

# A közösségi közlekedési rendszert érintő működtetési és üzemeltetési döntések támogatása az elektronikus jegyrendszerhez kapcsolódó közlekedési adattárház kialakításával

A hazai közösségi közlekedés megújuló koncepciójához és stratégiájához illeszkedő korszerű, interoperábilis közszolgáltatási rendszer kialakításának folyamatában létrejövő Nemzeti Elektronikus Jegyrendszer Platform (NEJP) a fokozatosan belépő közlekedési szolgáltatóknál (Volánok, MÁV, BKK) olyan mennyiségű valós idejű adatot keletkeztet, hogy ezen adattömeg hasznosítása elkerülhetetlenné teszi egy Nemzeti Közlekedési Adattárház kialakítását.

**Dr. Sárközi György\*** — **Roósz Tamás\*\*** — **Pusztai Gábor\*\*\***

vezető tanácsadó

innovációs irodavezető

üzletfejlesztési igazgató

\*e-mail: sarkozi.gyorgy@nemzetimobilfizetes.hu

\*\*e-mail: roosz.tamas@nemzetimobilfizetes.hu

\*\*\*e-mail: pusztai.gabor@nemzetimobilfizetes.hu

## 1. AZ ADATOK SZEREPE A KOR-SZERŰ KÖZÖSSÉGI KÖZLEKEDÉSI RENDSZEREKBE

Ma már szinte nincs olyan szakmai, tudományos rendezvény vagy publikáció, amely ne foglalkozna az adatok hasznosításának jelentőségével, ennek módszereivel és problémáival. A nagy tömegben keletkező adatok feldolgozására már jóval azelőtt kell a megoldásokat kigondolni, mielőtt azok keletkeznek. Fel kell készülni ezek tárolására, kezelésére, a belőlük származtatható információk körének felmérésére és a hasznosítás módjának kialakítására. A jelen cikk – támaszkodva és hivatkozva a

tárgykörben már megjelent forrásmunkákra – átfogó betekintést ad arra vonatkozóan, hogy az elektronikus jegyrendszer evolúciója során milyen kérdéseket kell kezelni.

A Közlekedéstudományi Szemle korábbi lapszámaiban több szerző is foglalkozott a témakör egy-egy kiemelt területével: az adathalmazokon végzett ad-hoc analitikák elvégzésére alkalmas megoldások vizsgálatával és fejlesztésével [1], a közlekedési adatok hasznosítását lehetővé tevő információtechnológiák megújításával [2].

Az elektronikus útdíjrendszer adatainak újrahasznosíthatóságának vizsgálata [3] és az ada-

tok fúziójára alapozó technológia bemutatása is [4] ráirányította a figyelmet e területen a hazai közlekedéstudományi munka eredményeire.

Az utóbbi időszakban szervezett közlekedéstudományi konferenciák előadásai között is kiemelt figyelem kísérte az adatvezérelt, igényvezérelt közlekedési közszolgáltatások témakörét.

A 2016. évi Nemzeti Közlekedési Napok adatokkal foglalkozó szekciójában bemutatott prezentáció [5] esettanulmány formájában már betekintést engedett a mobilfizetési rendszer egy adott időszakában keletkezett, az SAP HANA adatbázisában tárolt, anonimizált adatainak üzleti intelligencia célú hasznosítási lehetőségébe. A budapesti parkolási zónák igénybevételére vonatkozó Big Data ad-hoc analitika értékes összefüggésekre mutatott rá, amelyek a későbbiekben jól hasznosíthatók az utazási szokások megismerésében és a parkolási rendszer további fejlesztése során.

Az ad-hoc elemzések elvégzéséhez megfelelően strukturált adatbázisokra, adattárházra és modellekre (logikai és fizikai) van szükség, így ezen a helyen szükséges rámutatni a gyakran nem megfelelően használt fogalmakra, azok lényegi különbségére.

Az adattárházak fogalmának tisztázásakor talán legtöbbször Bill Inmon [6] és Ralph Kimball [7] definícióját idézik. A gyakran az adattárházak atyjaként is emlegetett Bill Inmon az adattárházat technológiák gyűjteményének definiálja: az adattárház az adatok egy témaorientált, integrált, tartós és időfüggetlen, a vezetői döntések támogatására létrehozott kollekcója.

Kimball szerint az adattárház egy adott szervezet azon adatgyűjtő és -szolgáltató részeit foglalja magában, ahol a működési adatokat újrastrukturálják riportkészítési, jó teljesítményű és egyszerűen kezelhető elemzésekhez.

Az adatbázis és az adattárház fogalmának helyes értelmezésére Halassy Béla definíciói is

jól alkalmazhatók [8]. Az adatbázis az adatok olyan rendszere, amelynek felhasználatán több felhasználó osztozik, és az ismeretek egymással alkotott természetes összefüggéseik szerint szervezettek, de függetlenek a feldolgozó programoktól. Az adattárház az a speciális adatbázis, amely az adatokat lekérdezési, elemzési műveletekre optimalizált szerkezetben tárolja; a kiszolgált vezetési szintek igényeinek megfelelően aggregált adatokat is tartalmaz; valamint különböző forrásokból nem tranzakciónként, hanem adott periódusonként és az adatértekek történetiségének megőrzésével frissül.

