

# Köznevelés az adatok bővületében – az adattermelés forradalma

Az adattermelés forradalma szinte kivétel nélkül végigsöpör minden tudományágon.

Hamarosan a pedagógia<sup>1</sup> is célkeresztbe kerül. Vajon feltettük azokat a kérdéseket és megadtuk azokat a válaszokat, melyek felkészültségünkről tanúbizonyságot tesznek? Van elképzelésünk arról, hogy mit jelent az adattermelés forradalma a pedagógiában? Esetleg már kész terveink vannak a közeljövő problémáinak és lehetőségeinek a hatékony kezelésére? Ha a kérdésekre igen a válaszunk, megnyugodhatunk. Ha esetleg mégsem, akkor – miként az alábbiakban is – folytassuk a problémák felvetését és a válaszok keresését mindezen területeken, még ha e válaszok olykor vitára is adnak okot.

Az adattermelés forradalma két – látszólag talán nem annyira jelentős – következménnyel jár a pedagógia vonatkozásában: míg a rendelkezésre álló adatok száma egyre nagyobb ütemben növekszik, addig az adatok aggregáltsági szintje folyamatosan csökken. A következmények végiggondolása olyan paradigmaváltást jósol, mely a pedagógia évszázados hagyományokra visszatekintő rendszerét alapjaiban fogja megváltoztatni.

## Az adatmennyiség növekedése a gyakorlatban

A rendelkezésre álló adatok számának növekedését jól példázza, hogy míg Galileo Galilei – korának egyik legnagyobb tudósa – 3 bájt/éjszaka sebességgel „termelte” az adatokat,<sup>2</sup> addig ma becslések szerint 2,2 millió terabájt adat keletkezik naponta.<sup>3</sup>

---

<sup>1</sup> A pedagógia fogalmát tanulmányunkban a tanítás – tanulás folyamatával összefüggésben értelmezzük, az adatfelhasználásban rejlő lehetőségeket elsősorban ebben a kontextusban kívánjuk hangsúlyozni. Ugyanakkor a tanulmány több pontján a pedagógiát is magába foglaló köznevelési ágazat egészére vonatkozó megállapításokat teszünk, amennyiben az adatvagyon csak ágazati szinten értelmezhető. Ennélfogva az ágazati szintű adatvagyon egyik felhasználási területe a pedagógia.

<sup>2</sup> Forrás: <https://indico.kfki.hu/event/103/contribution/0/material/slides/0.pdf>

<sup>3</sup> Forrás: [http://www.portfolio.hu/vallalatok/it/elkepeszto\\_mennyi\\_adat\\_letezik\\_mire\\_lehet\\_felhasznalni.186053.html](http://www.portfolio.hu/vallalatok/it/elkepeszto_mennyi_adat_letezik_mire_lehet_felhasznalni.186053.html)

► *Educatio* 2015/3. Salomvári György: *Köznevelés az adatok bővületében – az adattermelés forradalma*, 73–85. pp.

Hasonló jelenség a köznevelésben is tetten érhető: míg az 1997-ben először említett Köznevelés Információs Rendszere (továbbiakban: KIR)<sup>4</sup> kezdetben 3 adatbázissal indult, addig ma az adatbázisok száma közel 500 db. Természetesen a nyilvántartott adatok száma is folyamatosan bővül: például 2015. szeptember 3-án reggel 7:00 és délután 18:00 óra között 273 950 db elektronikus kérelmet fogadott be és dolgozott fel a KIR, melynek eredményeképpen milliós nagyságrendű adattal bővült a nyilvántartott adatok száma – csak ezen az egy napon. Ma a KIR hozzávetőlegesen 4,1 millió magyar állampolgárról tartalmaz részletes adatokat, és akkor még nem beszéltünk a nem személyekkel kapcsolatban keletkező adatokról.

## Az adattermelés során keletkező adatok jellemzői

Mielőtt az aggregátsági szint csökkenésére rátérünk, célszerű a termelt adatok definíció alapú kategorizálásával röviden foglalkozni – elsősorban logikai megközelítést alkalmazva.

Az *elemi adat* fogalmának definiálása tanulmányunk szempontjából rendkívül fontos. Az „elemi adat” valójában egy olyan tulajdonság, mellyel a meghatározott feltételeknek eleget tevő adatokat jellemezzük. Az elemi adatok az adott adatrendszer keretein belül nem állíthatók elő más adatokból, tehát nem matematikai művelet eredményeképpen jönnek létre, másrészt nem bonthatók tovább kettő vagy több további adatra. Az elemi adat tulajdonság időben dinamikusan változó jellemző: megszűnik, ha adatkör-bővítés során olyan adatok kerülnek be a rendszerbe, melyek biztosítják a korábban elemi adatként minősített adat műveletekkel történő előállíthatóságát.

A keletkezési helyükön létrejövő adatokat az egyszerűség kedvéért nevezzük *forrás-adatoknak*. A forrásadatok tipikusan elemi adatként minősülnek, vagyis műveletek segítségével nem állíthatók elő más adatokból. Például egy szóbeli felelet során szerzett érdemjegy elemi adatként minősülő, intézményen belüli forrás adat.

Ezzel szemben a *számított adat* egyszerű matematikai művelet segítségével előállított, új adat. A matematikai művelet vonatkozhat elemi adatok vagy számított adatok tetszőleges kombinációjára. Például egy írásbeli dolgozatra adott érdemjegy adott esetben már nem minősül elemi adatként, miután az egyes feladatokra adott részpontszámok alapján számítással hozható létre.

A *statisztikai adat* általában műveletek sorozatával jön létre. Az adatforrások „környékén” tipikusan még számított adat, azonban az adatfeldolgozás eredményeképpen adatrendszerekbe bekerülve már nagyon gyakran minősül elemi adatként, vagyis adott adatrendszeren belül nem bontható fel további elemi adatokra. (A definíciók értelmezését a következő, aggregátsági szinttel foglalkozó rész nagymértékben segíti.)

