, amelyre senki nem volt felkészülve. (Hasonló történt az Egyesült Államokban egy évvel korábban, amikor az 1892 és 1924 között az Ellis Islanden keresztül New Yorkba érkező 17 millió bevándorlóra vonatkozó, mikrofilmen őrzött hajórakomány- és utasjegyzékek váltak kereshető formában hozzáférhetővé a weben.) Később már elővigyázatosabbak voltak az angolok: a Nemzeti Levéltár 2006-ban egy családtörténeti kutatással foglalkozó céggel kötött szerződést az 1911-es népszámlálás 8,5 millió dokumentumának digitalizálására és feldolgozására (beleértve a saját kezűleg kitöltött egykori űrlapokat).

A gyorszemlélet azért kezdtek ezekkel az esetekkel, hogy kirajzolódhassanak a határvonalak, amelyek a korábbi paradigmától (számítógép a történettudományban, kvantitatív történetírás, kliometria) elválasztják a digitális történelmet.

Amióta világossá vált, hogy a nagygépes környezet korlátainak lebomlása után elérhetővé vált mikroszámítógépek alkalmasak bármilyen adattömeg kezelésére, a nagy adatfeldolgozási szükségletekkel rendelkező történeti kutatások értelem-

szerűen fordultak a problémáikat megoldó eszköz- és módszertanok felé.<sup>3</sup> Amiben a számítógép segíteni tudott, az az adatbázis-kezelő szoftverben rögzített adatok fogadása és tárolása, visszakeresése és feldolgozása, a feldolgozás eredményének grafikonos megjelenítése. A létrejövő alkalmazott történeti adatbázisok egy-egy kutatási témát szolgáltattak ki, a bevitt végző kutatócsoport tagjai fértek hozzá. A történeti munkát az általuk feltárt forráscsoportokból kinyerhető adatok rögzítésének kiszervezése (vagy megosztása) könnyítette meg, az egyszerű hozzáférés és a feldolgozást támogató szoftveres megoldások használata pedig gyorsabbá és egyszerűbbé tette az adatokból kinyerhető összefüggések és relációk feltárását, majd az arra ráépülő elemzést. A gépről tipikusan floppyra, később CD-re mentett állományok jelentették a hordozhatóságot, vagy tették lehetővé mások hozzáférését. Az adattáblák hasznos mellékleteknek (olykor forrásközlésnek) számítottak az eredmények megosztásakor, a statisztikává párolt nyersanyag bizonyos halmazainak szerkesztett változatai pedig színes illusztrációkká tudtak válni.<sup>4</sup>

Az 1990-es évek közepétől jelentett nagy ugrást mindehhez képest az online közzététel, az internetes megosztás gyakorlata, s mögötte a feldolgozási kapacitás megnövekedése, amelynek köszönhetően a korábbiakhoz képest nagyságrendekkel nagyobb állományok váltak gépi közvetítéssel kezelhetővé. Sok történeti műhely ismerte fel, hogy milyen plusz lehetőségek rejtőznek a hálózati kultúrába tagozódásban. Az adatbázisok a gépek merevlemezeiről és a CD-kről – korlátozásokkal<sup>5</sup> vagy anélkül, ingyenes vagy fizetős eléréssel – kiszabadultak a World Wide Webre, és evvel a fókusz is elkezdett eltolódni az alkalmazott adatbázisoktól a teljes körű digitális forrásközzététel irányába. A papír-, illetve mikrofilmalapú forrásokról alkalmi projektek részeként kézi erővel adatbázisba vitt numerikus világ helyét komplett forráscsoportok (régii újságok és folyóiratok, kéziratok és levelek, fotó- és térképgyűjtemények, levéltári őrzési egységek), terv- és iparszerű, szisztematikus, központilag támogatott digitalizálása vette át, a memóriaintézményekkel az élen, amelyek felismerték, hogy a digitális állományok részben

<sup>3</sup> A részletekre magyarul lásd Evan Mawdsley és Thomas Munck könyvét (1996), és benne a Benda Gyula és Halmos Károly által összeállított szakirodalmi szemlét (*Számítástechnika, kvantifikáció és történelem. A magyar irodalom bibliográfiája*. 279–286.).

<sup>4</sup> Ennek remek példája a görög poliszok társadalomtörténetének Németh György vezetésével Apple Macintosh számítógépre készült adatbázisa (Németh, 1998), amely a felhajtóerőt biztosította a téma szintetikus összegzéséhez (Németh, 1999), és számtalan kisebb közleményhez.