1. ábra: : Parkolási zónák közötti mozgások elemzése a mobilfizetési tranzakciók alapján [5]



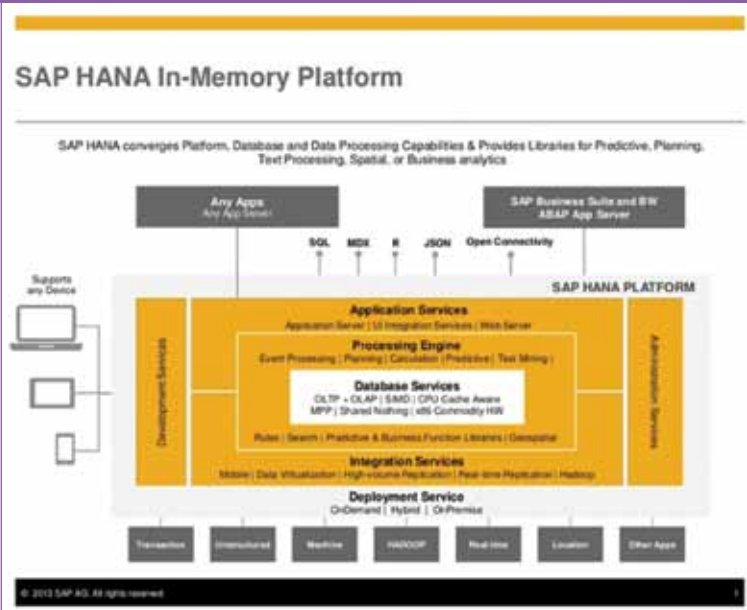
A Nemzeti Mobilfizetési Zrt. az általa működtetett mobilfizetési rendszerben keletkező közterületi parkolóhoz kapcsolódó tranzakciók adatainak hasznosítására elindított pilot projektben az SAP HANA platformot alkalmazta.

## A pilottal szemben a következők az elvárások [1]:

- a Nemzeti Mobilfizetési Rendszerben (NMFR) használatos vezetői statisztikák készítése,
- nagy mennyiségű adathalmazon végzett ad-hoc analitikák készítése,
- adathalmaz vizsgálat alapján korrelációk, összefüggések feltárása, prediktív modellezési lehetőségekre javaslattevés és ezek bemutatása,
- valós idejű feldolgozás lehetőségeinek vizsgálata,
- éles NMFR adatok az SAP HANA rendszerben történő feldolgozása.

Az adatok platformban való hasznosíthatóságának komplexitását és a rendszerben rejlő

2. ábra: Az SAP HANA In-Memory Platform felépítése (forrás: SAP)



lehetőségeket jól mutatja, hogy a pilot során összesen 27 mutatószám és 43 különféle dimenzió került meghatározásra (1. és 2. táblázat). Ezek a mutatószámok és dimenziók a későbbiekben a pilot tapasztalatai alapján az adattárház céljainak megfelelően természetesen módosíthatók, bővíthetők.

Az SAP Lumirában történő riportkészítésre, megjelenítésre a létrehozott és publikált di-

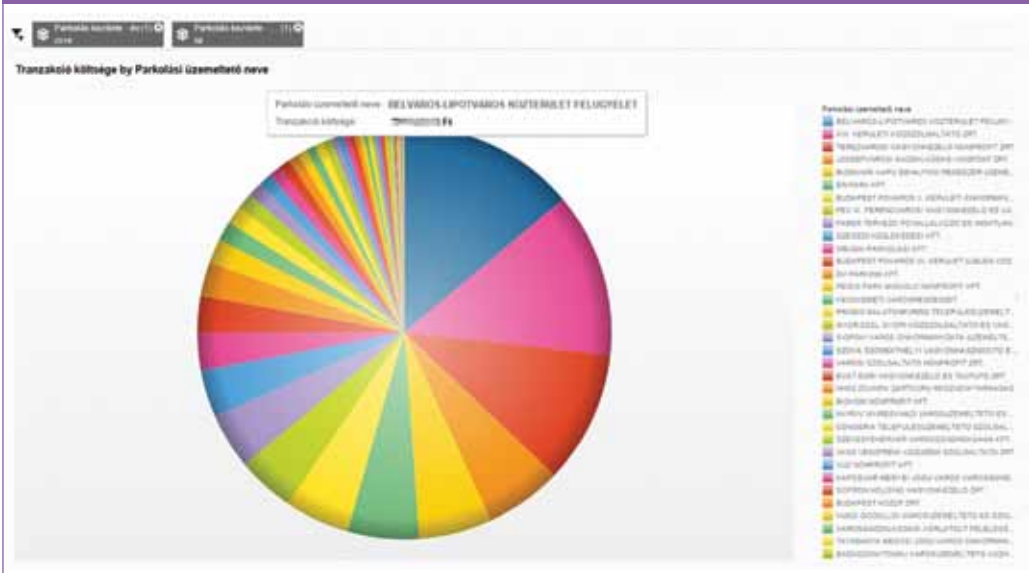
1. táblázat: A definiált 27 db mutatószám az NMFR rendszer adathalmazán

