

A CYBER-INFRASTRUKTÚRA MINT AKTUÁLIS KIHÍVÁS ÉS MINT TUDOMÁNYSZOCIOLÓGIAI PROBLÉMA

Z. Karvalics László

a történettudományok kandidátusa, Budapesti Műszaki Egyetem Gazdaság- és Társadalomtudományi Kar
Információ-és Tudásmenedzsment Tanszék; George Washington University, Georgetown University
zkl@itm.bme.hu

(mi, tudósok) *Ellőtt nyilak vagyunk.
A mi agyvelőink fákhelyék, melyeket egy
istenség a másik istenség sötétjébe hajít.*
Balázs Béla: *Doktor Szélpál Margit*

2005 júliusában 41 neves tudós, a *Science 2020 Group* (főleg angol, kisebbrészt amerikai és francia) tagjai három intenzív napot töltöttek Velencében azzal, hogy a tudomány szerepével és jövőjével kapcsolatos nézeteiket vízióvá egyesítsék. Ezt követően több hónapos szerkesztőmunkával elkészült, és 2006 elején nyilvánosságra is került a '2020 tudománya felé' (*Towards 2020 Science*) című dokumentum, amely az elkövetkező tizennégy év legfontosabbnak tartott tudományos trendjeit és feladatait rendezte áttekinthető jelentéssé (Emmot et al., 2006).

A jelentés megállapításai túllépnek Paul A. David igazodási pontnak tekintett téziséen, miszerint az új digitális eszközvilág gyökerelesen átalakítja (profoundly alters) azokat a módokat, ahogyan a „normál” tudományos programok szerveződnek, támogatást kapnak, és lefutnak (David, 2000). A szerzői kollektíva ugyanis egyenesen azt állítja, hogy

az egyes diszciplínákkal összeolvadó információtechnológia *forradalmi* változást hoz, a *tudomány radikális átalakulását* eredményezi. A számítógépeknek, a hálózatoknak és a digitális eszközparknak szoftverestül, alkalmazásostul immár nemcsak metasztintú a hozzájárulása a jövő tudományához, hanem *tárgyszintű* (object level) is. Nem pusztán segítségül hívják valamely probléma kezeléséhez, hanem *a számítástechnika fogalmai, módszerei és tételei szervesen beépülnek az adott tudományterület művelésének szövetébe, új szakmai „minőségeket” hozva létre.*¹

Leginkább kiteljesedett módon mindez a biológia, a biotechnológia, a biokémia és a gyógyszer tudomány területén látszik, de az integráció meggyőző a környezeti tudományokban és az asztronómiában is. A biológiai tudományoknak a korszerű információtechnológiával való egyfajta „házasságát” sokol-

¹A jelentés „tartalmi párjának” tekinthető a *Nature* 2006. március 23-i száma (Vol. 440), amelyben *Computing 2020* cím alatt kilenc tematikus tanulmány foglalkozott a számítástudomány jövőjével, és ezt valamennyi a tudomány egy-egy dimenziójában találta meg. (Részletes bemutatásukat lásd Jéki, 2006)

dalúan elemző Rick Stevens pedig egyenesen azt állítja, hogy miközben a korszerű szimulációs és modellező eljárások révén máris prediktív elméleti felismerésekig lehet eljutni, a jelenlegi sejszintű modellektől hamarosan az organizmusszintűekig ugorva, az új biológiai felismerések legfőbb korlátjává mindinkább a számítógépes háttérrendszerek esetleges teljesítményelégtelensége válik (Stevens, 2006).

Nem véletlen, hogy a tudomány igényei érdekében fejlesztett informatikai közművek (information technology utilities) iránti érdeklődés az elmúlt években erősen megnőtt, és stratégiai kérdéssé is vált mind az Európai Unió, mind az Egyesült Államok, mind a délkelet-ázsiai versenytársak (elsősorban Japán, Kína és Korea) esetében. A kulcsszereplők nemzeti, kétoldalú, többoldalú és globális fejlesztési és kooperációs programjai mindinkább két kulcskategória, a *cyber-infrastruktúra* (cyberinfrastructure) és a tudományművelés kiteljesített *digitális környezete* (cyberenvironment) köré rendeződnek.²

A cyberkörnyezet olyan integrált, *end-to-end*³ szoftverrendszer, amely ahhoz hasonlóan teszi *online* és azonnal elérhetővé a cyberinfrastruktúrát, ahogyan egy böngésző az internetet.⁴ A jelen és a jövő cyberkörnyezete a tudósoknak arra ad módot, hogy

- nagy volumenű és bonyolult projekteket és folyamatokat gondolozzanak, diszciplináris és földrajzi határok nélkül,
- eltérő természetű és óriási mennyiségű kísérleti, számítási és más adatforrásokat

² A *Magyar Tudomány* hasábjain 1999-ben, – az akkori helyzetnek megfelelően – Vajda Ferenc tekintette át tudomány és informatika kapcsolatát (Vajda, 1999). Jellemző, hogy technológiai szinten (a teljesítményadatok kivételével) ma is szinte minden aktuális, de a fogalomkészlet és a problémák szerkezete mostanra teljesen átalakult.

³ Teljeskörű végfelhasználói szolgáltatásokra épülő

mozgósítsanak komplex problémák megoldásához a helyi, intézményi és nemzeti erőforrások „összeépítése” révén (beleértve a kutatótársakat is),

- kiteljesítsék a kétirányú kapcsolatokat a már publikált és a még nyers, de már megszületett eredmények, az alap- és az alkalmazott kutatás, illetve a kutatás és az oktatás kulcsszereplői között.

A cyberkörnyezet egyaránt támogatja a tudománygyár működésének átfogó újratervezését (re-engineering) és a kutatási folyamatok jobb programozását – evvel a *tudomány új korszaka* születik meg (next generation science), amit bátran nevezhetünk *adat-intenzív tudománynak* (data intensive science) (Grey et al., 2005). Az adatok mennyiségének, illetve sokféleségének ugrásszerű növekedését eredményező változásokkal és a kutatóközösségek teljes behuzalozásával lerövidül a megszülető eredmény és az ahhoz való hozzáférés közti idő. A számítási kapacitás megsokszorozódása révén megválaszolhatóvá vált kérdések száma és fajtái robbanásszerűen megnövekednek, és a válaszidők hihetetlenül lerövidülnek. Az *e-business* és az *e-government* (a digitális térbe „átkúszó” üzlet s kormányzat világa után) beléptünk tehát az *e-science* korszakába – amit mi sem bizonyít jobban, mint a különböző kormányok által e-tudományra szánt források növekvő összege (az Egyesült Államok egymaga évi 400 millió dollárt, a britek öt év alatt 500 millió dollárt fordítanak erre a célra).

⁴ A meghatározás és a gondolatmenet egy különösképp érintett személytől, Jim Myerstől, az USA *National Center for Supercomputing Applications* (NCSA) *Cyberenvironments and Technologies Directorate* társ-igazgatójától származik (Myers, 2006). Az NSF „a lakossági áram- és vízszolgáltatás használatának könnyedségéhez” hasonlítja az informáciotechnológiai és tudásmenedzsment erőforrásokhoz való hozzáférést.