<sup>5</sup> A Kominternnek az Orosz Állami Társadalom- és Politikátörténeti Archívumban őrzött 22 millió oldalas iratanyagának egymillió, leggyakrabban használt darbjáról készült nemzetközi együttműködéssel digitális másolat. Ezekhez egy ideig csak a moszkvai anyaintézet olvasótermében, később a partnerintézmények (köztük a latin átírást végző amerikai Kongresszusi Könyvtár) dedikált munkaállomásain lehetett ingyenesen hozzáférni (Magyarországon 2003 vége óta az Open Society Archive-nál, OSA). Az internetes megosztásról azóta folynak a viták, de a gyűjtemény még nem nyílt meg a hálózati polgárok számára.

mentesítik az őrzött példányokat a felhasználói terheléstől<sup>6</sup>, részben új utat nyitnak a tartalomban való kutakodásnak.

A digitalizálás és a szövegek (később: képek) kereshetősége olyan technológiai lehetőségekkel gazdagították a korábbi gyakorlatot, amelyek szintén csak a kilencvenes évek közepén váltak elérhetővé. A források digitális kópiája az esetek legnagyobb részében az eredeti állományok teljesen identikus, információvesztés nélküli megismerését teszi lehetővé. A releváns tartalom megtalálásának pedig ugrásszerűen megnő az esélye: részben az állományok pusztá beolvasásán túlmenő szakszerű, szabványokon alapuló *metaadatolásnak*, részben a szövegekben megnyíló kulcsszavas keresésnek köszönhetően. (Emiatt kellett sokszor a sebtében és ötletszerűen beszkenneletett állományokhoz másodszor vagy harmadszor is hozzányúlni, hogy megfelelő formátumban készült változatokkal lehessen kicserélni és használhatóvá tenni azokat.)

A jó metaadatok révén különböző őrzési helyeken tárolt állományok vonhatók össze virtuális gyűjteményekké (ilyenek például a – főleg középkori anyagra szakosodott – képkönyvtárak) (lásd URL2 vagy URL3), és a forrásokhoz linkkompozíciókkal kapcsolható a könyv vagy tanulmány formájában elektronikusan elérhető elsődleges szakirodalom.

A 'nagy adat' azonban nem írható le és érthető meg pusztán nagy „mennyiségként”. Legalább annyira lényeges, hogy az állomány hatalmas sebességgel növekszik (bizonyos területeken a képződéssel párhuzamosan, valós időben követhető és feldolgozható módon), s a létrejövő állomány elemeinek kapcsolódása pedig akkora mennyiségű variációt eredményez, hogy magasabb rendű összefüggések kinyeréséhez, felismeréséhez és következtetések levonásához új feldolgozási formákra van szükség. A technológiai fejlődés ma már elsősorban ezt a legutolsó szakaszt támogatja, hiszen az adattá tétel és az adatkezelés eljárásai régóta konszolidálódtak. Az újdonság így egyre inkább abban van, miképpen jeleníthetők meg úgy az adatok, hogy annak heurisztikus értéke legyen a megismerő elme számára. A nagy adat forradalma elválaszthatatlan a vizualizálás elegáns, látványos és rohamtempóban fejlődő alaprendszereitől. Elsősorban az információk térképes ábrázolására (a földrajzi információk rendszerek más adatkörökkel való termékeny összekapcsolásával), a háromdimenziós adatmegjelenítés eszközeinek az animációval való összekapcsolására, a kronológiai elrendezés konvencionális időegyenesen túllépő számtalan variációjára és az interaktív felületekre gondoljunk. Egészen különleges felhasználói élmény, amikor ki-ki egy egységben látott állomány számára releváns részletére „zoomolva” csak azt emeli ki és jeleníti meg, amire neki szüksége van. A szöveghez, a történelmi munka hagyományos narrációjához képest felértékelődik a kép, és minél inkább makrostruktúrákat jel-

<sup>6</sup>Norvégia például minden más országot megelőzve tűzte ki stratégiai célként, hogy teljes levéltári gyűjteményét minél hamarabb digitalizálja.

lemeznek az adatok, annál inkább az arányok, a ritmusok, a rejtett szerkezetek megelevenedése a megismerő élmény forrása. Különösen érdekes a vizuális rekonstrukciók, hajdanvolt települések, épületek megelevenítése grafikai eszközökkel, Trójától az alexandriai könyvtárig.