Mutatószám (27 db)	Típus
14	Átlag
15	Átlag
18-21	
<ul style="list-style-type: none"> <li>tranzakció száma</li> <li>tranzakció összege (fizetett díj)</li> <li>díjfizetés módja (szájtól egyéni/vízszelaladó)                             <ul style="list-style-type: none"> <li>esetlélaladó</li> <li>SMS</li> <li>bankkártya</li> <li>web</li> <li>vízszelaladó</li> </ul> </li> </ul>	

2. táblázat: A definiált 43 db dimenzió az NMFR rendszer adathalmazán

Dimenziók (43 db)	Típus
1-10	időpont
11-15	időpont
16-21	időpont
22-43	időpont

3. ábra: Parkolási közszolgáltatók bevételei - kördiagram



menziók, mutatószámok mentén van lehetőség. A riportkészítések alapja lehet a tranzakciószám és költség adat, amelyek a különböző dimenziók szerinti bontásban megjeleníthetők, elemezhetők.

A mutatószámok és dimenziók összekapcsolt lekérdezéseket és vizualizációt tesznek lehetővé az SAP Lumira moduljának alkalmazása révén. A következő ábrákon megismerhetők az eszköz képességei.

A 3. ábra szemléletesen mutatja egy tetszőlegesen kiválasztott időszakra a parkolási közszolgáltatók részesedését a mobilparkolásban. (Az összes adatvédelmi szempontok miatt kitaraktuk.)

A 4. és 5. ábra nem csupán a parkolás indítások időbeli eloszlását mutatja, hanem ezen mutatókon keresztül az NMFR terhelését is.

4. ábra: Tranzakciók száma a hét napjai szerint - oszlopdiagram



A HANA az adatbázisába áttöltött, az NMFR működése során keletkező adatok hasznosításával további lehetőséget kínál arra is, hogy a korábbi pénzügyi teljesítések idősorai (szezonális összefüggések), valamint egyéb paraméterek, adatok (pl. város nagysága, motorizáció mértéke) elemzése és vizsgálata alapján az üzleti tervezéshez megbízható alapot nyújtson, ezen felül a rendszerrel kapcsolatos jogszabály-módosítási elképzelések, szcenáriók hatásainak modellezéséhez is technikai háttérrel adhat.

5. ábra: : Tranzakciók száma 08-20 óra között félórás bontásban - oszlopdiagramok



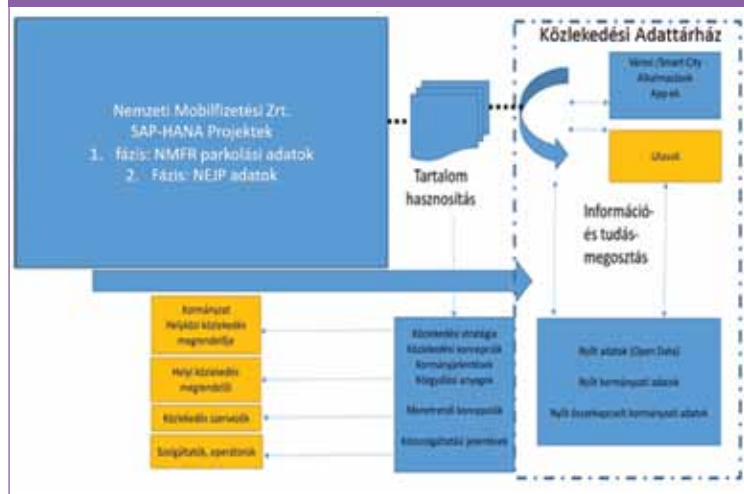
## 2. A NEJP ÉS NMFR ADATHALMAZAINAK HASZNOSÍTÁSI LEHETŐSÉGEI EGY KÖZLEKEDÉSI ADATTÁRHÁZ KI-ALAKÍTÁSÁVAL

Az SAP HANA pilot eddigi tapasztalatai alapján a NEJP valós idejű adatainak hasznosítá-

sa egy közlekedési adattárház kialakításával biztosítható a legegyszerűbb módon, az ehhez vezető folyamatot a 6. ábra mutatja.