Az Egyesült Államok Nemzeti Tudományos Alapjának (*National Science Foundation – NSF*) *Cyberinfrastructure: Vision for 21st Century Discovery* című stratégiai anyaga 2006. január 20-án látott napvilágot,⁵ és a cyber-infrastruktúra kiteljesítése érdekében öt kiemelt területet állított a középpontba:

1. A nagyteljesítményű számítástudományt (high-performance computing)
2. Az adatok, az adatelemzés és a vizualizáció világát
3. Az adatok cyber-infrastruktúráját komplex, globális kontextusban
4. A kutatói közösségek együttműködését, a megfigyelőállomások felértékelődő helyét és az új virtuális szervezeteket
5. Az oktatási és munkaerő-vonatkozásokat (erre önálló terv született)

A tudományos kutatások hagyományos infrastruktúrájaként szolgáló *könyvtári rendszerben* a fentiekkel tökéletesen analóg változások mennek végbe. A második generációs digitális könyvtárakkal szemben támasztott követelmények a mennyiségi és szervezeti kihívásokra is reflektálnak. Heiko Schuldt (2005) áttekintésében ezek az alábbiak:

- Megfelelően kialakított szolgáltatásokkal minél nagyobb tömegű felhasználó számára tegye elérhetővé az igényelt adatokat és digitális könyvtári alkalmazásokat.
- Célozza meg az egy hálózati térbe kerülve adatokat és szolgáltatásokat nyújtók közti kölcsönös kapcsolatok kiépítését.
- Olyan számítási/feldolgozási és adattárolási kapacitást érjen el a grid-technológia segítségével, mint amelyek korábban

csak a nagy hálózati szolgáltatásokat jellemezték.

A folyamat jelentősége és a kihívás mélysége láthatóan már messze túlmutat az érintett tudományterületek és a számítástudomány képviselőin: a teljes tudományos közösség mellett a felső- és közoktatás szereplői, valamint a politikai és gazdasági elit számára is fontos, hogy sokoldalúan foglalkozzon az alapfogalmakkal, az elterjedt megoldásokkal, a cyber-infrastruktúra természetével és megoldó erejének „anatómiájával”, a változások dinamikájával, a fejlesztések tétjével, s az egész *átalakulás rendszereszméletű elemzésével*.

A továbbiakban elsőként arra keressük a választ (számos példa segítségével), hogy egészen pontosan mit és hogyan változtat meg a cyber-infrastruktúra *a tudományban mint tudástermelő és elosztó „megagépezetben”*.⁶ Az elemző áttekintés segítségével bemutatjuk *a változások rendszerességét*, s ebből kiindulva (e tanulmány folytatásában) megfogalmazzuk több radikális, új hipotézist, amelyek új oldalról mutatják be *a tudomány „cyberkorszakának” strukturális és intézményi korlátait*. Ezt követően pedig sokoldalúan érvelünk amellest, hogy az „információs társadalom” tudománya és tudományos közössége nem (elsősorban) az információtechnológia csodái, hanem *társadalmi innováció, egy „humán forradalom”* révén léphet fejlődése következő szakaszába. A cyber-infrastruktúra önmagában még nem eredményez paradigmaváltást, de előkészítője, megalapozója és nélkülözhetetlen háttérfeltétele annak.

⁵ Ennek előzménye az ún. *Atkins Report* volt, amely 2003 januárjában tette közzé szakértői ajánlását *Revolutionizing Science and Engineering Through Cyberinfrastructure* címmel. Lásd www.communitytechnology.org/nsf_ci_report/report.pdf

⁶ Lewis Mumford termékeny terminusa remekül használható – sokkal jobbnak tűnik, mint allegorikusan beszélni a nagybetűs *Tudomány*-ról.

Funkció és folyamat: mi történik valójában?

A tudományt átalakító digitális födrengés öt „szintéren”, öt „támadási ponton” zajlik, ezeket egyenként bemutatjuk, majd egy szintetikus ábrában, modellszerűen összesítjük a részfolyamatokat.

I.

Az új-mexikói Apache Point ad otthont a *Sloan Digital Sky Survey* (SDSS) 1,5 méteres teleszkópjának. A program kezdete, 2000 óta *másodpercenként 5 megabájt* adatot termel a 120 megapixel-es kamera. Amikor az ambiózus program véget ér (a tervek szerint 2008-ban), az égbolt egynegyedének háromdimenziós képe lesz előállítható, benne több mint egymillió galaxissal és kb. félmilliárd űrobjectummal.⁷ És ez csak a kisöccse lehet a működését 2010-ben, Chilében megkezdő *Large Synoptic Survey Telescope*-nak (LSST), amely a maga 8,4 méteres mérete mellett *minden éjszaka 30 terabájtnyi képi adatot* termel, távoli galaxisok milliárdjait téve láthatóvá és vizsgálhatóvá. 2012-ben fejeződik be a gigantikus rádióantenna-rendszer, a Hubble űrtávcsőnél tízszer élesebb *ALMA 66* telepítése (szintén Chilében, az Atacama sivatagban), amely *3 másodpercenként fog egy új galaxist* felfedezni.

Európai és amerikai rádiócsillagászok ugyanakkor már 2004-ben meggyőzően demonstrálták a Világegyetem kutatásának egy új módját – az Internet segítségével. Különböző kutatási hálózatok felhasználásával egy gigantikus méretű „virtuális teleszkópot”

hoztak létre, amely néhány éve még elképzelhetetlen, soha nem volt részletességű és valós idejű képalkotást tesz lehetővé.⁸ Brit, svéd, holland, lengyel és Puerto Ricó-i rádióteleszkópok húsz órán vették célba a por- és gázfelhőbe burkolt, de erős rádióhullámokat kibocsátó, tőlünk több mint 15 ezer fényévnire, az Aquila (Sas) csillagképben keringő IRC+10420 kódjelű csillagot. A várhatóan hamarosan szupernovává váló égitestről a nemzeti kutatási hálózatokat felhasználva zúdultak az adatok a hollandiai feldolgozóközpontba, ahol a *9 terabájt* adattömeget a „Korrelátornak” becézett speciális szuperszámítógép dolgozta fel. A „minél nagyobb a teleszkóp, annál jobb a felbontás” elve a virtuális internet-teleszkóppal fényesen bebizonyosodott: már az első kísérletnél is ötször jobb képet lehetett produkálni, mint amit a Hubble űrteleszkóp nyújt. A távlatok felmérhetetlenek: nemcsak az immár *e-VLBI* néven számon tartott technológia finomodása, hanem az adatátviteli sebesség további növelése is nagyságrendi ugrásokat eredményezhet.⁹

Petabájtos nagyságrendben fogja előállítani és hozzáférhetővé tenni az új adatokat a genfi CERN-ben épülő *Large Hadron Collider* részecskegyorsító is, és már folyamatosan ontják a feldolgozóközpontokba a jeleket az ún. *szenzor- vagy érzékelőhálózatok* (sensory networks), amelyek radikális árcsökkenésük után szintén petabájt nagyságrendig jutnak majd. De már most is létező projektek tucatjai révén kapunk adatokat a nanoszinttől a kontinensnyi tartományokig nagy méretek-

⁷ 2004 vége óta az egyre gyarapodó adatbázis magyar nyelven is hozzáférhető.

⁸ Korábban a VLBI (Very Long Baseline Interferometry) technológiát alkalmazó teleszkópok adatait szalagon küldték feldolgozóközpontokba elemzésre, emiatt sokszor heteket vagy hónapokat kellett várni egy-egy

megfigyelés eredményére. Csakhogy a szélessávú adatátvitel forradalma új fejezetet nyitott a rádiócsillagászat történetében, amit a nagyszabású kísérlet meggyőzően igazolt.