Az ismerkedést olyan anyagokkal érdemes kezdeni, mint a University of Dallas kutatóinak ötperces filmje az emberiség 2600 évének kulturális centrumairól – amelyek tér- és időbeli váltakozása 120 ezer „történelmi személyiség” születési és halálozási helyszínének adatbázisba építésével vált mozgóképes formában megjeleníthetővé, olyan izgalmas tanulságokra való rácsodálkozással, hogy Párizs csak 1789-ben vette át Rómától a csomóponti szerepet (URL4). Vagy a közelmúltban elhunyt Hans Rosling Gapminderének (bárki által használható statisztikageneráló oldalának) talán leghíresebb videoetűdje, a *200 év, 200 ország, 4 percben*. Ebben a rendkívül látványos és meggyőző animált infografikai csúcsteljesítményben a kép és az emberi magyarázat együtt teljesíti ki a különleges élményt. Egyetlen mozgóképben kel életre az az elképesztő statisztikai anyag, amely az átlagéletkorra és jövedelemre vonatkozó adattömegből megrajzolja az egyes országok és országcsoportok jóléti mozgását 1800-tól napjainkig (URL5). S korábban szintén elképzelhetetlen volt olyan méretű kapcsolati hálózat feltárása és megjelenítése is, mint a kora modern (16. és 18. század közötti) Anglia 13 ezer nevezetes szereplője között megrajzolható 200 ezer összeköttetés (URL6) – vagy az e sorok írásakor 12 217 személy leveleit tartalmazó Early Modern Letters Online (EMLO), amelyre az Oxfordi Egyetem Cultures of Knowledge (Networking the Republic of Letters, 1550–1750) és a Stanford Egyetemnek az összefüggéseket térképen ábrázoló Mapping the Republic of Letters című projektje épül (URL7, URL8).

Ahhoz azonban, hogy a nagy adat és a digitális történelem metszéspontján kirajzolódó leglényegesebb változásokhoz közelebb kerülhessünk, újra hozzá kell fognunk a történetmeséléshez, de immár egészen máshonnan kezdve.

#### (ÉLET)IDŐ, KUTATÁSI ERŐFORRÁSOK ÉS SZAKMASZOCIOLÓGIA. A TÖRTÉNÉSZ ÉS AZ Ő MUNKÁJA

„Ne higgyetek a történetírásnak, ha nem a legkivételesebb szellemekből fakad, [...] (de) ne becsüljük le a történettudomány fuvarozó, főlhalmozó, rendezgető munkásait csak azért, mert bizonyára nem válhatnak nagy történészzé.”

(Nietzsche, 1989, 66.)

A nagy történészeket és régészeket mindig egyedül látjuk a fényképeken, ahogy galambósz (jellemzően hátrafésült) hajjal, kedves könyveik, kézírataik és tárgyaik között szigorúan tekintenek a nyugalmukat megzavaró kamerába. Néha feltűnnek a cédulásdobozok is, amelyekben az évtizedek alatt hangyaszorgalommal

összegyűjtött feljegyzéseik pihennek. Szinte kézzelfoghatóan érezzük, hogy amikor a tudós megszűnik majd alkotni, óriási űr keletkezik a helyén: hosszú időnek kell eltelnie még, amíg valamelyik tanítványa majd a helyére lép.

A történezmesterség mindig is feltűnően magányos, individuális tevékenységet jelentett. A levéltár vagy a könyvtár csendjében, kedves segítőkkel, de mégiscsak egyedül. Szeretett Mester(ek) és legendás szemináriumok ugyan szinte mindig akadnak az egyetemen (aki/k/re még sokáig jólesően és tiszta szívből lehet hivatkozni), de a forrásokkal folytatott csatában és az alkotás, a narratívák összeállításának pillanataiban a kutató tartósan és jellemzően önmagára van utalva. A szakmai találkozók, a tanulmányok és a könyvek megjelenése alkalmából elinduló viták, a Mesterré válással megjelenő igény a tanításra, a tudás átadására és felépítésére – mindezzel rövid időre megjelenik ugyan a történezt körülvevő közösségi tér is, de ennek jellemzően nem az új tudások létrehozásában, hanem az új tudások létrehozásához elengedhetetlenül szükséges szakmai közélet megteremtésében és életben tartásában van szerepe.<sup>7</sup>

Ez a békeidőket felidéző, részben idilli kép a múlté. Az a korszakváltás, amellyel a 19. század megteremtette az egyszerű eseménytörténet helyére lépő problémaközpontú elemzés<sup>8</sup> (s ennek révén a téren és időn keresztül is diskurzusképes, egymásra épülni tudó történeti vizsgálódások) szakmai-módszertani standardjait<sup>9</sup>, eltörpül a 21. század elején, szemünk láttára formálódó minőségi váltás mögött. A nagy adat- és a digitális eszközkészlet ugyanis nemcsak a történelmi munka minden eddiginél hatékonyabb infrastruktúráját kínálja: segítségükkel az *a mód* alakul át, ahogyan a történeti kutatás alapjául szolgáló forrásaink elrendeződnek, ahogyan kereshetőek, ahogyan viszonyulunk hozzájuk, ahogyan értékeliük őket, és ahogyan további feldolgozásra elrendezhetőek. Ráadásul avval, ahogyan egyre több forrás, dokumentum és szakirodalom áll rendelkezésre az azonnaliságot biztosító *online* térben, ezek elérhetővé tétele, áttekintése, előfeldolgozása, megfelelő forráskritikai apparátussal való ellátása, az indexelt hozzáférés folyamatos biztosítása *hatalmas emberi erőforrásigényt*