Az adattárházban előkészített és vizualizált adatok tartalomhasznosítása a közösségi közlekedés releváns szereplői számára lehetővé

6. ábra: Közlekedési adattárház kialakításának folyamata



teszi, hogy hazánkban elkezdődjön a digitális átállás a közlekedési szektorban, és létrejön a közlekedési szolgáltatók összehangolt tevékenységén alapuló együttműködő, intelligens közlekedési rendszer (C-ITS). A közlekedési adattárház egyaránt kiszolgálja adatokkal a közlekedési szolgáltatások megrendelőit (közlekedési kormányzat és önkormányzatok), az általuk megbízott közlekedés szervezőket, a közlekedési szolgáltatásokat nyújtó közszolgáltató-

kat, alvállalkozóikat és a közösségi közlekedési szolgáltatások igénybe vevőit, azaz az utasokat. Az adattárházhoz kapcsolódhatnak olyan, például felhőalapú szolgáltatások, amelyek a megrendelők és közlekedésszervezők számára stratégiai döntéseket, koncepciókat alapozhatnak meg, támogatják a közlekedési alágazatok együttműködését, a menetrendek összehangolását, a vizualizált adatokkal könnyen áttekinthetővé teszik a közszolgáltatásról szóló jelentéseket, beszámolókat. Az adatok meghatározott köre bevonható a nyílt és/vagy nyílt-összekapcsolt kormányzati adatok körébe, felhasználható a smart városi szolgáltatások kialakításához, korszerű utastájékoztatáshoz, utazástervezéshez. Az információk és a tudás megosztása révén szolgálhatja a digitális jólét program megvalósulását a közösségi közlekedés területén. Az adattárház nyújtotta komplex adatszolgáltatás elengedhetetlen az olyan korszerű közlekedési megoldások hatékony bevezetéséhez is, mint például az igényvezérelt közlekedés (DRT-rendszerek) és mobilitási szolgáltatás (MaaS – Mobility as a Service). Az összekapcsolható adatok körét növelheti például a városkártyák felhasználásával kapcsolatos információk bevonása a közlekedési adattárházba, amely a közlekedéshez közvetlenül vagy közvetve kapcsolódó tevékenységek megismerését is segítheti.

A közlekedési adattárház révén a szakadástmentes egyéni utazástervezés és optimális mobilitási megoldás kiválasztása kiegészülhet más, az adattárházban hasznosuló nyílt, összekapcsolt adat hasznosításával, azaz komplex szolgáltatási lánc válik tervezhetővé, lefoglalhatóvá (helyfoglalás, szolgáltatásfoglalás, parkolás, jegyigénylés) és korszerű mobilfizetési megoldással megvásárolhatóvá. A közhiteles adatok összekapcsolása révén az állam által nyújtott kedvezmények jogszerű igénybevétele és a közszolgáltató közlekedési rendszer fenntarthatóságára gyakorolt hatása is ellenőrizhető.

A Nemzeti Mobilfizetési Rendszer és a Nemzeti Elektronikus Jegyrendszer Platform fejlesztésének is kiemelt stratégiai célja a fenti példákhoz hasonló olyan lehetőségek kiaknázása, amelyek támogatják a technológiai és közpolitikai innovációt, az együttműködést, a pro-

duktivitást, a méretgazdaságosságot és a folyamathatékonyságot. Ennek részeként kiemelt szempont az operáció során keletkező adatvagyon felhasználásának támogatása az éles és pilot működésből származó anonimizált adatok innovatív módon, közpolitikai célból történő alkalmazási lehetőségeinek minél szélesebb körű feltárása és kiaknázása. A szélesebb körű feltárás és kiaknázás operatív megvalósítása történhet például a piaci szereplők számára szabad hozzáférésű adatok biztosításával vagy a belső elemzési kapacitás fejlesztésével (pl. szolgáltatók dinamikus árazásának vagy gördülő tervezésének támogatására).

Az MIT Sloan Management Review és az IBM közös New Intelligent Enterprise Global Executive kutatása [9] is kiemeli, hogy a fejlett adatelemzés versenyképességi kulcs tényező. A szervezetek többségénél – a közhiedelemmel ellentétben – azonban nem az adatgyűjtés módja jelenti az adatelemzés széles körű elterjedésének legnagyobb akadályát, hanem (1) azon ismeretek hiánya, hogy miként lehet az adatelemzést használni az üzleti működés támogatására, vagy mert (2) más menedzsment prioritások miatt háttérbe szorul.

Az NMFR és a NEJP adathalmazainak hazai hasznosítása abban az esetben válhat nemzetközi szinten is példaértékűvé, amennyiben az adatelemzés a múltbeli adatok vizsgálata mellett mielőbb alkalmassá válik (és alkalmazásra kerül) a jövőben szükséges stratégiai és operatív lépések megalapozására. Ennek érdekében az adatgyűjtés, előkészítés, elemzés-hasznosítás terén olyan erős gyakorlatot szükséges kiépíteni, amely az MIT kutatásában résztvevő élenjáró társaságokat jellemzi, azaz az adatelemzés napi alkalmazása köré szükséges építeni az operatív és a hosszú távú döntéselőkészítési folyamatokat.