Az adattermelés során létrejövő adatok egy része – különösen a pedagógiában – konkrét személyekkel kapcsolatban keletkezik, az adott személyeket jellemző információ. Ha személyek azonosítására – azaz az adott személy és a vele kapcsolatban keletkező infor-

<sup>4</sup> A „közoktatási információs rendszer” kifejezés először egy 1997-ben, az akkori Művelődési és Közoktatási Minisztérium által kiírt pályázatban jelent meg. A KIR 1999-től jogszabályi felhatalmazás alapján működik, jelenleg a 2011. évi CXCV. törvény a nemzeti köznevelésről, valamint a 229/2012. (VIII. 28.) Korm. rendelet a nemzeti köznevelésről szóló törvény végrehajtásáról szabályozza működését. Továbbá fontos megjegyezni, hogy a KIR jogilag nem tartalmazza az összes köznevelési adatot kezelő rendszert, ugyanakkor – az egyszerűség kedvéért – tanulmányunkban ezeket is a KIR fogalma alatt tárgyaljuk.

máció egyértelmű összekapcsolására – alkalmas és szükséges természetes adatok rendelkezésre állnak, *személyazonosító adatokhoz kötött egyéni szintű adatokról* beszélhetünk. (A természetes személyazonosító adatok a következők: családi és utónév, születési családi és utóneve. Amennyiben a felsorolt adatok csak részben állnak rendelkezésre, a személy azonosítása nem végezhető el 100%-os biztonsággal.)

Pedagógiai kutatási szempontból különösen fontos megérteni, hogy a személyazonosításra alkalmas adatok és az adott személlyel kapcsolatban keletkező adatok közötti kapcsolat megszüntethető – ebben az esetben beszélhetünk *egyéni szintű, de személyazonosításra nem alkalmas adatokról*. Másképpen abból, hogy egyéni szintű adatok birtokában vagyunk, még nem következik a konkrét személy beazonosíthatóságának lehetőségége – amennyiben az egyéni szintű adatok nem tartalmazzák a személy azonosítására alkalmas adatokat vagy a személy azonosítására alkalmas adatokhoz vezető kódokat.

### Mit értünk az aggregáltsági szint csökkenésén?

A korábban ismertetett KIR adatok egyértelműen megmutatták számunkra, hogy az adattermelés a köznevelés egész rendszerében jelen van, ezen belül a pedagógia is aktív részese a folyamatnak. Az előre jelzett paradigmaváltásnak ugyanakkor szükséges, de nem elégséges feltétele az adatok mennyiségének növekedése, az aggregáltsági szint csökkenése ugyanúgy nélkülözhetetlen.

Ezek után az aggregáltsági szint csökkenése nem más, mint az adatrendszerekben kezelt elemi adatok „átváltozása” statisztikai vagy számított adatból forrásadattá. Vagyis a magas aggregáltsági szintet képviselő statisztikai adatok helyett a lehető legalacsonyabb aggregáltsági szintet képviselő forrásadatok kerülnek előtérbe. Természetesen az adatmennyiség és az aggregáltsági szint szoros összefüggést mutat: amennyiben nem statisztikai adatokat gyűjtünk, hanem az összes olyan forrásadatot, ami lehetővé teszi különböző számítások elvégzését, lényegesen több adat kezelésére van szükség. (A jelenség nyilvánvalóan technológiafüggő, amennyiben a számítási, adattárolási és adatátviteli kapacitások jelentős bővülése peremfeltételnek minősül.)

A példánknál maradvá egy korszerű, pedagógiai célú adatrendszernek nem az osztályatlagra van szüksége, de még csak nem is a tanulónkénti érdemjegyekre, hanem tanulónként a szóbeli feleletek érdemjegyeire és az írásbeli dolgozatok feladatonkénti pontszámára. Vegyük észre, hogy mindkét esetben elemi adatokra „vadászunk”, melyek forrásadatként már nem bonthatók szét további adatokra. Másképpen fogalmazva, míg korábban az adatrendszerek statisztikai adatokat tartalmaztak elemi adatként, addig ma – az aggregáltsági szint csökkenésének következtében – az adatrendszerek elemi adatai leginkább forrásadatok.

Az aggregáltsági szint csökkenése a köznevelési adatrendszerekben is megfigyelhető: például a köznevelés statisztikai adatgyűjtés 1999-ben még papír alapú adatlap kitöltésével történt – ezáltal kizárólag magas aggregáltsági szintet képviselő adatbázis jött létre, addig ma már részben a gyakorlatban is megvalósuló erős törekvés, hogy a statisztikai adatok jelentős részét egyéni szintű elemi adatokból generáljuk.

## Az adattermelés forradalmának következményei a pedagógiában

A leírtak alapján nem lehet kétségünk afelől, hogy az adattermelés forradalma a köznevelési ágazatban, ezen belül a pedagógiában is jelen van. Mind az adatok mennyiségének növekedése, mind pedig az aggregáltsági szint csökkenése egyértelműen erre utal. Ugyanakkor még mindig jogosan tehetjük fel a kérdést: miért jelent változást a pedagógiában ez a jelenség?