<sup>7</sup> Szépen mesél erről Georges Duby és Guy Lardreau (1993, 36.): (a nyomok mellőzése és felszínre hozása) „nem egyetlen egyén ítélete, hiszen nyilvánvaló, hogy az én tevékenységem nem különülhet el néhány kortársamtól, elődömetől, akikkel tökéletes nézetazonosság köt össze. Nemcsak azokról van szó, akik körülöttem végeznek kutatásokat, akikkel hetente találkozom, hanem másokról, tőlem igen távol dolgozó pályatársakról is. Nemcsak közvetlen tanárainkkal [...] vállalom a közösséget, hanem többi mesteremmel is. Választásom nem szabad... én is egy közösségbe tartozom...”

<sup>8</sup> Ami természetesen nem jelenti azt, hogy Ibn Khaldún vagy Condorcet ne lépett volna fel evvel az igénnyel – csak azt, hogy nem tudtak diskurzusképzővé válni, életművük egyfajta historiográfiai torzó.

<sup>9</sup> Tegyük hozzá: a szintén a 19. században létrejövő szakmai testületek és társaságok is magát a szakmát, illetve annak normáit és infrastruktúráját termelik. Mindennek a történelmi tudás termelése felőli korai megközelítésére lásd Jacob Burckhardt (1943, 25–30.).



teremt. A digitalizálással, metaadatulással, webprogramozással kapcsolatos munkákhoz, majd a bevitel, az áttekintés, az elolvasás, a feltárás vagy a kivonatolás műveleteihez nagyon sok eszközre és kutatói életidőre van szükség – és akkor még nem szóltunk a rendezés, a feldolgozás, az értelmezés, az értékelés, kontextusba helyezés feladatairól. Az új eredmények mögött így egyre inkább a hagyományos földrajzi és szakmai határokon átívelő kutató közösségek (Lewis Mumfordot parafrázálva: tudománytermelő *megagépezetek*) állnak, magát az alkotás folyamatát is kooperatív tevékenységgé változtatva, a szakmai információk mind hatékonyabb cseréjével és folyamatosan újrendezett munkamegosztással.

Mindez valóban arányban áll a megoldandó feladatok nagyságával. A figyelembe veendő források, adatok, szempontok, szakirodalmak és rokon területek óriási száma miatt valódi (vagy akár csak részleges) feldolgozási teljességig a legnagyobb történészek egy élet munkájával is csak ritkán tudtak eljutni – miközben a legnagyobb, leginkább maradandó monográfiák és tanulmányok a megformáltság, a „megoldottság”, a megszüntetett kérdőjelek, a megnyitott dimenziók élményével természetesen ugyanúgy a kerek Egész érzését közvetítették, mint mondjuk a műalkotások.<sup>10</sup> S noha egy-egy nagy összegzést zajosan ünnepelt a szakmai és „társasági” közélet, az eredendő, lényegi hiányérzet megformálására (sokszor ugyanazoktól a szerzőktől) a történetfilozófiai reflexiók és a történetpszólya végén írott módszertani útravalók, illetve számvetések rejtett vonulataként került sor. A történész legszűkösebb erőforrása ugyanis – kimondva-kimondatlanul – mindig *az idő* volt: a levél- és könyvtári források felkutatása, feldolgozásra előkészítése, (kézirásos) másolása. A releváns részek keresése, rögzítése, visszakeresése. A megtalált (sokszor óriási mennyiségű) szöveg, kép, tárgy elolvasása, végignézése. A problémák megoldásához esélyt kínáló társtudományok irodalmának fürkészése. A teljesítmény garanciájává (az invención felül) így vált a történész szorgalma, kitartása, a kutatásra fordított idő mértéke.

Technológiai oldalról – bármilyen meglepő is – a 80-as évek közepétől nem is a számítógép, hanem még a tömegszolgáltatássá lett *fénymásolás* megjelenése kezdte enyhíteni a levéltári és könyvtári munka keserveit. Az ekképpen szerzett időnyereség teljesítménynövekedésbe fordult: a „megnyert” időt szélesebb spektrumú forráskeresésre lehetett felhasználni. A *számítógép* esetében pedig a kvantitatív kutatások során a nagy mennyiségű adat feldolgozhatóvá tétele váltotta ki azt az időt, amelyet a hagyományos „kézi” módszerekkel kellett volna a gondozásukra fordítani. A rendkívüli méretű digitális állományokon végzett szöveg-bányászati-analitikus eljárások esetében azonban már nem csak az időnyereség