A közlekedési adattárház hasznosításának egyik legnagyobb kihívása, hogy valódi üzleti és/vagy közpolitikai értékteremtést biztosítson, ezért az adatelemzés hasznosításának kiemelten eredményorientáltnak kell lennie. Az MIT kutatói ezt az adat-elemzés-cselekvés hármasának folyamatos körforgásával látják biztosíthatónak, amely minden esetben az üz-

leti hasznosulást tartja szem előtt. Ennek értelmében, az adatelemzési modellek, folyamatok és alapadatok az üzleti igények szerint kerülnek elfogadásra, elvetésre vagy átdolgozásra, és az elemzések eredményei folyamatosan egyre bővülő mértékben épülnek be a mindennapi folyamatokba. A 6. ábrában szemléltetett „tartalom hasznosulás” tehát nem egy statikus állapot és inkrementálisan módosuló eredménytermékek (pl.: közszolgáltatási jelentések, menetrendi koncepciók, közlekedésforgalmi kimutatások) összessége, hanem egy elemzésorientáción alapuló újszerű állami vállalati működést jelent.

Szintén fontos eredménye az MIT kutatásnak, hogy az elemzésorientált vállalati menedzsment által alkalmazott eszköztár folyamatosan átalakuláson megy keresztül, amelyet az adatelemzések üzleti hasznosításának igénye indukál. Így folyamatosan háttérbe szorulnak a sztenderd adatszolgáltatások és azok vizualizációjára épülő megoldások, és előtérbe kerülnek az üzleti szimulációk, szcenárió-képzések, az üzleti folyamatokba közvetlenül beépülő elemzések. A vizualizáció mellett tehát kiemelt szereppel bír a magasabb szintű statisztikai módszereken (pl. regresszió elemzés) alapuló döntési alternatívák előkészítése.

Kiemelt szempontként kell kezelni, hogy a közlekedési adattárházat tényleges kihívások mentén, fókuszáltan alakítsák ki, azaz olyan adatok elemzésére koncentráljanak, amelyek hasznosítása a legnagyobb lehetőséget tartogatja az ágazat szereplői számára. Ennek megalapozására Lavalée és szerzőtársai „Proof-of-value” típusú pilot projekteket [9] javasolnak, amelyek becsült biztosítanak a potenciális megtakarítás vagy additionális értékkeremtés tekintetében. A lehetséges opcióknak pedig illeszkedniük kell a vállalati/közpolitikai stratégiához, és az aktuális képességekhez olyan célokat kell választani, amelyekre reálisan választ adhatnak az ismert adatelemzési módszerek.

Hiába fejlődtek azonban az általános adatelemzési képességek (elérhető adatok mennyisége, adattárolási és számítási teljesítmény, adatelemzési módszerek és eszközök) exponenciálisan, az előzőekben részletezett szem-

pontok egyeztetése és az optimális megoldások megtalálása nemzetközi szinten is óriási kihívást jelent. A McKinsey kutatása [10] szerint a legtöbb vállalat az adatgyűjtés-elemzés potenciális értékének csak töredékét realizálja, és az európai közszektorban, amely a közlekedési adattárház szempontjából releváns összevetés, a lehető legrosszabb az eredmény, mindösszesen a potenciális érték 10-20%-a teljesül.

A McKinsey szakértői – amellet, hogy az MIT kutatáshoz hasonlóan szintén jelentős korlátként azonosítják, hogy sok vállalat csak nehezen integrálja napi operatív folyamataiba az adatelemzéseket – kiemelik, hogy az adatelemzés egyben a versenypiaci feltételeket is jelentősen átszabja. Az élenjáró vállalatok nem csak operációjukat építik fejlett adatelemzésre, hanem teljes üzleti modelleket, digitális platformokat építenek azokra. Ezeket a platform alapú piacokon pedig a piacvezető mellett a további szereplők csak korlátozottan érvényesülhetnek. A McKinsey tanulmánya arra hívja fel a figyelmet, hogy bár az adatok egyre inkább elérhetőek, azonban az értékteremtés azon adatgazdánál csapódik le nagyobb mértékben, amelyek ritka adatokkal rendelkeznek és/vagy azokat újszerűen aggregálják és/vagy azokból értékes adatelemzéseket készítenek.

Az előzőek gyakorlati átültetése érdekében a közlekedési adattárház kialakításáért és működtetésért felelős társaságnak birtokában kell lennie annak a tudásnak, hogy miként lehet a fejlett adatelemzési technikákat az értékteremtést szolgáló döntések megalapozásához alkalmazni. A Boston Consulting Group (BCG) szakértői [11] is rámutatnak, sok vállalat azért nem tudja megfelelően kiaknázni az adatelemzési technikákat, mert már a lehetőségek feltérképezése is kihívást jelent számukra. A BCG szakértői által alkalmazott adatelemzési kategóriák a múltbeli adatok elemzésétől, a modellezésen keresztül az optimalizálásig terjednek. A közlekedési adattárháznak képesnek kell lennie minden ezen felhasználási céloknak megfelelni, azaz az információ/tudás megosztásán túlmenően olyan hatékony modellezést kell támogatnia, amely lehetővé teszi az üzleti környezetet befolyásoló változók hatásának vizsgálatát

(pl. mobilparkolásnál a kényelmi díjnak az alkalmazott parkolási óradíjtól függő, vagy a parkolási óradíjak napszaktól, forgalomtól függő eltérésének bevételekre gyakorolt hatása). Fontos az is, hogy a modellezési tapasztalatok alapján tényleges optimalizálási megoldások szülessenek.