„Olyan mennyiségű és olyan minőségű adat, amivel ma az egyéni tanulói nyomkövetést lehetővé tevő rendszerek rendelkeznek, soha nem állt a rendelkezésünkre. Ezeknek az adatoknak köszönhetően az iskolai folyamatok „mélyrétegeibe” tudunk belelátni, az oktatási rendszerek immár szinte semmit nem tudnak „elrejtteni”. Látni lehet minden egyes iskolát, minden egyes tanulót, és bizonyos rendszerekben kisebb-nagyobb pontossággal minden egyes tanárt és minden egyes kurzust. Látjuk ezek teljesítményének, eredményességének időbeli mozgását.” (Halász, 2010)

Tehát az adattermelés eredményeképpen létrejövő adatvagyon kifejezetten alkalmas közvetlen pedagógiai célú alkalmazásra, miután a statisztikai megközelítést felváltja az egyéni szintű elemi adatok világa, ami alapvető és egyben sok mindent felforgató újdonságnak tűnik – így méltán beszélhetünk paradigmaváltásról.

A továbbiakban túllépünk az adattermelés következményeinek a feltárásán, és a létrejövő adatvagyon pedagógiai célú alkalmazási lehetőségeit és az alkalmazással kapcsolatos anomáliákat próbáljuk körüljárni.

## Az adatvagyon-hasznosítás új lehetőségei a pedagógiában

Az adatvagyon olyan rendezett adathalmaz, melyet az adattermelést követő adatfeldolgozás hoz létre. Az adatvagyon – akár pénzben is kifejezhető – értéket képvisel, mely a felhasználása során jelenik meg és válik mérhetővé. Az adatvagyon létrehozásáról és jellemzőiről később szó lesz, egyelőre nézzük meg, hogy az adatvagyon felhasználása milyen előnyökkel kecsegtet a pedagógiában, milyen konkrét pedagógiai tevékenységek véghezvethetők el adatok alkalmazásával – a teljesség igénye nélkül, felsorolásszerűen:

- állapotok jellemzése
- időbeli változások követése, trendelemzés
- változások irányának és nagyságának felmérése, trend-előrejelzés
- előzetes várakozások és/vagy célkitűzések megfogalmazása
- folyamatok modellezése, várható hatások előrejelzése
- folyamatok induló és záró állapotának összehasonlítása
- összehasonlítások széles köre
- indikátorok előállítás
- hatékonyság, eredményesség és méltányosság kifejezése (Lannert, 2004)
- kompetenciák mérése
- gyors és egyszerű kommunikáció pl. a szülők felé
- döntések előkészítése, hatásaik feltérképezése.

A felsorolt lehetőségeket alaposan szemügyre véve könnyen juthatunk arra a következtetésre, hogy valójában az adatvagyon-hasznosítás semmi olyan újdonságot nem hoz be a pedagógia rendszerébe, amit egyébként nem végzünk hagyományosnak mondott eszközökkel. Ha ez valóban így van, akkor mégis miért beszélünk paradigmaváltásról?

A probléma feloldásához az erőforrás-hatékonyság környékén szükséges kutakodnunk. Ahhoz, hogy a felsorolt tevékenységeket hagyományos pedagógiai módszerekkel színvonalasan elvégezzük, meglátásunk szerint minden egyes gyermek és tanuló körül kiválóan képzett, széles körű módszertani repertoárral felvértezett pedagógusoknak és egyéb szakemberek hadának kellene tevékenykedni – ráadásul sok esetben az egyes személyekkel külön-külön foglalkozva. (Lehet, hogy ezt hívjuk ideális iskolának?) Ezzel szemben az adatvagyon alkalmazása biztosan nem nyújtja a személyes kapcsolatokra épülő pedagógia minőségét – ellenben kiegészítve azt hatékony erőforráskezelést tesz lehetővé úgy, hogy közben egyetlen gyermeknek vagy tanulónak sem kell kimaradni az adatvagyon felhasználására is építő pedagógiai rendszerből. (Halász, 2007)<sup>5</sup>

A megfontolásaink alapján úgy tűnik, hogy az adatok felhasználása nem abban jelent újdonságot, hogy különleges és újszerű pedagógiai tevékenységeket hoz létre (bár ez sem zárható ki), hanem sokkal inkább abban, hogy a korábban is végzett tevékenységek tömeges, lényegében minden egyes tanulóra kiterjedő alkalmazhatóságát biztosítja elfogadható színvonalon. (Az „elfogadható színvonal” elsősorban társadalmi közmegegyezés tárgya, de természetesen szorosan összefügg az adatvagyon formájában rendelkezésre álló adatok mennyiségével és aggregáltsági szintjével is.)

## Az adatok pedagógiai célú felhasználásának globális kockázata

Mielőtt önfeledten hátradölnénk és megelégedve szemlélnék a kibontakozó adatalapú „szép új világot” – ami ráadásul meglehetősen költséghatékonyak is tűnik –, azonnal ejtsünk szót a legnagyobb globális kockázatról, mely álláspontunk szerint ugyan nem kérdőjelezi meg az adatok széles körű felhasználhatóságát, viszont igencsak óvatossá kell, hogy tegyen bennünket. A legnagyobb globális kockázat felismeréséhez a következő természettudományos szemléletű idézet pedagógiára vonatkozó értelmezését javasoljuk:

„A feltételezés az, hogy ha egyszer megértjük a részeket, akkor könnyű lesz az egészet felfognunk. [...] Közel vagyunk ahhoz, hogy majdnem mindent tudjunk, amit a részekről tudni lehet. A természet egészének megértésétől azonban ugyanolyan messze vagyunk, mint bármikor korábban. [...] Az ok egyszerű. Az egyszerűsítést erőltetve beleütközünk a komplexitás kemény falába. Megtanultuk, hogy a természet nem egy jól megtervezett összerakós játék, amely csak egyféleképpen rakható össze. [...] Manapság egyre inkább felismerjük, hogy semmi nem történik elszigetelten. A legtöbb esemény és jelenség része egy komplex, univerzális kirakójátéknak, amelynek sok-sok darabja egymással kapcsolatban és kölcsönhatásban áll, egymást befolyásolja.” (Barabási, 2013)

A pedagógia ilyen értelemben szintén egy kirakójáték, ahol semmi nem történik elszigetelten, az események és jelenségek egy komplex rendszer részei, melyek sok-sok darab-ból állnak össze, ráadásul az egyes darabok kapcsolatban és kölcsönhatásban állnak, egymást befolyásolják.