<sup>10</sup> „Mint a nagy műalkotást, a nagy történelmi művet is átélheti más ember” (Braun, 1922, 82.)  
Kicsit később: a történelem újra átélése szoros rokonságban van a világ művészi felfogásával (88.).

a célfüggvény: ezek új összefüggések felismerését teszik lehetővé hagyományos kutatói megközelítéssel nehezen felderíthető anyagon. Ehhez azonban csoportmunkára van szükség. Ami nem volt elvégezhető nagy, individuális tetteként, az most lépésről lépésre megvalósíthatóvá válik az azonos kérdésekkel foglalkozó és tevékenységüket munkamegosztás-szerűen összehangoló történészek és a mind professzionálisabb háttértámogatást nyújtani képes kutatási asszisztencia összekapcsolásával kiterjesztett cselekvési térben. De nem csak egy megnövelt méretű és a korábbi időknél kooperatívabb történészcéh tagjai a részesei az új tudások előállításának. Onnan is érkeznek a segédcsapatok, ahonnan a 19. század nagy pozitívista iránymutatói elképzelni sem tudták volna.

### MICROTASKING, CROWDSOURCING – ÚJ ERŐFORRÁS A TÖRTÉNETI KUTATÁSBAN

„Mint mindig, most is egy csoport középpontjában vagyok, a csoporttal együtt haladok előre [...] már ők teszik fel a kérdést. Nem nekem, hanem nekik van esélyük arra, hogy a kérdést megválaszolják.”

(Duby, 2000, 155.)

A nagy adat világában a számolásteljesítményt elvégzi a brutális processzorero. De hogyan lehet minden egyes rekordhoz értelmezésre képes aggyal összekapcsolt emberi szemet rendelni? Miképp lehetne elvárni, hogy a szaktörténészek egyenként szembesüljenek a gigászivá nőtt adatuniverzum megszámlálhatatlan egyedi objektumával? Hogy kódolják, értelmezzék, ellenőrizték, metaadatulják, más adatokkal összekapcsolják őket?

Ott, ahol egy nagy volumenű feladat elvégzése sok-sok kicsi részfeladatra bontással megoldható, nem egy embernek kell embertelenül sok feladatot elvégeznie, hanem nagyon sok ember teljesít fejenként csekély mennyiségű munkát. Ez a *microtasking* világa, ahol a hálózati térben a feladat elvégzésétől remélt előrelépés ígéréteben önkéntesek ezrei (nagyobb projektek esetén milliói) kínálják fel saját életidejük kis részét, hogy számukra fontos ügyeket előrevigyenek. A más tudományterületeket is mind markánsabban jellemző tömeges erőforrásszerzés, a *crowdsourc*e a történeti tudás termelésének értékláncában szinte minden ponton feltűnhet.

Önkéntesek vesznek részt dokumentumok, fényképek vagy akár folyóiratok digitalizálásában. (Egyetlen nyugdíjas amerikai mérnök a Kongresszusi Könyvtár teljes folyóirat-digitalizálási teljesítményének harmadával bír.) Mindennek neve is van már: *civil archiválás*. Históriaját 2009-ig szokás visszavetni, amikor az Ausztrál Nemzeti Könyvtár a nemzeti történelmi dokumentumok interaktív portáljának (Trove) építéséhez programszerűen vette igénybe a hálózati polgárok segítségét. Azóta száznál is több ilyen projekt indult már, az önkéntesek tipikusan az archívumok építésében, a meglévő gyűjtemények

kiegészítésében, a hibaigazításban és a metaadatulásban segítenek. A leginkább jellemző a nyilvánossá tett iratok tömegének feldolgozásában nyújtott segítség. A British National Archives 1,5 millió oldalnyi feljegyzés közzétételére vállalkozott a honlapján, kivétel nélkül az első világháborúban harcolt katonák feljegyzéseiből, akik sok dokumentumot kézzel, a fronton írtak meg. Az archívum önkénteseket kért fel, hogy segítsenek a történészeknek feldolgozni (sok esetben a gyorsírás, illetve az olvashatatlan kézírás miatt: egyáltalán elolvasni) a naplókát és leveleket, hogy minél több ember megismerhesse a fronton harcoló rokonát. A *Moving Here* oldalon harminc helyi, regionális és országos brit múzeum gyűjteményéből több mint 150 000 forrásanyag érhető el digitálisan, és a gyűjtemény azáltal is gyarapodik, hogy a látogatók feltölthetik saját családi emlékezetük témába vágó darabjait. A levéltárakba zárt feldolgozatlan kéziratok egyes oldalainak „kiosztását” segítő, bármilyen kutatási programhoz ingyenesen igénybe vehető szoftvereszközt az „Írd át Benthamet” (Transcribe Bentham) projekt oldalvén fejlesztették ki, amelyik 2016. április 18-án 15 788, önkéntesek által elolvasott és számítógépes szövegszerkesztőbe bevitt, ellenőrzött kéziratoldalnál tart. Magyarországon az OSA, a Nyílt Társadalom Archívuma reméli a második világháború „csillagos házai” történetének megismerését az érintett lakosság adatközléseinek a segítségével előbbre vinni (a 2000 épületből 1600 ma is létezik) (URL9, URL10, URL11).