A BCG szakértői is rámutatnak azonban arra, hogy nem elegendő matematikai algoritmusokkal alátámasztott döntési opciókat előállítani (pl. szolgáltatási lánc tervezésére vagy annak dinamikus árazására), hanem alkalmazásukat tudatosan és következetesen kell kiterjeszteni, ami szükségessé teheti az adattárház kialakító és felhasználó állami tulajdonú társaságok részére új típusú, a fejlett adatelemzési folyamatok alkalmazásához kapcsolódó teljesítményindikátorok (KPI-k) kidolgozását is.

Az OECD 2015. évi jelentése [12] alapján a közlekedésben keletkező és felhalmozódó adatmennyiség alapvetően átalakítja az ágazatot. Az alacsony költségű adatgyűjtés és -feldolgozás, valamint az új elemzési algoritmusok a valóság jobb megértését fogják biztosítani, és a jelenlegi adatforrásokat idővel nem kiegészítik, hanem leváltják. Ezen adatok értelmezéséhez azonban az adatgazdáknál is ki kell építeni a szükséges kompetenciákat, azaz mind forrás-, mind felhasználói oldalon bővíteni kell a humán erőforrás „adatműveltségét”, amivel lehetővé válik, hogy az adatelemzés valós problémákra keressen reális opciókat és ebben az állami és piaci szereplők innovatív adatmegosztási együttműködésekkel alakítsanak ki és túlmutassanak a klasszikus szolgáltató-felhasználó típusú viszonyrendszeren.

Mindez miért nem kerülhető meg egy közlekedési adattárház kialakítása kapcsán?

A McKinsey szakértői szerint a közlekedési ágazatban a „hyperscale” digitális platformok nyújtanak megoldást a kereslet-kínálat gyakori fluktuációjára, a szolgáltatói kapacitások megfelelő kiaknázására és a gyenge jelzési mechanizmusok ellensúlyozására. Amennyiben az állami oldalon ezen kihívásokra nem sikerül hatékony válaszokat adni, az ágazat vagy tartósan eselik jelentős növekedési potenciáltól,

vagy helyét piaci szolgáltatók platform megoldásai fogják biztosítani. Utóbbi esetben az Uberhez hasonló technológia cégek disruptive megoldásait kellene rendszerszinten integrálni, amely még a fentieknél is nagyobb kihívást jelent, tekintettel arra, hogy ez hasonló rendszerszintű kompetenciákat igényel az ágazat szereplőitől, azaz minden lehetséges szcenárió közös jellemzője, hogy a jelenlegi adatelemzési folyamatok és képességek fejlesztése nem megkerülhető.

### 3. A NEMZETI ELEKTRONIKUS JEGYRENDSZER PLATFORM MŰKÖDÉSE SORÁN KELETKEZŐ KÖZLEKEDÉSI ADATHALMAZOK FAJTÁI, JELLEMZŐI ÉS KEZELÉSÜK

A pilot projekt tapasztalatainak felhasználásával a Nemzeti Elektronikus Jegyrendszer Platform (NEJP), sőt ezen túlmenően az országosan egységes elektronikus jegyrendszer üzemszerű működését megelőzően definiálni szükséges az adatstruktúrákat, a mutatószámokat és dimenziókat.

Áttekintve a kialakuló adathalmazokat, a mai Volán járműállománnyal kalkulálva, a szolgáltatások megvalósulásával és a flottakövetési feladatokkal közvetlen összefüggő adatok éves szinten nagyságrendileg mintegy 5-10 TByte nyers (7000 járművet és nagyságrendileg 150 MByte járművenkénti havi adattömeget feltételezve), mérésből származó adatot eredményeznek. Az adatábrázolás megvalósításától függően, ezen adatcsoportok természetesen kiegészülnek a díjtermékek megvásárlásának és felhasználásának – évi többszázmillió – tranzakciós adataival, amelyek volumene fokozatosan emelkedik az elektronikus utasmédiák elterjedésével párhuzamosan.

A közlekedési adattárház a különböző jellegű – például hálózati, menetrendi, járműkövetési és a jegyrendszeri tranzakciós, kedvezményfelhasználási – adatok összevezetése által építhető fel, kiszolgálva a komplex adatelemzési és adatbányászati célokat.