⁹ http://www.naic.edu/~astro/aovlbi/press_release/eVLBI_AR.htm

ben, folyamatosan és a történésekkel egyidejűleg a világról az óceánokban, gleccserekben, patakparti talajban, a légkörben, mozgó járműveken, az emberi testben telepített szenzorhálózatoktól. A kutatók folyamatosan új kérdéseket tehetnek fel, vagy új feltevéseket ellenőrizhetnek az egymással összhangban dolgozó, adatokat megosztó hálózatokon. Ez Gaetano Borriello szerint pontosan olyan hatású, mint a kísérleti természettudományok megjelenése volt az újkorban, a létrejövő új „minőséget” Vernor Vinge bátran el is nevezi *digitális Gaiának*, az „input adatok globális ökoszisztémájának”. (A magunk részéről talá-
lőbb metaforának érezzük a korábban Bill Gates ajánlott digitális bőrt – *digital skin*). Jellemző, hogy a korábban kizárólag adatfeldolgozásra használt számítógépek világába is „betör” az érzékelők forradalma, egyre több számítógép fog szenzorként is üzemelni egyéb feladatainak ellátása mellett.

Az ezt bizonyító egyik legizgalmasabb és leghasznosabb fejlesztés a merevlemez olvasófejének mozgásait követő hibajavítási és biztonsági alkalmazást használja ki. A *Tsunami Harddisk Detector*¹⁰ nagyobb rázkódások, így szeizmikus mozgások esetén képes az egyszerre több számítógép winchestere által mért mozgásokat összesíteni. A felhasználóknak csupán egy apró programot kell Internetre kapcsolt gépükre telepíteni. Az atomórákkal szinkronizált alkalmazás pontosan jelzi a rezgések terjedését, így lehetőséget kínál egy esetleges földrengés epicentrumának meghatározására. Az egyszerű algoritmus megvizs-

gálja, hogy a rengés a tenger alatti földlemez mozgásából ered-e, amennyiben igen, az adott területen automatikusan riaszt, figyelmeztetve a lehetséges szökőárra.

Csak emlékeztetőül: a feldolgozórendszerek jelenleg a terabájt szintű adatkezeléssel is küszködnek, miközben jól láthatóan itt van már a petabájt nagyságrend. Az input adat-tömeg „nyomása” alatt roskadoznak a feldolgozórendszerek – nem véletlen, hogy a tudomány cyber-infastruktúrájának második „epicentruma” a *számításteljesítmény növelése*.

2.

Az újonnan képződő adatmennyiség tárolásával és átvitelével kapcsolatban számottevő, a felhasználás folyamatosságát akadályozó kapacitorikus korlát a rendszerbe állított adathordozók fizikai „szaporíthatósága” és a szélessáv detonációja miatt ebben a pillanatban nem mutatkozik. Annál inkább előtérben áll a számítógépes adatfeldolgozás műveleti sebességének (computation performance) a megsokszorozása, amelynek egy hagyományos és egy erősen új módja az elmúlt néhány évben párhuzamosan indult hatalmas fejlődésnek: a gigaflop-tartományból indulva megérkezett a teraflop-szintre, hogy szinte azonnal a petaflop-skálát kezdje ostromolni.¹¹

Az egy géppel elérhető számításteljesítmény növelésének mintegy harminc éve a *szuperszámítógép-fejlesztés* az elsődleges terepe, a jelenlegi architektúra határainak folyamatos tágításával.¹² A hetvenes évek közepén a

¹⁰ <http://www.tsunami-alarm-system.com/>

¹¹ A Flop (Floating Point Operation, lebegőpontos művelet) a számítógépek teljesítményének szokványos mérőszáma. A petaflop (SI-szabvány szerint: petaFLOP) teljesítményű számítógép 1 millió milliárd, vagyis 10¹⁵ lebegőpontos műveletet végez másodpercenként.

¹² A alternatív architektúrák (bioszámítógépek, kvantumszámítógépek), amelyek a jelenlegi működési elvektől gyökeresen eltérő módon és annak fizikai határait átlépve kívánják és ígérik az új nagyságrendek elérését, még csak kísérleti fázisban vannak. Az ATI nemrég bejelentett *stream computing* technológiája

Cray erre „szakosodott” első gépei 1-200 megaflop körüli teljesítményt tudtak produkálni, és már akkor is légkörkutatókat, majd különböző szimulációkra használták őket. A „repülőrajt” után a fejlesztés dinamikája (és a szuperszámítógépes üzlet) megtorpanni látszott, hogy aztán az ezredforduló után soha nem látott tempóban induljon meg a verseny az egyre gyorsabb masinák egyre nagyobb számú gyártója között.

Sokáig az NEC által Yokohamában, 2002-ben, klimatológiai megfigyelésekre üzembe állított japán *Earth Simulator* elnevezésű gép vezette a listát a maga közel 36 teraflopos sebességével, de mivel az Egyesült Államokban kisebb „szputnyiksokkot” okozott a japán térnyerés, 2004-ben a NASA *Columbia*-ja (az SGI gyártásában), majd 2005-től az USA Lawrence Livermore Nemzeti Laboratóriumába telepített IBM *Blue Gene* szériája futott fel a dobogó tetejére. A jelenlegi listavezető, a széria „L” jelű darabja, elsőként és ezidáig egyedülként átlépve a százás küszöböt, már 280 teraflopnál tart. A következő generációt most telepítik ugyanott, ebben 130 000 processzor dolgozik majd egy időben, sebessége pedig el fogja érni a 360 teraflopot.

Az IBM-nek egy másik fejlesztése is közel jár a befejezéshez. A *Roadrunner* az amerikai energiaügyi minisztérium Los Alamos-i Nemzeti Laboratóriumának új superkomputere lesz, és teljesítménye a tervek szerint eléri az 1 petaflopot. A Cray, amely egy géppel „visszakúszott” a térképre, 2007-re ígéri 1 petaflopos gépe, az *Oak Ridge* üzembe helyezését.

azonban izgalmas valóság: a korábban pixelszámításokra kifejlesztett grafikus processzorok alkalmazásának kiterjesztésével újszerű módon növelhető meg a számításteljesítmény, azonnal komoly ígéreteket jelentve az élettudományok (különösen a járványmodellelés) és a klimatológia számára.

Eközben a japán RIKEN kutatólaboratóriuma már el is készítette, és üzembe is helyezte a maga 1 petaflopot tudó gépét, a *MDGRAPE-3*-at, amellyel molekuladinamikai szimulációkat, proteinszerkezeti előrejelzéseket végeznek (a listát csak azért nem ez a szuperszámítógép vezeti, mert nem „általános célú”, hanem kifejezetten erre a feladatra fejlesztették). A talaj „forró”: az eddigi versenyzőkhöz csatlakozó franciák és németek után immár Kína is tervbe vette, hogy 2010-ig a világ legerősebb szuperszámítógépe az övé lesz. S miközben az *Earth Simulator* 2006 júliusában még mindig tizedik tudott maradni a számítógép-teljesítmények Top 500-as ranglistáján a maga 36 teraflopjával, aközben Japán még egy nullát kíván a jelenlegi számok mögé írni: bejelentette a 10 petaflopos gép terveit.