A webre feltöltött műholdképek százazreire vagy a Google Earth megadott szelvényein belül fürge önkéntesek ezrei próbálnak megtalálni nevezetes objektumokat (az érdeklődők látogassanak el a Field Expedition: Mongolia – Valley of the Khans Project oldalára [URL12], hogy képet kaphassanak arról, hogyan működik a Lengyelországtól Japánig terjedő mongol birodalomhoz köthető, ma is látható felszíni képződmények – egykori folyók, utak, ősi építmények – azonosítása).

Ez a lendület nemhogy nem látszik kifulladásra, hanem szinte naponta kapunk híreket régi és új szereplők friss projektjeiről. Az egyelőre láthatatlan (már létező, de digitálisan elérhetetlen) úgynevezett „sötét adat” (dark data) birodalmából is újabb óriásállományok felszabadulása és megosztása remélhető, sokszor a legváratlanabb helyekről. Ilyenek voltak legutóbb például a klímátörténészek számára felbecsülhetetlen jelentőségű 1964-es képek a Földről, amelyeket a NASA Nimbus nevű műholdja készített. A mágnesszalagra rögzített felvételeket tévéképernyőn játszották egykor vissza, s ekkor készítették fényképeket róluk, amelyeket a sorozatjelleg miatt filmszalagra vittek. Az azóta is ömlesztve, rendezetlenül tárolt anyag talán soha nem lett volna közkinccsé, ha egy elkötelezett kutató, David Gallagher nem szervezi meg az irtózatossá adatmennyiség digitalizálását. A több mint 200 ezer fotó ma szabványosított formában hozzáférhető és kereshető. (Jellemző, hogy ez a több gigabájtnyi anyag elfér egy pendrive-on, de a korban nem volt akkora tárolási kapacitás, amely képes lett volna kezelni, ezért kellett a fotókat elkészíteni; URL13.)

## ÖSSZEGZÉS, KITEKINTÉS

Foglaljuk össze, immár pontokba szedve, hogy a nagy adat, a digitális történelem és a crowdsourcing hol és mit változtat *s mit nem* a történeti kutatásban.

1. A korábbi kutatási feltételrendszerben elképzelhetetlen méretű adattömeg egyidejű kezelése vált praktikusán megoldhatóvá, a megértést szokatlan vizuális erővel támogató megjelenítési eljárásokkal karöltve. Mindez egy új kooperatív térben zajlik: kutatóközösségek határokön átívelő munkamegosztásával és önkéntesek tömegének bekapcsolásával kialakuló „tudástermelő óriásgépek” létrejöttével.

*Mérlege:* az új kutatási eredményekhez való eljutásnak egy új osztálya jött létre, a jelen történetének kutatása pedig olyan instrumentumokkal gazdagodik, amelyek későbbi korok történészei számára majd különleges felbontásban teszik lehetővé a retrospekiót.

2. A kutatásokat tervező és lebonyolító történészi munka szempontjából az egyéni eredmény eléréséhez szükséges időteher nagy része átkerül a gépi számolásteljesítmény és a spontán munkaszervezetbe tömörült emberi elmék oldalára.

*Mérlege:* részben megnő az immár reálisan tervezhető kutatások száma és formagazdagsága, másrészt a történészi munkaidő felszabadul a nagyobb hozzáadott érték, a magasabb rendű műveletek (absztrakciók, kontextusok, összehasonlítás, értékelés, diskurzusképzés) számára.

3. A memóriaintézmények forrástömegének esetében, s főleg az ókori és középkori anyagnál messzemenően felértékelődik a „klasszikus” forráskritikai és segédtudományi (a digitalizáláshoz szükséges magas szintű diplomatikai, paleográfiai, idegen nyelvi, ikonológiai és más) tudás.

*Mérlege:* a digitalizálás intenzív szakaszában a jelenlegi kibocsátásnál nagyobb mértékben volna szükség felkészült szakemberekre, hogy a következő tudósgeneráció megbízhatóan rendszerezett, feltárt és bővíthető állományokra mint alapra építhesse feldolgozó munkáját. Ennek messzemenő tudomány- és felsőoktatás-politikai következményeiről azonban még nem indult el a párbeszéd.

