

# Információtechnológia és a tudomány gyermeke: az eScience



Adatözönben élünk. Műholdak térképezik fel Földünk felszínét és időjárását, távcsövek pásztázzák automatikusan a Világegyetem egyre távolabbi régióit, elektronikus berendezések rögzítik a nagyenergiás részecskeütőközésekből kijövő számtalan részecske összes adatát. Személyes adatainkat hitelfelvételi, vásárlási és utazási szokásainkról magán- és közintézmények gyűjtik, a humán genom projekt keretében pedig elkészült az első emberi géntérkép. Az interneten felgyülemelő adatmennyiség ma már csak *exabájt* nagyságrendben mérhető, és ez is kétévente duplázódik. Az összegyűjtött adatok kiértékelése általában nagyobb gondot okoz, mint maga az összegyűjtésük, mert sokszor nehéz átlátni az adatok közötti összefüggéseket, és kiszűrni belőlük a lényeges információkat: előre jelezni esetleges katasztrofális jelenségeket a természetben vagy a hálózatokban, felismerni a terroristagyanús személyeket, megtalálni az öregedésért és örökletes betegségekért felelős géneket, azonosítani a potenciálisan piac képes termékeket, megbecsülni egy-egy döntés kockázatát.

Felvetődik a kérdés, hogy miként lehetne a – fentiekben példaként kiragadott esetekben meglevő – hatalmas adatözönt kezelni. A választ az *eScience* jelentheti.

## Mi az eScience?

Az eScience kifejezést egy új, „feltörekvő” technológiára használják, melynek révén nagyléptékű, komplex tudományos tevékenység fejthető ki a modern információtechnológia felhasználásával. Legfőbb jellemzője a rendkívül sok, gyakran különböző helyekről elérhető adattal operáló kiértékelő munka, melynek eredményes

véghezviteléhez az adatok automatikus gyűjtésére, optimális adatbázisba rendezésére, rendkívül nagy számítástechnikai kapacitást igénylő feldolgozására, és a lényegyet megragadó vizualizációra van szükség.

Az eScience az alkalmazott tudomány, az alapkutatás, valamint a modern információtechnológia (IT) interdiszciplináris egymásra hatásából született, és egy új minőséget képvisel a kiértékelésben. Az IT kifejleszti a korszerű eszközöket és informatikai rendszereket az alap- és alkalmazott kutatások mai problémáinak megoldásához. Az így keletkezett rendszerek, illetve az általuk szolgáltatott adatok, komplexitásuk miatt, már csak az eScience segítségével vizsgálhatók. Az eScience – jelenlegi fejlettségi szintjén – elsősorban a szervezett adatgyűjtés anyagának feltárására irányul, ugyanakkor a spontán felhalmozódó adatmennyiség feldolgozásához is előkészíti a szükséges technológiát.

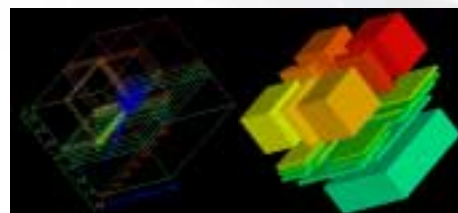
## Miért fontos az eScience?

Az elmúlt évtized jelentősen megváltoztatta szinte minden tudomány módszertanát. Ahhoz hasonlóan, ahogy a matematika egy évszázaddal ezelőtt szétáradt szinte minden

A Sloan Digital Sky Survey 120 megapixeles kamérajája az elmúlt években 40 terabájt adatot termelt



tudományban, és az elméleti modellek alapeszközévé vált, ma már szinte lehetetlen bármelyik tudományágat a modern információs technológiák használata nélkül művelni. Ezzel párhuzamosan egy másik irányú folyamat is megindult: a modern információs rendszerek anynyira összetetté és bonyolulttá váltak, hogy vizsgálatukhoz, fejleszté-



Galaxisok paramétereinek adaptív cellázása bináris keresőfával

sükhöz, valamint az általuk szolgáltatott adatok kiértékeléséhez az informatika eddigi módszerei már nem elégségesek. Szükség van a hagyományos, de bonyolult természeti jelenségek vizsgálatára kidolgozott módszerek eszköztárára is.

Sem a tudomány, sem az elektronika történetében nem ez az első forradalom. Az olcsó személyi számítógépek megjelenése megváltoztatta az informatika viszonyát a tudományhoz, az üzleti és a mindennapi élethez. Az olcsó internet forradalmasította a munkacsoportok működését, az itt közölt eredmények mindenki számára gyorsan elérhetővé váltak. Ma pedig a szemünk előtt zajlik az adattechnológia forradalma: rohamosan nő a műszerek, érzékelők, felvevő berendezések sávszélessége és mennyisége, valamint az internet-felhasználók száma, akik hála a ma már könnyen kezelhető webtechnológiának, szinte ontják a különböző értékű adatokat. Mindez oda vezetett, hogy napjainkra mind

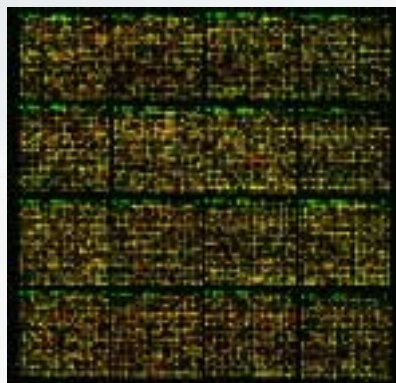
az igények, mind a már létező kapacitások kinőtték az informatika által biztosított eddigi kereteket: a már meglévő adatbázisok sokszor szét-szórtnak, „ömlesztve” tárolják az adatokat, az új adatgyűjtési eljárásoknak egyre nagyobb mennyiségű adatot kell automatikusan begyűjteniük. A meglévő adatokat egyre nagyobb számítástechnikai teljesítménnyel lehetne csak feldolgozni, és az adatbázist használók sokszor sötétben tapogatóznak, mert nem mindig tudják, hogy pontosan milyen adatokra van szükségük, illetve, hogy mi mindenre lehetne használni az összegyűjtött információt.