A NEJP projekt megvalósítási szakaszában a megépülő NEJP-központ és a szolgáltatóknál

## 3. táblázat: Tervezett mutatószámok a NEJP rendszer adathalmazán

	Mutatószámok	Típus
1	Utasszám (felszálló, leszálló, tovább utazó, várakozó, lemaradó)	összeg
2	férőhely km, utas km	összeg
3	Menetek (járatok) száma	összeg
4	díjbevételek értéke	összeg
5	Értékesített ill. felhasznált díjtermékek mennyisége	összeg
6	Utazási távolság, díjkezesi távolság	átlag
7	SLA értékek - pontos, késett, siető járatok száma - normál, zsúfolt, utashiányos járatok száma - incidensek száma (human, műszaki, forgalmi) - utaspanaszok száma	összeg
8	Kirótt ill. beszedett pótdíjak értéke és darabszáma	összeg
9	Indokolt költségek megtérítése	összeg

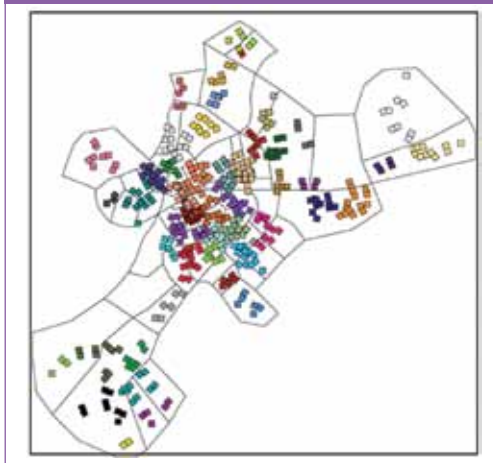
(Volánbusz és DKV) kiépülő pilot projektek, a teljes Volán szolgáltatói portfólióra történő kiterjesztése és az országos szintű integráció (MÁV, BKK) folyamatos feladatot jelent a közösségi közlekedési adattárhálózattal szembeni elvárásrendszer megfogalmazásában és fejlesztésében mind az adatok előállítás, kezelése, előkészítése és tárolása, mind az üzleti intelligencia által nyújtható eredmények szempontjából.

A 3. és 4. táblázatok az adatstruktúrák mutatószámain és dimenzióin keresztül szemléltetik a rendszerben rejlő potenciális lehetőségeket.

Az előzőekben bemutatott lehetőségeket figyelembe véve nem meglepő, hogy a közlekedési szektorban működő tudományos kutatók és nem utolsósorban az üzleti hasznosítók – pl. diszruptív szolgáltatások fejlesztői – már türelmetlenül várják egy komplex közlekedési adatszolgáltatás létrejöttét és a „big data” adatokhoz való hozzáférés lehetőségét, mivel ezen adatokat felhasználva a korábbi módszereiket és szolgáltatásaikat továbbfejleszthetik, illetve új megoldásokat és a jelenleginél lényegesen komplexebb szolgáltatásokat fejleszthetnek.

Egy ilyen rendszerből származó információk felhasználhatóságát példázza a győri Széchenyi István Egyetem korábban már több konferencián és szakkonferencián is publikált kutatása [13], amely rámutat a nagy mennyiségű és jó minőségű adat rendelkezésre állásának szükségességére.

7. ábra: Megállók klaszterei tulajdonságok és távolság alapján



4. táblázat: Tervezett dimenziók a NEJP rendszer adathalmazán

	<b>Dimenziók</b>	<b>Típus</b>
1	Időpontok (hierarchikus dimenzió) év; negyedév; hónap; hét; nap; óra; félóra; negyedóra; perc;	időpont
2	Időszakok (pl. reggeli csúcsidő, csúcsidőn kívül, esti csúcsidő)	időszak
3	Megálló csoport (hierarchikus dimenzió) megálló, megálló terület, zóna, település, közigazgatási terület, elővárosi térség, járás, megye, közlekedési régió, ország	terület
4	Vonali csoportok (hierarchikus dimenzió) vonalcsoportok, vonal, járat	terület
5	Viszonylat	terület
6	Utazási távolság, díjővezeti távolság	távolság
7	Közlekedési mód (vasút, közút, vízi)	felsorolás
8	Közlekedési eszköz (hierarchikus dimenzió) jármű típus, jármű altípus, jármű	felsorolás
9	Szolgáltatási kategória (helyi, elővárosi, regionális, országos)	felsorolás
10	Ellátásért felelős, Közlekedés szervező	felsorolás
11	Közlekedési szolgáltató (hierarchikus dimenzió) Vállalat, üzemegység, gépjárművezető	felsorolás
12	kedvezmény kategória	felsorolás
13	Komfort fokozat	felsorolás
14	Járatok fajtája (menetrendszerinti, másodrész, szerződéses)	felsorolás
15	Díjtermék fő típusok (pl. jegy, bérlet, idő alapú jegy)	felsorolás
16	Díjtermék fajták	felsorolás
17	Értékesítés csatornák (fedélzeti, jegypénztári, online)	felsorolás
18	Alkalmazott utasmédiák (papír, elektronikus kártya, mobil applikáció, stb.)	felsorolás
19	Értékesítő szervezet (hierarchikus dimenzió)	felsorolás

A közforgalmú közlekedésben generálódó adatok elemzését és felhasználását bemutató fenti esettanulmányban a Ward-módszert alkalmazták a megállók klaszterezésére. A pontosabb eredményekhez és a módszer verifikálásához azonban a kutatóknak nagyobb adathalmazra van szüksége, amely ideális esetben alapulhat a közlekedési adattárházon keresztül elérhető, elektronikus jegyrendszerből származó adatokon. Az ilyen adatokhoz a hozzáférés a közlekedési adattárház szolgáltatásaként alakítható ki.