4. Az új emberi és gépi erőforrások megjelenése elkezdte szolidan átrendezni azt a hatalmi teret, amely részben a tudománypolitikával párosult fiskális eszközök, részben az intézményi és tekintélystruktúrák révén meghatározta azt, hogy mikor, milyen kutatási programok és milyen történészek által irányítva élveznek támogatást.

*Mérlege:* a társadalomnak a történeti kérdések iránt érzékeny és elkötelezett tagjai a történeti eredmények pusztá fogyasztóiból a kutatások tematizálójává

(agenda setter) tudnak válni azáltal, hogy milyen programok mögé sorakoznak fel. Evvel összességében erőforrás-bevonás történik a történettudomány művelésébe, amely megnöveli az alternatív, kísérleti programok elindíthatóságának esélyét is.

5. A korábban a társadalomtól elkülönült, arisztokratikus *céh*be tömörült történészközösség és a történelemfogyasztó nagyközönség közti határvonal így számos formában elmosódik. Közelebb kerül egymáshoz családtörténet és köztörténet, nemzeti és családi emlékezet. A személyes és a köztörténeti vonatkozások.

*Mérlege:* a részvétel kultúráján és ethoszán keresztül a társadalomban újfajta (részben emancipatorikusnak tekinthető) érzékenység bontakozik ki a történelmi múlt iránt, s ez a tudományos műhelyeket egyúttal arra is készíti, hogy a korábinál is hangsúlyosabban formálják meg eredményeiket a tudományos ismeretterjesztés csatornáin.

6. A digitális eszköztár, az attraktív megjelenítés és a népszerűsítő történelmi tárgyak megállíthatatlan beáramlása a hálózati világ népszerű tartalomgyáraiba oda vezet, hogy a történelem tanítása és tanulása eltolódik a hagyományos (premediális, frontális) műfajoktól, s a történelmi tudat formálódását alakító csatornák, szereplők és hatások száma is megnő (Kee, 2014).

*Mérlege:* a történelemhez való viszony, a történeti gondolkodás is átalakul. (Mindennek az is része, hogy a számítógépes és hálózati forradalom története mint megélt közelmúlt már maga is történeti tárgy.)

7. A fentiek együttesen azt ígéri, hogy *bizonyos kutatási kérdések esetén* a műkedvelő és a szakember közti határok is máshová kerülnek. Egykor a helytörténetírásban vagy a gyógyítás- és technikatörténeti tárgyak kapcsán a minőségi kritériumoknak megfelelni tudó iskolaigazgatók, papok, orvosok és mérnökök minden további nélkül válhattak teljes értékű történeti művek létrehozóivá, a „hobbytörténészek” kasztjából kiváló professzionális amatőrre (professional amateur, ProAm). Ma egy egészen más történeti szocializációs térben felnövekvő nemzedék egyre több tagja számára nyílik meg az esély, hogy ne csak önkéntessé, hanem megfelelő kompetenciák megszerzése esetén az állampolgári tudomány (citizen science) képviselőjévé is váljon.

*Mérlege:* a történeti kutatások világába az elemi szintű adatműveleteken túl is vezethet út. A nyilvánosság, a megosztás és a participáció új közösségi kultúrájában a párbeszéd, a vita, a kritika, a hipotézisgenerálás és a hipotézisvizsgálat is megnyílik olyanok számára, akiknek eddig ez nem volt lehetőségként adott (Rosenzweig, 2010). A történész szakma demokratizálódik, a szakmai közélet új formái jönnek létre.

8. *Nem változik* a szakmai standardoknak az a hosszú idő alatt kiformált együttese, amelyekhez megfelelő előképzettséget társítva az egyetemi és akadémiai minőségbiztosítás rendszerén átjutóknak, a történész munkát életpályaszerűen folytatóknak igazodniuk kell. A forráskutatás, feltárás, kezelés rutinműveletei, a filológiai pontosság, a forráskritika, a verifikálhatóság és falszifikálhatóság kritériumának való megfelelés, a meghaladni kívánt diskurzusok elmélyült (historiográfiai) ismerete, a kellő mélységben feltárt szakirodalomra épülő következtetések, az állítások plauzibilitásának vizsgálata és számos más, a gyakorlat sodrában normaként kialakult eljárás továbbra is az új történeti ismeret megtermelésének aranyfedezetét jelenti. Ezek a minőségi követelmények mindenkire vonatkoznak, a tudományos újságíróktól a professzionális amatőrökig.

*Mérlege:* A történelmet hivatásrendként művelők esetében a tények kompozícióba rendezése, a szisztematizálás, a táguló témauniverzumok monográfiákkal való időszakos összebroncsolása, új modellek, absztrakciók és kontextusok keresése továbbra is az egyéni szorgalom, tehetség és invenció kérdése marad – ám a közösségi beágyazás és a megosztás erősödő imperatívuszával.