A másik fejlesztési irány, a *kisebbségi teljesítményű számítógépek hálózati összekapcsolása* egyelőre még (csak) a *teraflopos* nagyságrendnél tart. A számítástechnikában már korábban is létező elosztott rendszerek (distributed computing) világának új fejleményeként először (a kilencvenes évek közepétől) önkéntes alapon, a kihasználatlan gépidő átengedésével jöttek létre óriási kapacitások – a Földön kívüli életformák rádiójelei után kutató pionír *SETI@home* mára elérte a négymillió gépet és a 100 teraflopos teljesítményt. A korábbi *genome@home*-t magába olvasztó, proteinstuktúrákat szimuláló *Folding@home* pedig kevesebb géppel is már 150 teraflop körül jár. Az első eredmények láttán három irányban fejlődnek tovább a feldolgozóhálózatok.

(1) Az önkéntes felajánlásokra épülő programok mindinkább összehangoltan, közvetítetten és számos rész megoldással támogatottan futnak tovább; (2) Miután a „hivata-

los” tudománypolitika is felfedezte magának az összekapcsolásban rejlő lehetőségeket, elindultak a szervezett, nagy intézményi kapacitásokat összeépítő tudományos célú hálózatok, az ún. *grid*ek (GRID-ek); (3) Állandó a törekvés a grid-architektúra informatikai innovációjára, új típusú *grid*ek építésére.

(1) A Berkeley Egyetem nyílt forrású szoftverplatformra épülő BOINC programja (Berkeley Open Infrastructure for Network Computing) tucatnyi, külön-külön is jól ismert tudományos projekt számára „közvetíti” ki a gépidő átengedését (a legtöbb jórészt környezeti modellezéssel, klímaváltozással foglalkozik, de megtaláljuk benne a mi SZTAKI-nk *Desktop Gridjét* is). A SETI@home mellett ebbe a körbe tartozik például a bolygó klímájának előrejelzésén dolgozó *Climateprediction.net*, illetve a pulzárókból kiinduló gravitációs jeleket kereső *Einstein@home*. A felhasználó a számítógépe paragon heverő képességeivel keresheti betegségek nyitját (*Rosetta@home*) és különféle fehérjék várható tulajdonságait (*Predictor@home*).¹² A BOINC lehetővé teszi, hogy ugyanaz a szoftver a számítógépet többféle probléma megfejtésével is elfoglalja. A megosztott műveletekre felajánlott számítógépek erejét így több projekt párhuzamosan használhatja. Természetesen a BOINC-től függetlenül továbbra is jönnek létre önállóan szerveződő hálózatok: eddig 170 ezer számítógéppel kapcsolódtak be például önkéntesek a világ különböző tájairól, hogy az óriási számítógépes kapacitást kihasználva a *FightAIDS@*

¹² Az ismertebb hálózatok közül még néhány: *LHC@home*, *BBC Climate Change Experiment*, *Quantum Monte Carlo at Home*, *World Community Grid*, *Seasonal Attribution Project*, *SIMAP*. További részletekre lásd <http://boinc.berkeley.edu/>

Home programban részt vállalva AIDS-terápiákat keressenek

(2) Az „intézményi” *grid*ek „első generációját” (Grid 1.0, computational grids) a nagy volumenű szimuláció jellemezte, „klasszikus” számítási erőforrás allokációval. Az újabb generációs *grid*eknél (Grid 2.0, data grids) az adatáramlás iránya megfordul, a cél immár a nagy volumenű együttműködés¹⁴ támogatása, ahol a megtermelődés helyéről nagy kapacitású feldolgozóközpontokba jutnak el az adatok.¹⁵

Az Egyesült Államok a hálózati erőforrások és az elosztott rendszerek új generációjától az innovációt és a gazdasági növekedést egyaránt stimuláló infrastruktúra kialakítását reméli, és erőfeszítéseit a GENI Initiative¹⁶ (*Global Environment for Networking Innovations*) köré csoportosítja, ami a gridvilág minden biztonsági, platform-, menedzsment-, felhasználhatósági, illetve szolgáltatási/alkalmazási kérdésében előre kíván lépni.

Az amerikai *Open Science Grid* (OSG) és az európai *Enabling Grids for E-Science* (EGEE) összekapcsolásával *minden idők legnagyobb tudományos megagépezetének* (*Worldwide LHC Computing Grid* –

¹⁴ <http://www.tryscience.com>

¹⁵ A szakirodalomban nem harmadik generációként, hanem a *grid*ek harmadik típusaként tartják számon az ún. *equipment grid*eket, mint amilyenek például az összekapcsolt teleszkópok vagy akár az előző részben tárgyalt intelligensebb szenzorhálózatok is. A gridirodalom elképesztően részletes és adatgazdag összefoglalását lásd a Wikipédián, a <http://www.gridcomputing.com/> oldalon, a friss fejleményeket a <http://www.gridtoday.com/> oldalon lehet megtalálni.

¹⁶ A programot a National Science Foundation informatikai igazgatósága (*Directorate for Computer and Information Science and Engineering* – CISE) indította. <http://www.nsf.gov/cise/cns/geni/>

WLCG) sikeres próbaüzeméről értesülhetünk 2006 elején. A kísérletben bebizonyosodott, hogy amikor a Genfben épülő óriás részecskeütköztető 2007-ben elkezdi dolgozni, a mintegy két tucat partnerintézmény képes lesz az innen érkező terabájtnyi adat fogadására és tárolására, illetve elérhetővé tételére.

- (3) Alacsony költségű alternatív gridtechnológia a Dr. Rajkumar Buyya által a Melbourne Universityn fejlesztett *Alchemi*,¹⁷ amely a Windowsszal futó gépeket szervezi virtuális szuperszámítógéppé.

Egyre több kutatói hálózat tervezésébe épül már eleve bele a cél és a lehetőség, hogy egyidejűleg több gridet is ki tudjon majd szolgálni – illetve különböző hálózati erőforrásokat igénybe véve párhuzamosan lehessen end-to-end szolgáltatásokat futtatni. A kulcsszavakkal – *Articulated Private Networks* (APNs), *Concurrent Architectures are Better than One* (Cabo) – célszerű mihamarabb megismerkedni, a kanadai *CA*net 4* vagy az amerikai *National Lambda Rail* már ezek egyfajta „prototípusaként” üzemel.

3.

Mielőtt az elemi és az (elő)feldolgozott adattömeg elérné felhasználása legfontosabb célállomását, a kutató agyát, működésbe lépnek a kutatói tevékenység támogatásának háttérrendszerei is. Mint egy tisztító zuhanyosoron, úgy mennek át a „nyers” adatok különböző algoritmusok, átalakítások, rendezések és érték-hozzáadott műveletek láncolatán, ahol valamennyi fázis egyetlen célt szolgál: *az értelmezést végző elme számára tenni áttekinthetőbbé, rendezettebbé, kezelhetőbbé az input információk hatalmassá növekedett áramát.*

¹⁷ <http://www.alchemi.net/>

Abból a hátrányból, hogy a terabájtnyi adattömegek már nem futtathatóak „privát” gépeken, és a „birtoklás” helyett a hozzáférhetőség válik kulcskérdéssé, a tudományos közösség úgy kovácsolt előnyt, hogy új típusú, a tárolást és az elérést biztosító szakosított célintézményeket, ún. *science centereket*, tudományos központokat hozott létre.