A leírtak tehát arra irányítják rá a közlekedési és közlekedésinformatikai szakma figyelmét, hogy bár számos innovatív és tudományos megoldás vált ismertté a hazai szakemberek ez irányú tevékenysége révén, ideje ezeket integrálni és a közösségi közlekedés javítására rendszerbe állítani.

Reményeink szerint a jelen szakmai összefoglaló is elősegíti az adatvezérelt intelligens megoldások térnyerését a nemzeti közlekedési rendszerben.

## FELHASZNÁLT IRODALOM

- [1] VERES Mihály-SÁRKÓZI György Tibor: A Nemzeti Mobilfizetési Rendszerben (NMFR) keletkező nagy mennyiségű adathalmazon végzett ad-hoc analitikák vizsgálata és fejlesztése, Közlekedéstudományi Szemle LXV évfolyam 5. szám, 2016. október
- [2] SZÚCS Lajos: A közlekedési adatok és az alkalmazott információtechnológiák megújulása, Közlekedéstudományi Szemle LXV évfolyam 5. szám, 2016. október
- [3] CSEREPES Tamás-SZÚCS Lajos: A HUGO megtett úttal arányos elektronikus útdíjrendszer adatainak újrahasznosítása, Közlekedéstudományi Szemle LXVI évfolyam 3. szám, 2016 június
- [4] HORVÁTH Márton Tamás – TETTAMANTI Tamás – VARGA István: Közúti eljutási idő becslésének lehetősége adatfúziós technikával városi úthálózaton
- [5] ROÓSZ Tamás: Országos egységes elektronikus jegyrendszer -NEJP, előadás a Nemzeti Közlekedési Napok konferencián, Debrecen, 2016.
- [6] INMON, Bill: Building the Data Warehouse, Wiley, (1992), ISBN 0-471-56960-7
- [7] KIMBAL, Ralph: The Data Warehouse Toolkit, Wiley, (2011), ISBN 9780470149775
- [8] HAVASSY Béla: Az adatbázis tervezés alapjai és titkai, (1994)
- [9] LAVALLEE, Steve et al: Big Data, Analytics and the Path From Insights to Value. 2011. MIT Sloan Management Review Vol. 52 No. 2
- [10] HENKE, Nicolaus et al: The Age of Analytics: Competing in a Data-Driven World. 2016. McKinsey & Company
- [11] SRIVASTAVA, Ravi et al: Using Advanced Analytics to Improve Operational Decisions. 2016. BCG Perspectives.
- [12] OECD: Big Data and Transport: Understanding and assessing options. 2015.
- [13] NAGY Viktor - HORVÁTH Balázs: Bigdata a közforgalmú közlekedésben a megállóhelyi utasforgalmak példáján. Széchenyi István Egyetem, Győr, 2015.



### The support of management and operation decisions affecting the transport system, with the development of a traffic data warehouse connected to an electronic ticketing system

The National Electronic Ticketing Platform (NEJP) has been established in the framework of a modern, interoperable public service system that aligns with the renewable concept and strategy of domestic public transport. NJEP creates such an amount of real-time data with the gradually entering traffic service providers (Volán bus services, MÁV, BKK), that the utilisation of this mass of data makes the establishment of a National Transport Data Warehouse inevitable.

This paper attempts to remind transport and traffic information professionals that although a great number of innovative and scientific solutions have become known due to the efforts of domestic experts in this field, it is time to integrate them and synthesize them in order to improve public transport.

We hope that this technical summary also promotes the spread of data-driven, intelligent solutions for in the national transport system.



### Die Unterstützung von den Management-und Betriebsentscheidungen für den öffentlichen Verkehr mit der Entwicklung einer Verkehrsdatenbank, die mit einem elektronischen Fahrkartensystem verbunden ist

Der Nationale Elektronische Fahrkartenplattform (NEJP) wurde im Rahmen eines modernen, interoperablen öffentlich-rechtlichen Systems eingerichtet, das sich an das erneuerbare Konzept und an die Strategie des öffentlichen Verkehrs im Ungarn anpasst. NJEP erstellt mit den schrittweise eintretenden Verkehrsdienstleistungs-Unternehmen (Volán-Betriebe, MÁV, BKK) eine solche Menge von Echtzeitdaten, dass die Nutzung dieser Datenmasse die Einrichtung einer nationalen Transportdatenbank unvermeidlich macht.

Auf Grund der hier beschriebenen, es wird für die Fachleute im Verkehrswesen und in der Verkehrsinformatik betont, dass obwohl eine große Anzahl von innovativen und wissenschaftlichen Lösungen aufgrund der Bemühungen der einheimischen Experten in diesem Bereich bekannt geworden sind, es ist höchste Zeit, sie zu integrieren und synthetisieren und für die Verbesserung des öffentlichen Verkehrs in System zu stellen.

Wir hoffen, dass diese technische Zusammenfassung auch die Verbreitung von den datengetriebenen, intelligenten Lösungen im nationalen Verkehrssystem fördert.