## IRODALOM

- Aiden, E. – Michel, J.-B. (2013): *Uncharted: Big Data as a Lens on Human Culture*. Riverhead
- Braun, O. (1922): A történettudomány ismeretértéke. In: Braun, O.: *Bevezetés a történetfilozófiába. (Filozófiai Könyvtár)* (ford. Pukánszky B.) Budapest: Pfeifer Ferdinánd (Zeidler testvérek) Nemzeti könyvkereskedésének kiadása, [http://mtdaportal.extra.hu/books\\_kulf/braun\\_otto\\_bevezetes\\_a\\_tortenetfilozofiaba.pdf](http://mtdaportal.extra.hu/books_kulf/braun_otto_bevezetes_a_tortenetfilozofiaba.pdf)
- Burckhardt, J. (1943): The Qualifications of the Nineteenth Century from the Study of History. In: Burckhardt, J.: *Reflections on History*. London: George Allen & Unwin Ltd.
- Carr, E. H. (1993): *Mi a történelem? (Századvég Könyvtár)* (ford. Bérczes T.) Budapest: Századvég–Osiris Kiadó
- Duby, G. (2000): *Folytonos történelem.* (ford. ifj. Benda K.) Budapest: Napvilág Kiadó
- Duby, G. – Lardreau, G. (1993): *Párbeszéd a történelemről.* (ford. Szilágyi G.) Budapest: Akadémiai Kiadó
- Galgano, M. J. – Arndt, C. – Hyser, R. M. (2012): *Doing History: Research and Writing in the Digital Age.* Cengage Learning; 2<sup>nd</sup> edition
- Kee, K. (2014): *Pastplay: Teaching and Learning History with Technology.* Ann Arbor: University of Michigan Press
- Mawdsley, E. – Munck, T. (1996): *Számítógép a történettudományban.* (ford. Turi László) Budapest: Osiris Kiadó
- Nawrotzki, K. – Dougherty, J. (2013): *Writing History in the Digital Age. (Digital Humanities Series).* Ann Arbor: University of Michigan Press, DOI: 10.3998/dh.12230987.0001.001, <https://quod.lib.umich.edu/cgi/t/text/text-idx?cc=dh;c=dh;idno=12230987.0001.001;rgn=full%20text;view=toc;xc=1;g=dculture>
- Németh Gy. (1998): A POLIS-program. *Ókortudományi Értesítő*, 2, 33–34.
- Németh Gy. (1999): *Polisok világa. Bevezetés a görög archaikus és koraklasszikus kor társadalomtörténetébe.* Budapest: Korona Kiadó

- Nietzsche, F. (1989): *A történelem hasznáról és káráról.* (ford. Tatár Gy.) Budapest: Akadémiai Kiadó
- Pentland, A. (2014): *Social Physics: How Good Ideas Spread - The Lessons from a New Science.* Penguin Press
- Pentland, A. (2015): *Social Physics: How Social Networks Can Make Us Smarter.* Penguin Press
- Rosenzweig, R. (2010): *Clio Wired: The Future of the Past in the Digital Age.* Columbia University Press
- Weller, T. (2012): *History in the Digital Age.* Routledge
- Z. Karvalics L. (2015): A nagy adat-jelenség társadalomtudományi lehorgonyzásához. *Replika*, 3–4, 92–93., 189–202. [http://replika.hu/system/files/archivum/92-93\\_14\\_zkl.pdf](http://replika.hu/system/files/archivum/92-93_14_zkl.pdf)
- URL1: <https://datascience.berkeley.edu/erez-lieberman-aiden-and-jean-baptiste-michel-what-we-learned-from-5-million-books/> (Letöltve: 2016. 04. 14.)
- URL2: [www.bl.uk/catalogues/illuminatedmanuscripts](http://www.bl.uk/catalogues/illuminatedmanuscripts)
- URL3: <http://utu.morganlibrary.org/index.htm>
- URL4: <http://www.nature.com/news/humanity-s-cultural-history-captured-in-5-minute-film-1.15650>
- URL5: <https://www.youtube.com/watch?v=jbkSRLYSojo>
- URL6: <http://sixdegreesoffrancisbacon.com/>
- URL7: <http://www.culturesofknowledge.org/>
- URL8: <http://republicofletters.stanford.edu/>
- URL9: <http://blogs.ucl.ac.uk/transcribe-bentham/>
- URL10: <http://www.movinghere.org.uk>
- URL11: <http://www.csillagoshazak.hu/>
- URL12: <http://exploration.nationalgeographic.com/mongolia>
- URL13: <http://barentsobserver.com/en/2014/10/american-scientists-unearth-lost-1960s-polar-satellite-images-worth-billions-14-10#VD1bCUcp6oU.twitter>