Ezek talán legismertebb képviselője a Sanger Intézet genetikai adatbázisa, a *Trace Archive*, amely a világ tudományos közössége által felderített összes génszekvenciát tartalmazza: 2006 végén körülbelül 40 terabájtnyi adatot. Mérete tízhavonta megduplázódik. Az adatbázis 2006 elején egymilliárd rekordot számlált (egy rekord átlagosan 864 bázis sorrendjét tartalmazza).¹⁸ A tudományos központok nagy előnye, hogy „specializálódhatnak” egyes adattípusokra és a kollégák kiszolgálására. A tárolt adattömeg mellett alkalmazásokat futtatnak és támogatnak, és felkészült személyzettel tartják karban a rendszereiket. Elemi érdekük az innováció. *Storage Resource Broker* (SRB) néven például párhuzamosan fejlesztett sikeres, gridekre optimalizált adatkezelő alkalmazást egy magáncég (a Nirvana) és a San Diegó-i szuperszámítógép-központ. Utóbbi – közel félszáz együttműködő partnert kiszolgálva – élenjár a metaadatok használatát megkönnyítő megoldások fejlesztésében is (*MetaData Catalog* – MCAI).

Nem véletlen, hogy a cyber-infrastruktúrára elkezdte saját céljaira alkalmazni az üzlet

¹⁸ 800 milliárd bázis a szuper-génadatbázisban <http://index.hu/tech/tudomany/trco60118/> További ismertebb tudományos központok: a *Fermilab* kezeli a már bemutatott virtuális teleszkóp (az SDSS) adatait. A B-mezonok millióinak tanulmányozását támogatja a BaBar projekt (*Stanford Linear Accelerator Center* – SLAC). A *Biomedical Informatics Research Network* (BIRN) adatainak a 6 petabájtnyi archív adatot tároló *San Diego Supercomputer Center* (SDSC) ad otthont.

világára kifejlesztett, tranzakció-orientált munkafolyamat-vezérlő szoftvereket (workflow tools) is. A hosszú adatfolyamok és összetett műveletegyüttesek vezérlését, áttekintését megkönnyítő, egyedi és vagy standard felhasználói felületek (interfészek) komponálását lehetővé tévő eszközöket egyszer csak birtokba vette a bioinformatikusok közössége, de rohamos terjedésnek indult a csillagászok és az ökológusok között is (legismertebb példái a kanadai *BioMoby* és *Taverna*, valamint a *Triana* és a *Keppler*¹⁹).

A korszerű szoftveres megoldások és munkafelületek azonban csak kényelmesebbé teszik a kutatók életét, de önmagukban még nem szüntetik meg az adatok „előfeldolgozásával” kapcsolatos szűk keresztmetszeteket. Az adatállomány elemzéséhez még mélyebbre, *algoritmus szintig* kell „lenyúlni”, hogy a megnövekedett mennyiség kihívását az *analitikus eszközök* követni tudják (Gray et al., 2005). Az új nagyságrendek pokoli nyomásvizonyai között hasonlóképpen alakul át az adatok megjelenítésének, *vizualizációjának* világa: fejlesztés alatt állnak azok az új *adatprezentációs architektúrák* (Agutter – Bermudez, 2005), amelyek már kifejezetten az új méretek „leképezésére” hivatottak.

A cyber-infrastruktúra változásainak ezen a szintjén az eszközöknek egy olyan, *az elsődleges adatok előfeldolgozására „szakosított” családja formálódik ki, amely a digitális platformon találkozik és olvad össze a tudományos munka korábbi, a másodlagos adatokhoz való hozzáférésre kifejlesztett óriásrendszereivel* (a szakirodalmi, bibliográfiai adatokkal és magukkal a különböző formában közzétett tudományos tartalmakkal). Amikor Ian Foster bátran bevezeti és népszerűsíti a *szolgáltatásorientált*

¹⁹ <http://gchelpdesk.ualberta.ca/services/services.php>
és <http://kepler-project.org>

tudomány (service-oriented science) kategóriáját (Foster, 2005), nem pontosan erre a „fúzióra” utal, de hogy találébb kifejezést nehéz volna keríteni rá, az majdnem biztos.

4•

Az előző három szinten létrejött hálózati, átviteli, tárolási és műveleti kapacitások technológiai oldalról szinte mindent előkészítenek ahhoz, hogy létrejöjjön az adatok hatékony találkozási az azokat értelmező és összefüggésbe helyező kutatói közösségekkel. *A technológiai kihívás tehát még tovább tolódik a szakma-szintű interakció támogatása felé.* A megnövelt méretű adattömeg magas értékhozzáadású feldolgozása a kooperáció újszerű formáit igényli, egyidejűleg minél több szakértő összehangolt munkáját, amellyel elkerülhetőek a felesleges párhuzamosságok, és biztosítható a szükséges „humán infrastruktúra”, a megfelelő minőségben, de még inkább mennyiségben és diszciplináris sokszínűségben. *A korszerű hálózati technológia soha nem látott méretű és rugalmasságú tudásközösségekbe kovácsolja az elsődleges adatokra érzékeny kutatókat, akiket saját „tárgyuk” természete húz a globális együttműködés felé.* Elsősorban a genetika és az asztronómia az a két tudományterület, amelyben elkerülhetetlennek látszik a tudósok részleges „feloldódása” világméretű kutatóközösségekben, amelyeket a fundamentális és megkerülhetetlen adatbázisok közös gyarapítása és felhasználása abroncsol össze.

A nyomás egyszerre több irányból érkezik. Az egymással versengő nemzeti, illetve szupranacionális tudománypolitikák, az üzlet és maga a tudomány is ugyanabban az irányban keresi a megoldást.

Az Európai Bizottság 2006. április 28-án jelentette be egy nagysebességű hálózati infrastruktúra terveinek jóváhagyását, melyhez

45 ezer európai és kínai kutató, illetve diák férhet majd hozzá. Az ORIENT (*Oriental Research Infrastructure to European Networks*) névre keresztelt hálózat több mint kétszáz kínai egyetemet és kutatóintézetet fog összekapcsolni Európával, 2,5 Gbps adatátviteli sebességgel, többek közt a radioasztronómia, a fenntartható fejlődés, a meteorológia és a gridszámítás területein segíti majd a már meglévő²⁰ és a majdani európai-kínai tudományos együttműködések. Az infrastruktúra-építés ebben az esetben elsősorban a kutatók összekapcsolását szolgálja, jól érzékelhető geostratégiai céllal.

Az *Innocentive* kezdeményezés²¹ – ami nem más, mint egy ügyesen megkonstruált weboldal – világszínvonalú kutatók (problem solverek) és fejlesztési kérdéseiket tudományra orientáltan megoldani kívánó vállalatok (seekerek) együttműködését szervezi. 2006 végén több mint 175 országból több mint 100 ezer tudós, illetve tudományos szervezet állt már csatasorban, elsősorban komplex kihívásokra adott innovatív megoldásokkal, főleg gyógyszerkutatói, biotechnológiai, kémiai, élelmiszeripari és műanyagipari (nemritkán a 100 ezer dolláros összehatárt is meghaladó) projektekben.