Az adatok használatának legnagyobb kockázata tehát a komplexitás elvesztése: az egyes részeket le tudjuk írni adatokkal, de közben az egészet – vagyis a gyermeket és tanulót, mint komplex rendszert – szem elől tévesztjük. A kockázat valós, minden egyes

<sup>5</sup> Az Egyesült Államokbeli No Child Left Behind átfogó oktatásfejlesztési program – mely korszakos jelentőségű – legfontosabb üzenete éppen az, hogy „kivétel nélkül minden egyes gyermek tanulása sikeressé válhasson”. L.d. bővebben: HALÁSZ G.: Tényekre alapozott oktatáspolitikai, [http://halaszg.ofi.hu/download/Evidence\\_based\\_study.pdf](http://halaszg.ofi.hu/download/Evidence_based_study.pdf)

adatfelhasználónak minden egyes alkalommal tisztában kell lennie az adatfelhasználás korlátaival. Az adatvagyon-alkalmazás feltételeinek megteremtése nem lehet cél, csak és kizárólag eszköz.

## Az adatvagyon hasznosításának sajátosságai a pedagógiában

A meteorológia az adatfelhasználásra erősen építő tudományág. Minél nagyobb számú elemi adat áll rendelkezésre, és minél pontosabbak a folyamatok leírására alkalmazott matematikai modellek, annál pontosabban és részletesebben leírhatók és előre jelezhetőek az időjárásban bekövetkező események.

Az egészségügy szintén épít az adatfelhasználásra: ha egy egészséges ember állapotát leíró adatokkal tisztában vagyunk, az állapot leírására használt adatokban bekövetkező változások azonnal jelzésértékűek.

A pedagógia esetében hasonló eljárást kellene követni: egyéni szintű elemi adatok segítségével részletesen leírni az adott személy aktuális állapotát, majd megfogalmazni egy jövőben célként elerendő állapotot. Ha ez megvan, a pedagógiai tevékenység a kezdeti állapotból kiindulva a célállapot elérésére irányul. A pedagógiai tevékenység megtervezése és végrehajtása épülhet adatvagyon felhasználásra: ha a tudásbázis tartalmaz olyan korábbi – elemi szintű adatokkal részletesen dokumentált – esetet, amikor hasonló kezdeti állapotból különböző személyek már „eljutottak” a megfogalmazott célállapotba, nincs más dolgunk, mint a „célravezető” tevékenységet újra lefolytatni. Minél nagyobb és részletesebb a rendelkezésre álló adatvagyon, annál nagyobb valószínűséggel választunk korábban már bevált pedagógiai tevékenységet – legalábbis az adatvagyon-hasznosítás általános módszertana szerint.

Végtelenül leegyszerűsítve (és ezzel egyben a szakszerűséget is kockáztatva) a betegellátás sok esetben hasonló logika szerint működik: ha a testhőmérsékletre vonatkozó adat nem megfelelő, beavatkozás kezdődik a kiinduló (lázás) állapot megváltoztatására, a kívánt célállapot elérésére. A szükséges beavatkozás kiválasztása korábbi esetek adataira épül, amikor is a kiinduló állapotból sikerült a célállapot elérése.

Ezzel az egyébként rendkívül technokrata megközelítéssel alapvetően két probléma is van a pedagógia vonatkozásában: legyen társadalmunk demokratikus, ami esetünkben azt jelenti, hogy a követendő példák sokféleségével számolunk, és a sokféleséget elismerjük. Még ha egy adott személy aktuális állapotát kellően sokszámú elemi adattal le is tudnánk írni, az elerendő célállapot meghatározása lehetetlenül nagy vállalkozásnak tűnik. Ráadásul további problémát okoz az időbeliség: pedagógiai rendszerekben a beavatkozások hatása sok esetben rendkívül hosszú idő elteltével jelentkezik, vagyis soha nem lehetünk abban biztosak, hogy egy időben lényegesen később megfigyelhető állapot milyen beavatkozások és egyéb hatások eredményeképpen jött létre.

Következtetésünk ezek után az, hogy a pedagógia esetében olyan speciális körülmények is fontos szerepet játszanak az adatvagyon hasznosítása során, melyek feltérképezése rendkívül komoly kihívást jelent a szakterület kutatói és egyéb szakemberei számára. A pedagógia álláspontunk szerint nehezen hasonlítható össze más, adatvagyonra építő szakrendszerekkel az adatvagyon-hasznosítás módszertanát illetően, ugyanis az adatmennyiség további növelése és az aggregáltsági szint csökkenése – ellentétben más rendszerekkel – számos kérdésre elvileg sem adhat kielégítő választ, többek között az itt felvetett problémákra sem.



## Az adatvagyon-hasznosításra is támaszkodó pedagógiai rendszerek működtetése

Amikor egy történelemtanár barátomnak megemlítettem, hogy micsoda lehetőségek vannak a pedagógiát támogató adatvagyon tantermi hasznosításában, meglehetősen furcsán nézett rám...

Pedig ezen a területen is alapvető szemléletváltozásra van szükség.

Első gondolatunk talán az, hogy az adatok kezelése informatikai szakemberek feladata – holott ez ma már messze nincs így. Először is rengeteg olyan önkiszolgáló rendszer létezik, mely lényegében informatikai tudás nélkül is lehetővé teszi a rendelkezésre álló adatok korlátlan felhasználását. Másrészt az informatikusok két lényeges dologban nem tudnak segíteni: a szakterületi igények megfogalmazása és az adatokból előállítandó információk meghatározása, valamint a kapott eredmények értelmezése és kiértékelése egészen bizonyosan nem az informatika hatáskörébe tartozik, márpedig éppen ezek a kulcsfontosságú momentumok. Meggyőződésünk, hogy informatikai megvalósíthatóság és kiszolgáltatósság ma már nem jelent korlátot az adatvagyon pedagógiai célú hasznosításában.