Az eScience technológiája, bár a tudomány eredményeiből gyökerezik, a mindennapi élet számos területén is alkalmazható. Mindenütt alapkövetelmény a jó trendek felismerése és kiaknázása. A megfelelő irányban tett gyors elmozdulást általában siker kíséri: példa erre a Google-portál, vagy az Amazon e-kereskedeleme több – az összegyűjtött adatok alapján megtervezett és jól célzott – reklámakciója. Ugyanakkor az e-buborék kipukkanása az ezredforduló körül jelezte, hogy számos technológia még nem érett be.

## Mit gondolnak mások?

Az eScience névvel jellemzett új technológia kifejlesztésének és alkalmazásának szükségessége már felértékelődött az USA Nemzeti Tudományos Alapjának, a kanadai Innovációs Alapnak, valamint az angol Nemzeti eScience Központ bizottságainak ülésein, ahol megállapították, hogy az adatgyűjtés, tárolás és karbantartás kezelésére új megközelítést kell kidolgozni, mivel az eddigi

eljárások rohamosan bonyolódnak, és egyre több pénzt fognak felémeszteni. Az IBM kivonul a hardver-üzletágból, helyette adatbázis-rendszerekre és a bioinformatikára helyezi a hangsúlyt, a Microsoft Research-nél hasonló irányvonal figyelhető meg. A *New Scientist* 2005. áprilisi száma a rákkutatás jelenlegi helyzetéről közöl áttekintést, melyből kitűnik, hogy ezen a területen szűk keresztmetszetet jelent olyan szakemberek hiánya, akik egyrészt



A génchipek által termelt adatmennyiség rohamosan növekszik

birtokában vannak bizonyos biotechnológiai ismereteknek, ugyanakkor szakszerűen képesek kezelni az eScience eszköztárát.

## Mi a jövő?

A számítástechnika és a kutatás-fejlesztés küszöbön álló forradalma a következő technológiákra fog épülni: az egyre nagyobb adatözmön olcsó begyűjtésére, strukturált tárolására, a különböző helyeken elszórt kapacitásokhoz való egyszerű hozzáférésre, valamint az adatok és keresések megjelenítésére. Mivel hatalmas adattömegek a tudósok műszereinél jelentek meg először, a kutatók már sok tapasztalatot szereztek ezeken a területeken, ebből adódik, hogy a tudományinformatika az eScience élén járhat olyan technológiák előállításában – illetve hazai átvételükben és továbbfejlesztésükben –, amelyek az élet egyéb területein is kamatoztathatóak.

Az adatbázisok összekapcsolásával létrejövő struktúra (ezt a csillagászatban Virtuális Obszervatóriumnak nevezik) egy „virtuális szervezet” definiál a hálózaton, melynek tagjai távolsági korlát nélkül képesek kihasználni annak erőforrásait. Az ilyen jellegű szervezetségre a közeljövőben a tudomány mellett elengedhetetlen szüksége lesz a nagyvállalati munkának, az államigazgatásnak, környezetgazdálkodásnak, a nemzetbiztonságnak, de hamarosan még a kulturális szférának is. Egy ilyen szervezet ugyanakkor jelentősen megkönnyítheti a kis- és középvállalkozások számára is az adatokhoz, valamint az olcsó feldolgozókapacitáshoz való hozzájutást, ezáltal javítva esélyeiket a versenyben.

## Hol tart a technológia?

A fenti technológiák intenzív fejlődési szakaszban vannak. Léteznek nagyon jól használható alapmodulok, melyekre már lehet építeni, de még nem következett be az a komoly átöröklés, amely a fenti problémák hatékony megoldását nyújtaná. Ennek oka egyrészt az, hogy az ipar még nem állt rá egy kiforrott technológia „gyártására”, így a kutatás-fejlesztésnek elsődleges szerepe van ezen a téren. Másrészt, ugyanezen okból, nagyon nagy lehetőségek rejlenek a témában: az, aki az első használható rendszer létrehozásában vezető szerepet játszik majd, könnyen válhat akár kis „garázscégből” is piacvezetővé, mint ahogy annak idején ezt a folyamatot a mobil- vagy internetes technológiáknál, illetve a személyi számítógépek piacán láthattuk.

A fenti témakörben folyó kutatások előmozdítására jött létre a Nemzeti Kutatási és Technológiai Hivatal Pázmány Péter programjának keretében 2005 végén az eScience Regionális Egyetemi Tudásközpont, melyről részletes információ a <http://ret.elte.hu> címen található.

Csabai István, Papp Gábor  
Eötvös Egyetem, eScience  
Regionális Egyetemi Tudásközpont

Galaxisok paramétereinek Voronoi cellázása adatbázis-indexeléshez