Egy másik weboldal „keretfelületet” épített bármilyen, esetlegesen felmerülő tudományos együttműködési szándékhoz. A (most már tizenöt nyelven, köztük magyarul is elérhető) www.academici.com lehetővé teszi, hogy a világ bármely pontján lévő tudósok és kutatók határok és kötöttségek nélkül bármikor megoszthassák egymással a tapasztalataikat, vagy megvitathassanak bármilyen kérdést, javaslatot, előterjesztést, problémát. Az Academicin az érdeklődők többfajta szempont, például kutatási terület vagy tudományos érdeklődés alapján is kereshetnek, és így akár percek alatt felvehetik a kapcsolatot egymással a világ legtávolabbi helyein lévő oktatási vagy tudományos intézmények.

Az Európai Unió DILIGENT projektje²² „virtuális tudományos közösségek” számára fejleszt biztonságos, koordinált, dinamikus és költséghatékony kísérleti felületeket a tudás megosztására és az együttműködésre, a Grid és a DL (digital library) technológiák kombinálásával. Ez a kombináció a jövő e-tudományos infrastruktúrájának kulcsa, számos további kutatási, sőt ipari alkalmazási lehetőséggel. Két valós idejű alkalmazási területen kísérleteznek: egy környezetvédelmi és egy kulturális örökségvédelmi témában.

Érzékelhető, hogy a tudományművelés hagyományos erőforráskorlátai mellé (intézményi kapacitások, pénzhiány) felzárkózott a „humán infrastruktúra” mint szűk keresztmetszet: a virtuális kollaborációs felületek és szolgáltatások világa annyiban segít, hogy a kreativitást is szolgáló kiterjesztett kommunikáció és a hatékonyságot fokozó koordinációs mechanizmusok révén racionalizálni és orientálni képes a feldolgozó munkába bevonható kutatói állomány teljesítményét és figyelmét: jobb gazdálkodást tesz lehetővé a már meglévő alkotó erővel; a tudományos megismerést mozgásban tartó agymunkával.

Érzékelhető, hogy a tudományművelés hagyományos erőforráskorlátai mellé (intézményi kapacitások, pénzhiány) felzárkózott a „humán infrastruktúra” mint szűk keresztmetszet: a virtuális kollaborációs felületek és szolgáltatások világa annyiban segít, hogy a kreativitást is szolgáló kiterjesztett kommunikáció és a hatékonyságot fokozó koordinációs mechanizmusok révén racionalizálni és orientálni képes a feldolgozó munkába bevonható kutatói állomány teljesítményét és figyelmét: jobb gazdálkodást tesz lehetővé a már meglévő alkotó erővel; a tudományos megismerést mozgásban tartó agymunkával.

5.

És még mindig nincs vége a beavatkozási lehetőségeknek. Szinte minden kutató megélt tapasztalata a felismerés, hogy a különösen

²⁰ Mint például az Európa tudományos gridszámítási kapacitását Kína irányába meghosszabbító EUChinaGrid, vagy az európai és kínai rádióteleszkópokat összekötő EXPReS projekt.

²¹ <http://www.innocentive.com/about/seeker.html>

²² <http://www.diligentproject.org/>

értékes erőforrássá lett emberi életidő rosszul hasznosul a tudományban, és a rutinmunka kényszerű végzése a kreatív tevékenység elől rabolja a figyelmet és az energiát. Az igényelt „humán erőforrások” biztosításának jól bevált módja a szellemi műveletek további automatizálása – másképp: *az algoritmizálható agymunka*²³ *gépesítésének újabb hulláma*. Az így felszabaduló idő hatékonyan forgatható vizsgáló és problémamegoldó folyamatokba.

Az agymunka automatizációjának első, forradalmi fejezetét a számítógépek története írta, itt már nem maradt tér további beavatkozásra. Ismeretelméleti értelemben a szenzorok is részben az agymunka gépesítésének tekinthetőek, hiszen *az in situ* érzékelés-teljesítményt váltják ki, és sokszorozzák meg. Az információtechnológia harmadik rendszerszintjén a szövegalapú adatok tárolása, keresése, indexelése stb. is már régóta sikeres automatizálási célpont, de az online világ a

²³ Az „agymunka” itt nem metaforaként, hanem terminusként szerepel. A *brain work* az „új gazdasággal, Internet-gazdasággal” foglalkozó közgazdasági irodalom kedves kifejezése (Dent, 1998; Harney, 2005), noha eredete a 19. század végére megy vissza. Minden bizonnyal Alfred Marshall, a nagyhatású *Principles of Economics* (1890) szerzője (Marshall, 1890) használta egy korai esszéjében először (a *manual work* párjaként), de igazán népszerűvé az anarchizmus nagyhatású teoretikusa, Pjotr Alekszejevics (Peter) Kropotkin tette egy tanulmányával (Kropotkin, 1890), majd számos további kiadást megért könyvével (Kropotkin, 1899). Konferenciák címadó kulcsszava (az amerikai nyelvészek 2000-ben Washingtonban rendeztek tanácskozást az „agymunkáról”), matematikai közgazdászok termelékenységi és értékszámításainak célpontja. Megszületett már az e-agymunka (e-brain work) kifejezés is, elsősorban online tanulási és kutatómódszertani kontextusban. (<http://www.edscuola.it/archivio/lre/alis/metod.htm>). Stephen M. Dent (2000) pedig az egész információs forradalmat az agymunka (pontosabban a rutinműveletekre szakosodott *bal agyfélteke* munkájának [left-brain work] automatizálására vezeti vissza.

kulcsszavas kereséssel, a professzionális bön-gészőkkel ehhez még sokat hoztatt az elmúlt években. A szolgáltatások szintjén azonban még nyílik tér a beavatkozásra: a *robotkönyvtárosok*, például, arra szabadítanak fel időt, hogy a könyvtárosok magasabb értékhozzáadású tevékenységgel segíthessék a tudástermelésben érintett kollégáikat.

A MTI 2004. február 17-én jelentette, hogy napi 24 órás szolgálatot teljesítő robotkönyvtárost állítanak szolgálatba az USA Indiana államában lévő Valparaiso egyetemének könyvtárában. A megadott könyveket az okos masina megkeresi a polcokon (a könyvtárnak jelenleg 450 ezer könyve van), és átadja az emberi személyzetnek. Alig másfél hónap múlva online források számoltak be arról, hogy Spanyolországban is hasonló szerkezet készült. A castellóni I. Jakab Egyetem (Univer-sitat Jaume I.) robotikacsoportja által tervezett gépkönyvtáros infravörös sugarak segítségével tájékozódik, azonosítja az olvasók által kért könyv címét a kötet borítóján, sőt a gerincén is, mi több, meg tudja különböztetni az eltérő grafikájú címeteket is. Miután megtalálta a kért könyvet a polcon, leemeli, majd átnyújtja az igénylőnek. Eközben kikerüli az akadályokat, s nem ütközik össze a figyelmetlen könyvtárlátogatókkal sem.