Az adatvagyon-hasznosításra építő pedagógiai rendszerek szempontjából kulcsfontosságú a kutatással foglalkozó szakemberek tevékeny részvétele a folyamatokban. Miért kellene minden egyes pedagógusnak és a neveléssel–oktatással kapcsolatba hozható szakembernek saját magának kitalálni, hogy adott helyzetben milyen elemi adatokra van szüksége, azokat feldolgozni és a kapott eredményeket módszertani szempontból kiértékelni hogyan célszerű? Persze a kész megoldások konfigurálása és helyi viszonyokhoz illesztése már más lapra tartozik, ehhez az oktatáskutatás inkább csak általános módszertani segítséget tud biztosítani.

Ezek után a feltett kérdésre javasolt válaszunk a következő: az adatvagyon-hasznosításra támaszkodó pedagógiai rendszereket a pedagógiai tevékenységben közvetlenül érintett pedagógusok és egyéb szakemberek működtetik – szoros kapcsolatban a támogatást nyújtó, kutatásokkal rendszer szinten foglalkozó szakemberekkel.

Mint ahogy azt oly sokszor látjuk, a pedagógusok ismét csak kulcsfontosságú szereplői a pedagógiai rendszereknek, az adatvagyon pedagógiai célú hasznosításának.

### Kihívások a pedagógusok számára

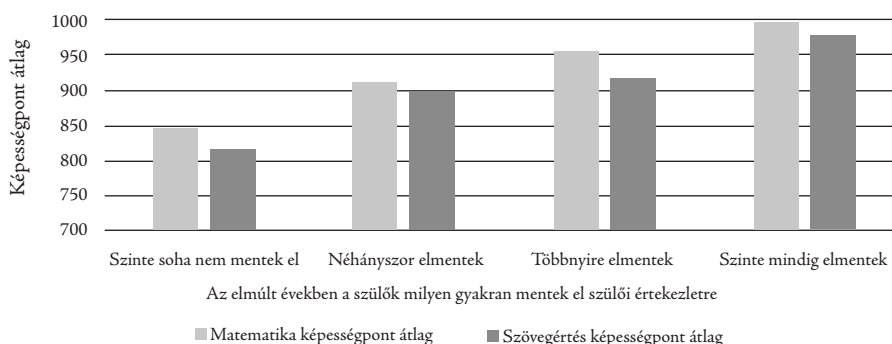
Az adatvagyon hétköznapi felhasználását biztosító információsrendszer-szolgáltatások kezelése valamennyi pedagógus számára könnyedén elsajátítható, ezáltal egyszerűen és gyorsan létrehozhatók a legkülönfélébb kimutatások – ez tehát nem jelent kihívást. Ezzel szemben a kapott eredmények értelmezése és pedagógiai célú alkalmazhatóságának megítélése alapos felkészültséget és nagy körültekintést igényel.

Éppen ezért talán az egyik legnagyobb dilemma napjainkban, hogy az oktatáskutatással foglalkozó szakemberek és a pedagógusok között hol húzódjon az adatvagyon-felhasználás határa. Mekkora veszélyt jelent, ha a pedagógusok olyan eszközt kapnak a kezükbe, mely eszköz alkalmazására egyelőre még nincsenek felkészülve?

A probléma érzékeltetésére nézzünk egy példát. A következő táblázat a tesztelési fázisban lévő KIMER<sup>6</sup> segítségével készült, mely a felhasználók számára sok tekintetben korlátlan adatkonfigurációs lehetőséget biztosít. A használata rendkívül egyszerű, a táblázat elkészítése 1-2 percet vesz igénybe.

2014. évi adatok	Az elmúlt években a szülők milyen gyakran mentek el szülői értekezletre?			
	Szinte soha nem mentek el	Néhányszor elmentek	Többnyire elmentek	Szinte mindig elmentek
Matematika képességpont átlag	846,4	912	956	998
Szövegértés képességpont átlag	815	901	918,8	978

Kompetenciamérés átlageredmények a szülői értekezlet látogatási gyakorisága függvényében



A számadatok és grafikon alapján akár azt a következtetést is levonhatnánk, hogy a kompetenciamérés eredményeit úgy tudjuk javítani, ha növeljük az értekezleteken részt vevő szülők számát – ami nyilvánvalóan téves következtetés még akkor is, ha egyébként a tényadatok ilyen jellegű kombinációja éppenséggel erre utaló eredményt ad.

Vegyük észre, hogy a bemutatott példával ellentétben nem mindig ilyen nyilvánvaló a téves megközelítés, ezért is szeretnénk felhívni a figyelmet az adatértelmezés jelentőségére.

Meggyőződésünk, hogy az adatvagyon szakszerű felhasználásának kialakulása hosszú ideig tartó evolúciós folyamat, mely rendkívüli körülményt igényel – ugyanakkor az is nyilvánvaló, hogy a szűk kutatói kör által alkalmazva az adatvagyon által biztosított lehetőségek kiaknáztatlanok maradnak.