De a robotkönyvtáros-szenzációk eltörpültek egy még korábbi bejelentés mögött: az agymunka algoritmizálása 2004 óta immár a negyedik rendszerszinten, a feldolgozó munkában is valósággá vált.²⁴

²⁴ A *Nature* 2004. január 15-i számában megjelent, a kutatók teljesítményével vetekedő robottudósról szóló beszámolót a *New Scientist* ismertette, majd nyomában szinte minden jelentős hírportál összefoglalót közölt róla. A leíráshoz az sg.hu-n *Hírszerkesztő* aláírással közölt szövegből és az Index.hu-n 2004. január 20., kedd 12:00 órakor megjelent terjedelmesebb cikkéből (Babucsik Péter: Robottudós a laboratórium-

A Walesi Egyetem, a Manchesteri Egyetem, az aberdeeni Robert Gordon Egyetem, valamint a londoni Imperial College számítástudományi szakembereiből álló brit kutatócsoport *robottudósnak* „elkeresztelt” rendszere egyes *genomikai laboratóriumi rutinvizsgálatokat*²⁵ a valódi tudósokéval vetekedő eredményességgel képes elvégezni. Segítséget nyújt a genetikai kísérletek lefolytatásakor, ám ezt követően önállóan elemzi és értékeli az eredményeket, sőt ha kell, új hipotéziseket is felállít, és hozzájárul a legmegfelelőbb kísérleti módszer kiválasztásához is.

A kutatócsoport ráadásul az adott kísérleteket szintén lefolytató végzős egyetemi hallgató teljesítményével is összevetette az új „kolléga” eredményeit, s meg kellett állapítaniuk, hogy a kapott eredmények az automatizált rendszer esetében ugyanolyan pontosak voltak. A 98 %-os megbízhatósággal teljesítő robot ugyanakkor kevesebb kísérlet lebonyolításával jutott ugyanarra az eredményre, mivel hipotézisgeneráló szoftvere előbb találta meg a megoldást.

A „robottudós” megkonstruálása Stephen Muggleton professzor számára nemcsak arról szólt, hogy a fáradságos aprómunkát „kiváltásák”. Demonstrálni kívánták, hogy a rutinjellegű laboratóriumi vizsgálatok automatizálásával a tudósok rengeteg teher alól szabadulhatnak fel, s *tudásukat és figyelmüket kreatívabb, magasabb szintű tudományos feladatok megoldására fordíthatják*.²⁶

ban) használtam fel szövegeket, néhol változatlan formában. Lásd <http://index.hu/tech/tudomany/labor040120/?print>.

²⁵ A laboratóriumi tesztek során az előre beprogramozott, szükséges genetikai és biokémiai háttérinformációk ismeretében sikerrel állapította meg a vizsgált élesztőgombában (*Saccharomyces cerevisiae*) található bizonyos gének funkcióját.

A cybertudomány taxonómiája és érvényességi köre: az e-science diadala és korlátai

A tudomány aktuális mozgásirányait áttekintve Paul A. David és Michael Spence az adott résztvevőkészségek elsődleges célja alapján osztályozták a fejlesztéseket. Ennek alapján különítették el négy kategóriát, amelyek tökéletesen egybevágnak az általunk bemutatott rendszerszintekkel (David – Spence, 2003).

- Adatközpontú (Data centric)
- Számításközpontú (Computation centric)
- Interakció-központú (Interaction-centric²⁷)
- Közösségközpontú (Community-centric)

Társítsuk e négy szinthez azokat a műveleteket, amelyek elvégzése érdekében a kapacitások csatornákba állnak, és rendeljük melléjük azokat a mennyiségeket, amelyek alapján a „kimeneti” teljesítmény mérhető és összehasonlítható – és megkapjuk a cybertudomány „rétegmódeljét”. (I. táblázat)

Mint láttuk, a technológia könnyűszerrel átlépi a kategóriahatárokat (a gridek például gyakorlatilag mind a négy műveleti szintet kiszolgálják már), és erős „csomósodások” figyelhetőek meg az egyes szintek között is. A 2. és a 3. szintet a tudományos központok (science centerek) közelítik egymáshoz, és határozott jelei vannak annak, hogy ez az integrációs mozgás nem áll meg az 1., illetve a 4. szint határainál. A 3. és a 4. szintek is olyan erősen „vonzák egymást”, hogy közös minőségükként Paul A. David a „kollaboratív e-tudomány” kifejezést (collaborative e-science) javasolja (David, 2004).²⁸ Az 1. és a

²⁶ A gép – fogalmaz egy interjúban Muggleton (Law, 2004) – soha nem helyettesíteni fogja az emberi kutatókat, hanem lehetővé teszi számukra, hogy gyorsabban és kevesebb kényzsertevékenységgel érjék el eredményeiket.

²⁷ Az „interakció” Davidék szóhasználatában az „alkalmazás, irányítás, vizualizáció” műveleteit fedti le.

Szintek és javasolt elnevezések	A támogatandó művelet	Központi mozzanat	Egyenértékes
1. szint Adat-input	gyűjtés (collection) létrehozás (production)	adat	mennyiség (elemi e., byte)
2. szint Számításteljesítmény	feldolgozás (processing)	számítás	feldolgozási idő
3. szint Előfeldolgozás	rendezés (ordering, integration) elemzés (analysis)	értékhozzáadás interakció	felhasználhatóság
4. szint Kreatív agymunka	átalakítás, elméleti innováció (transformation)	közösség	újdonosságérték

1. táblázat

2. szint a kihívás analóg természete és számos közös hardver- és szoftvereleme miatt „csúszik össze” a *petaszintű tudományos adatmenedzsment* (peta-scale scientific data management) gyűjtőkategóriába. A távolságok valóban rövidülnek: ma már semmi meglepő nincs

²⁸ Az együttműködés (kollaboráció – *collaboration*) és a laboratórium (laboratory) szavak összekapcsolásából képzett korábbi terminus, a *kollaboratórium* (col-laboratory) nem tudott meggyökeresedni (Vajda, 1999).

abban, ha egy tudós a 4. szinten azonnali igényt fogalmaz meg egy 1. szintű szenzor adatra, amit az összekapcsolt rendszerek a lehető legrövidebb időn belül megpróbálnak biztosítani. Nem véletlen tehát, hogy mind többen vetik fel a *cybertudomány következő fejlődési szakaszaként az egyetlen óriásrendszerben való összeolvadás* hipotézisét.

A svájci anyagtudományi és technológiai kutatóintézet, az *Empa* Lorenz Hilty által

Szintek	„Expanziós nyomás” középről	Létező integrált minőségek	Teljes fúzió
1. szint Adat-input		Petaszintű tudományos adatmenedzsment	VARRATMENTES TUDOMÁNYOS SZUPERHÁLÓZAT
2. szint Számításteljesítmény	↑ Tudományos központok		
3. szint Előfeldolgozás	↓		
4. szint Kreatív agymunka		Kollaboratív e-tudomány	

2. táblázat

vezetett kutatócsoportjának előrejelzése szerint tíz éven belül körülbelül egymilliárd ember mintegy egybillió elektronikusan összekapcsolt „tárgy” segítségével kapcsolódik össze egy soha nem volt méretű „mega-gépezetbe”. Sir Tony Hoare és kutatótársai is abból az alapfelvetésből indulnak ki, hogy a világ idővel egyetlen végső globális számítógépes hálózatba forr össze, s időben el kell kezdeni az ennek megépítéséhez szükséges mérnöki-tervezői munkát.

Út a cyber-infrastruktúra szintjeinek teljes összeolvadásához

Ez a vízió olyan erős, hogy sok elemzőt tol egy *eszközközpontú jövőkép* felé. Nem mindenki elégszik ugyanis meg az olyan, még tartható állításokkal, mint a *Science 2020* szerzői, hogy „a számítástudomány fogalmi világa új absztrakciós szintekkel gazdagítja az egyes diszciplínák kutatóit”. Visszatérően találkozhatunk avval a véleménnyel, hogy az alaptudományok elkerülhetetlen informatizálása a kutatói munkához szükséges informatikai jártasságok és készségek olyan fokú „megjelenítését” igényli a jövő szakembereinél, a programozási ismeretektől az algoritmus-generálásig, ami már-már azonos értékű magára az érintett tudományterületre vonatkozó tudásokkal. Vagy fordítva: magas szintű informatikai ismeretek nélkül nem lesznek művelhetőek a leginkább érintett tudományok (a biológia, az asztronómia és a klimatológia).

Ezek a nézetek a kibontakozó folyamatok természetének többszörös félreértésén alapulnak, és már most látszik, hogy a tudományos nagyüzem egészen másképp oldja meg a gondokat. Miközben minden informatikai jellegű problémának kialakulnak a maga specialistái,²⁹ a kutatóközösség esetében a

bonyolódó eszközkörnyezet kezelhetővé tételének bevált megoldása, az *alkalmazói felületek és programok* fejlesztése bizonyul eredményes útnak. A munkamegosztási lánc érzékelhető hosszabbodása mellett a tudóstársadalom is két részre oszlik (Gray, 2005): az *adatmenedzsmentre szakosodottakra és a feldolgozásban érintettekre*, s közben a szükséges tudások megfelelő kombinációinak előállítása minden további nélkül biztosítható *interdiszciplináris kutatócsoportok (team-ek)* és új, gyakorta virtuális együttműködési formák kialakításával is.

A jelenlegi feldolgozási kapacitásokra méretezett tudomány hatékony intézményi és technológiai megoldásainak közös „metszete” az *intenzifikálás. Az informatikai innovációkkal a meglévő kutatói közösség megnövelt projektméreteiben is alkalmassá válik arra, hogy a kialakult pénzügyi és irányítási mezőben megbirkózzon a saját maga által célirányosan megtermelt új adatok tömegével.* Csakhogy a „mintegy mellékesen” megtermelődő, jóval nagyobb számosságú adat feldolgozása és a felmerülő problémák természetét által igényelt még nagyobb méretű projektek gondozása már meghaladja az intenzifikálásban rejlő lehetőségeket. A megnövelt informatikai teljesítmény ettől a ponttól kezdve a lassan kibontakozó *extenzív fejlődés* támogatójává lesz – vagyis eddigi feladatai közé bekerül a tudományos problémamegoldásba újonnan bevonásra kerülő csoportok mindenoldalú menedzsmentje is.

²⁹ A bioinformatika és a bioinformatikus például mára jól definiált szakmává lettek, de jellemző, hogy mostanra a bioinformatika és az immunológia „hibridje”, az immuninformatika (immunomika) is önálló identitással (és az ehhez illeszkedő kiadvány, illetve konferenciavilággal) bír.

Kulcsszavak: *cybertudomány, cyber-infrastruktúra, cyberkörnyezet, adat-intenzív tudomány, peta skála, GRID, agymunka, tudományos adatközpontok, kollaboratív e-tudomány, varratmentes tudományos szuperhálózat*

IRODALOM:

- Agutter, Jim – Bermudez, Julio (2005): *Information Visualization Design: The Growing Challenges of a Data Saturated World* (AIA Report on University Research) http://www.aia.org/SiteObjects/files/Agutter_color.pdf
- David, Paul A., (2000): *The Digital Technology Boome-nang: New Intellectual Property Rights Threaten Global 'Open Science'*. Paper presented at the World Bank ABCDE (Europe) Conference (Paris) <http://www-econ.stanford.edu/faculty/workp/swp00016.html>
- David, Paul A. – Spence, Michael (2003): *Towards Institutional Infrastructures for E-science: The Scope of the Challenge* <http://129.3.20.41/eps/get/papers/0502/0502028.pdf> Appendix 1.2. 73–74.
- David, Paul A., (2004) *Towards a Cyberinfrastructure for Enhanced Scientific Collaboration: Providing Its Soft Foundation May Be the Harder Part*. SIEPR Discussion Paper No.04-01 <http://129.3.20.41/eps/le/papers/0502/0502004.pdf>
- Dent, Harry S. Jr. (1998): *The Roaring 2000's: Building the Wealth and Lifestyle You Desire in the Greatest Boom in History*. Simon & Schuster, New York
- Dent, Stephen M. (2000): *The New Economy and Partnering*. The CEO Refresher, 2000 <http://www.refresher.com/partnering1.html>
- Emmott, Stephen et al. (2006): *Towards 2020 Science*. Microsoft Corp. 1–86. és: <http://research.microsoft.com/towards2020science/>
- Foster, Ian (2005): *Service-Oriented Science: Scaling the Application and Impact of eResearch*. Keynote Speech on 1st IEEE International Conference of E-science and Grid Computing (2005, Melbourne) <http://www.gridbus.org/science/keynoter.pdf>
- Gray, Jim – Liu, D. T. – Nieto-Santisteban, M. – Szalay, A. S. – DeWitt, D. – Heber, G. (2005): Scientific Data Management in the Coming Decade. *CTWatch Quarterly*. 1, 1, February, <http://www.ctwatch.org/quarterly/articles/2005/02/scientific-data-management/>
- Guernsey, Lisa (2003): Making Intelligence a Bit Less Artificial. The New York Times. (01/05/2003), <http://www.nytimes.com/2003/05/01/technology/circuits/oireco.html?pagewanted=print&position>
- Harney, John O. (2005): Coming Right-Brain Economy. The Connection: New England's Journal of Higher Education, Summer 2005, http://www.findarticles.com/p/articles/mi_qa3895/is_200507/ai_n14800625
- Jéki László (2006): **Számítástudomány 2020-ban.** <http://www.origo.hu/tudomany/technika/20060508szamitastudomany.html>
- Kropotkin, Peter (1890): *Brain Work and Manual Work*. The Nineteenth Century. March, 456–475. http://dwardmac.pitzer.edu/anarchist_archives/kropotkin/brainmanualwork.html
- Kropotkin, Peter (1899): *Fields, Factories and Workshops: or Industry Combined with Agriculture and Brain Work with Manual Work*. Houghton, Mifflin and Co., Boston
- Law, Gillian (2004): *U.K. Researchers Create Robot Biotech Scientist*. IDG News Service, London Bureau 1/16/2004
- Marshall, Alfred (1890): *Principles of Economics*. Macmillan and Co., Ltd. A könyv szövege: <http://www.econlib.org/LIBRARY/Marshall/marPContents.html>
- Myers, Jim (2006): *Re-Engineering the Research Process*. <http://www.interactions.org/sgtw/10/05/2006>.
- Schuldts, Heiko (2005): *Peer-to-Peer Architectures, Grid Infrastructures, and Service-oriented Architectures for Digital Libraries*. <http://agenda.cern.ch/fullAgenda.php?ida=2051871>, (előadás a *Grid Technologies for the Digital Libraries* című szakmai napon)
- Stevens, Rick (2006): Trends in Cyberinfrastructure for Bioinformatics and Computational Biology. *CTWatch Quarterly*. 2, 3, 1–5., <http://www.ctwatch.org/quarterly/articles/2006/08/trends-in-cyberinfrastructure-for-bioinformatics-and-computational-biology/>
- Vajda Ferenc (1999): Tudományos kutatás és együttműködés informatikai bázison. *Magyar Tudomány*. 4, 459–463.