## Létezik-e komplex adatvagyon-kezelő rendszer?

Előjáróban: a köznevelés területén belül az adatvagyon pedagógiai célú alkalmazása talán a legkevésbé kidolgozott, egyben a legizgalmasabb feladat is. Ugyanakkor a rendelkezés-

<sup>6</sup> A Köznevelés Indikátor Menedzsment Rendszere (KIMER) az Oktatási Hivatal által a TÁMOP-3.1.8-09/1-2010-0004 Átfogó minőségfejlesztés a közoktatásban című kiemelt uniós projekt keretében készült. Fontos megjegyezni, hogy az időközben a *kimer.oktatas.hu* oldalon publikált rendszer végleges változata már nem tartalmazza a bemutatott táblázat előállításának lehetőségét.



re álló adatok között megfigyelhető szoros és szerteágazó összefüggérendszer nem indokolja, hogy az adatvagyonot részekre szabdalva képzeljük el, és így önálló tanügyigazgatási, kutatási vagy akár önálló pedagógiai adatrendszert hozzunk létre. A továbbiakban az adatvagyon egységes eszkézként kezeljük – miközben tisztában vagyunk az alkalmazási lehetőségek és jogosultságok széles spektrumával.

Ezek után nézzük a köznevelési adatvagyon létrehozásának logikai folyamatát, mely álláspontunk szerint a következő lépésekből áll:

1. Koncepció létrehozása és elfogadtatása
2. KIR Adatvagyon-kezelő Központ (továbbiakban: KIRAK) létrehozása
3. Adatvagyon kataszter létrehozása és feltöltése
4. Központi forrás-adatbázis kialakítása
5. Adatellenőrzési rendszer kialakítása
6. Adatfrissítő rendszer létrehozása
7. Adatkapcsolati háló kialakítása
8. Adatpublikációs megoldások kidolgozása
9. Finomhangolás
10. Információbiztonsági Irányítási Rendszer kialakítása, ISO 27001 tanúsítvány megszerzése

Az adatvagyon-létrehozás viszonylag egyszerű koncepcióra épül: a forrásadatok első sorban a keletkezési helyük által támasztott igényekre reflektálnak, ezért logikailag helyhez kötöttek. Ha az egyes alrendszerekből a forrásadatokat „kiszabadítjuk”, lehetővé válik egy komplex adatkapcsolati háló kialakítása. Ha tehát egy alrendszer a benne keletkező forrásadatokat „beadja a közösbe”, adott esetben ugyanazokat az adatokat már egy teljesen más kontextusban kapja vissza – kiegészülve az erőforrásként megjelenő, nem az adott alrendszerben keletkező adatokkal. A koncepció része továbbá, hogy statisztikai jelleg helyett minden esetben elemi szintű forrásadatokra épüljön az adatvagyon.

A koncepció elfogadtatása egyrészt felsővezetői döntést, majd a döntést követően folyamatos támogatást jelent. Másrészt a forrásadatokat „birtokló” alrendszer-gazdák számára világossá kell tenni az együttműködés előnyeit. Olyan adatvagyon-kezelő rendszert kell kiépíteni, ahol a haszon kézzelfoghatóvá válik számukra.

Az előzőekből egyenesen következik, hogy a forrásadatokat birtokló alrendszerek nem képesek és nem is lehet feladatuk az alrendszer független – nem az adatok létrehozására, hanem szerteágazó felhasználásukra koncentráló – adatvagyon-kezelő rendszer létrehozása és működtetése. Erre a feladatra önálló szervezeti egységet kell létrehozni – lehetőség szerint a KIR adatkezelőjeként működő Oktatási Hivatalon belül. Fontos, hogy tisztán lássuk: a KIRAK (vagyis KIR Adatvagyon-kezelő Központ) nem azért jön létre, hogy az értékes adatokat kivegye az alrendszerek belső szakmai műhelyeinek a kezéből, hanem éppenséggel ennek az ellenkezője igaz: minden elérhető adatot szerezzen meg és tegye ezeket az adatvagyon részévé, ezáltal biztosítva az alrendszerek összes külső adatigényének teljes körű kielégítését garantáltan jó minőségű, széles kapcsolódási lehetőségekkel rendelkező adatokkal.

Az egyébként rendelkezésre álló adatok felhasználhatóságát nagymértékben korlátozza, hogy egyszerűen nem ismerjük az adatokat, és főleg nem a tulajdonságaikat. Az adatvagyon kataszter szerepe éppen az, hogy a köznevelés számára releváns adatokat kata-

lógusserűen nyilvántartásba vegye, és minden érdeklődő számára kereshető formában megismerhetővé tegye.

A központi forrás-adatbázis adatszinkronizációs és -tárolási feladatokat lát el, az adatok közötti kapcsolati háló kiépítése már nem ezen a szinten történik.

Ha adatvagyon-hasznosításról beszélünk, az adatellenőrzési rendszer kialakítása, működtetése és folyamatos fejlesztése kiemelten fontos terület. Minden egyes tárolt adatról kivétel nélkül meg kell tudni mondani, hogy adatminőség szempontjából hogyan kategorizálható. Ugyanis hiú ábránd, egyben alapvető tervezési hiba lenne, ha feltétel nélkül megbíznánk a központi forrás-adatbázisba beérkező adatok minőségében. Az adatvagyon-kezelő rendszernek saját, önállóan működő hibadetektálási és protokollokra épülő hibakezelési rendszerre van szüksége, melynek működése az adatmező szintű értékészlet-problémák vizsgálatától kezdve egészen a kapcsolati háló segítségével elvégezhető globális anomáliák felfedéséig terjed.

Az adatvagyon természetesen nemcsak tökéletes minőségű adatokból áll, lehetőség szerint minden felhasználásra alkalmas adatot be kell gyűjtenünk – pontosan megadva az adatminőségre vonatkozó jellemzőket. Ezáltal az adatellenőrzési rendszer célja kettős: megkísérelni a hiba kijavítását, szükség esetén az adatforrás bevonásával, illetve pontosan megjelölni a minőségi problémával rendelkező adatokat. További fontos követelmény, hogy a felhasználók gyakorlati tapasztalatait is visszacsatoljuk a rendszerbe, ezáltal minél szélesebb körű a felhasználás, annál nagyobb az esélyünk a hibák felfedésére és kijavítására.

Az adatfrissítő rendszer a központi forrás-adatbázisban tárolt adatok naprakészségéért felel.

Az adatkapcsolati háló az adatvagyon központi eleme, amennyiben a tárolt adatok közötti kapcsolatok kiépítéséért és fenntartásáért felelős. (Az adatok közötti kapcsolatok kialakítási lehetőségeivel külön foglalkozunk.)

Az adatpublikációs rendszer célja, hogy a felhasználók számára a nekik megfelelő formátumban és feldolgozottsági szinten rendelkezésre álljon a szükséges adat. Míg egy kutató tipikusan további feldolgozásra alkalmas adatbázis-részletet tud letölteni, addig az intézményi felhasználók önkiszolgáló webes felületen keresztül jutnak adatokhoz, az alrendszerek felé pedig adatinterfészen keresztül történik az adatáramlás.

A finomhangolás lényegében folyamatosan végzett tevékenység, mely a külső és belső körülmények mindenkori változásaihoz igazítja a rendszert.

A KIR 2005. és 2008. között rendelkezett ISO-tanúsítvánnyal. A köznevelési adatvagyon jelentőségét és értékét figyelembe véve célszerű feléleszteni ezeket a nemes hagyományokat.

## Az adatkapcsolati háló létrehozásának lehetséges megoldásai

Az adatkapcsolati háló az adatvagyon „idegrendszere”, mely az adatvagyon alkalmazhatóságát – az adatok minőségén túlmenően – döntően befolyásolja. Az adatkapcsolati háló olyan logikai és fizikai réteg, mely az adatvagyon minden elemét a gyakorlatilag elérhető összes módon összefüggésbe hozza egymással. Az adatkapcsolati háló fontos tulajdonsága, hogy semmilyen módon nem függ össze az adatok keletkezési helyeként szolgáló alrendszerekkel, azoktól függetlenül létezik – többek között ezáltal is biztosítva az adatok „kiszabadítását”.

A különböző alrendszer adatvagyon-felhasználással kapcsolatos igényeit végső soron az adatkapcsolati háló „mozgósításával” lehet kiszolgálni.

A folyamatosan változó felhasználási igények okán az adatkapcsolati háló is dinamikusan fejlődik – összhangban az adatok számának és minőségének növekedésével.

Az adatkapcsolati háló értelmezésünk szerint csak komplex adatvagyon-kezelő rendszerekben létezhet – amiről egyelőre még nem beszélhetünk –, ugyanakkor már most is megfigyelhető adatbázisok közvetlen összekapcsolása különböző köznevelési rendszerekben. Az adatbázisok összekapcsolására háromféle, alapvetően különböző módszert azonosíthatunk, melyek a következők:

- 1 A legkézenfekvőbb eset az országos kompetenciamérés, amikor a személyazonosításra alkalmas adatokat, valamint a Hivatalban keletkező és tárolt személyazonosításra nem alkalmas egyéni szintű adatokat a köznevelési intézmény helyben össze tudja kapcsolni. Ekkor jogosultság alapú adatkapcsolatról beszélhetünk, amennyiben a köznevelési intézmény tanulónként nevesítve megismerheti a mérési eredményeket. A Hivatal ezzel szemben nem jogosult az egyébként nála keletkező egyéni szintű mérési eredmények konkrét személyekhez kötésére, ezért ún. mérési azonosító alkalmazásával, személyazonosításra alkalmatlan módon kezeli az adatokat.
- 2 „A tanulók egészségi állapot iránti közös felelősség megélését nagyban támogatni tudja, ha a szülők hozzájárulnak ahhoz, hogy a testnevelők a diákok fittségi eredményeihez közvetlenül hozzákapcsolhassák a tanulók releváns személyes adatait is.”<sup>7</sup> Miután a személyes adatok között az országos kompetenciamérésben alkalmazott mérési azonosító is megadható, ezért a fittségi adatok és a kompetenciamérés eredményei egyéni szinten, személyazonosításra alkalmas módon összekapcsolhatók. Az adatbázisok összekapcsolása tehát abban az esetben is lehetséges, ha erre az adatok tulajdonosa felhatalmazást ad.
- 3 Az előzőektől eltérő megoldás a KIR és a felsőoktatási információs rendszer, valamint az Országos Nyugdíjfolyósítási Igazgatóság adatbázisainak az összekapcsolása. Ebben az esetben egy harmadik, felhatalmazással rendelkező szervezet (Nemzeti Infokommunikációs Szolgáltató Zrt.) kapcsolati kód segítségével végezte az összekapcsolást, melynek eredményeképpen olyan egyéni szintű adatbázis jött létre, mely nem tartalmaz személyazonosításra alkalmas adatokat, így személyekhez elvileg sem köthető, ugyanakkor a benne szereplő egyéni szintű adatok okán rendkívül jól használhatónak tűnik.

Az adatbázisok összekapcsolása a felsorolt példák alapján többféle módszer szerint is történhet. Az adatkapcsolati háló hasonló funkciókat lát el, de nemcsak két adatbázis között, hanem az adatvagyon egészére vonatkozóan.

## Az adatvagyon kezelését biztosító információs rendszer felépítése

Az adatvagyon-kezelés természetszerűleg kiszolgáló információs rendszer létrehozását és működtetését feltételezi. Az információs rendszer felépítése többféle, egymásra épülő réteggel modellezhető.

<sup>7</sup> Az idézet a NETFIT® rendszer honlapján ([www.netfitweb.hu](http://www.netfitweb.hu)) elérhető tájékoztatóban szerepel, a 7. Miért jó, ha szülőként hozzájárulok gyermekem adatkezeléséhez? témakörnél.

Az adatgyűjtési réteg biztosítja a forrásadatok folyamatos beáramlását a rendszerbe – az adatforrások felépítésétől és földrajzi elhelyezkedésétől, valamint szervezeti hovatartozásától függetlenül. Miután köznevelési adatvagyonról beszélünk, a legfontosabb adatforrásnak az intézményi adminisztrációs rendszereket tekintjük, melyek főtevékenységként az intézmények helyi igényeinek teljes körű kiszolgálására hivatottak.

Fontos megjegyezni, hogy a forrásadatok sok esetben valamilyen tanügyigazgatási rendszer működése során „melléktermékként” keletkeznek, vagyis kifejezetten nem optimalizáltak univerzális adatvagyon-hasznosításra – ettől függetlenül persze még egy tanügyigazgatási rendszer teljes mértékben képes betölteni jogszabályban előírt feladatát. Ha egy ilyen rendszer forrásadataira szükségünk van – márpedig minden adatra szükségünk van –, akkor kiemelt feladatunk az adatminőség kontrollálása és hibajavítási folyamatok beépítése, mely a rendszer következő rétegének tekinthető.

A harmadik réteg a begyűjtött adatvagyon-elemek fizikai tárolására szolgál.

A következő logikai rétegnek az adatkapcsolati hálót tekinthetjük, melyről korábban már volt szó.

Végül a „legfelső” réteg az adat- és információ-publikációs réteg, mely a felhasználók igényeinek teljes körű kiszolgálására hivatott.

Az adatvagyon-kezelést biztosító információs rendszer leírása alapján azt hihetnénk, hogy ugyan a rendszer létrehozása sok munkával jár, viszont a különböző rétegek kialakítása és egymásra építése mégsem túlzottan bonyolult feladat. Ha pedig ez így van, akkor máris kezdjük neki: minden létező adatot szerezzünk be és hozzuk létre a köznevelési adatvagyon.

Ezzel szemben az az igazság, hogy valójában rendkívül összetett problémahalmazzal állunk szemben, különösen, ha figyelembe vesszük a köznevelés korábban is említett specialitásait.

Meggyőződésünk szerint egyetlen járható út létezik: evolúciós szemlélettel kezeljük a köznevelési adatvagyon létrehozását, vagyis nem a nagy egész gyors létrehozására törekszünk, hanem azokat a folyamatokat és tevékenységrendszereket definiáljuk, majd folyamatosan finomítjuk, melyek alkalmazása – „icipici”, de jól definiált kezdő állapotból kiindulva – elvezet a rendszer teljes körű kiépítéséig. Ellenkező esetben rövid időn belül egy minden adatot tartalmazó, de valójában használhatatlan és fenntarthatatlan rendszerrel fogunk szembesülni.

Az evolúciós szemléletű építkezés további előnye lehet, hogy mivel – valljuk be őszintén – jelenleg nem rendelkezünk a pedagógiai célú alkalmazhatóságához szükséges tudással, az adatvagyon kialakításával és folyamatos fejlesztésével párhuzamosan az alkalmazás lehetőségeit és korlátait is megtanulhatjuk. Törvényszerűnek tűnik, hogy az evolúciós folyamat gyakran fog visszalépésekkel járni, amikor kénytelenek vagyunk felülvizsgálni korábban biztosnak hitt elképzeléseinket, mert az eredmények nem igazolják vissza a fejlesztési irányok kiválasztását.

## Az adatvagyon pedagógiai célú felhasználásának jövője

A jövő többféle forgatókönyvet is tartogat számunkra. Nézzük az ideális esetet: a közeljövőben létrejön a köznevelési adatvagyon és ennek a menedzselésére hivatott KIR Adatvagyon-kezelő Központ (KIRAK), mely többek között hatékonyan támogatja az egyéni

adatok pedagógiai célú felhasználását és az ezzel kapcsolatos oktatáskutatói tevékenységeket, de természetesen ezen túlmenően kiszolgálja a tanügyigazgatás teljes adatigényét – a döntés-előkészítéstől kezdve egészen a beválás-vizsgálatokig.

A pedagógia gyermek és tanulóközpontú komplex megközelítést alkalmaz, melynek integráns eleme a pedagógusok adatvagyon-felhasználásra épülő tevékenysége – szakértők és oktatáskutatók által „kívülről” támogatva. Az egyéni szintű idősoros adatok jelentősen megnövelik az egyéni differenciálás lehetőségeit, egyben a pedagógiai tevékenységek hatékonyságát és eredményességét.

\*\*\*

## Hogyan foglalhatók össze a tanulmányban leírtak?

A Köznevelés az adatok bűvöletében elfogult írás.

Az adatfelhasználás pártján áll.

Hogy gyűjtjük, feldolgozzuk, ellenőrizzük, tároljuk, értelmezzük és alkalmazzuk őket.

Tudatosan, tervszerűen, széleskörűen, a korlátokra és kockázatokra figyelve.

### IRODALOM

HALÁSZ G. (2010): *Beszámoló az OECD oktatási információs rendszerekről tartott munkaértekezletéről, 2010. október 14-15, New York. Kézirat*

LANNERT J. (2004): *Hatékonyság, eredményesség és méltányosság. Új Pedagógiai Szemle, 54. évf. 12. sz., 2004. december.*

HALÁSZ G. (2007): *Tényekre alapozott oktatáspolitikai. Háttér tanulmány az Oktatási Kerekasztal munkájához. Kézirat. [http://halaszg.ofi.hu/download/Evidence\\_based\\_study.pdf](http://halaszg.ofi.hu/download/Evidence_based_study.pdf)*

BARABÁSI A. L. (2013): *Behálózva – a hálózatok új tudománya. Helikon Kiadó Kft., Budapest.*