

**E-Traffic közlekedési használati előrejelző modell megalkotás**

**Az adatforrások csatolási felületének implementálása**

Készült az Hírközlési és Informatikai Tudományos Egyesület számára

©2013 TFSZ. Minden jog fenntartva

**Tartalomjegyzék**

Bevezetés 3

Adatbetöltés általános leírása 3

Adatbetöltések ütemezése 4

Transzformáció 4

Job 4

Állandó, konstans adatok 4

Területi információk beolvasása 5

Beolvasás lépéseinek leírása 5

Strukturális felépítés 7

KSH által közzétett település lista 7

Koordináta adatbázis 7

Hazai úthálózat adatbázisa 8

Beolvasás lépéseinek leírása 8

Strukturális felépítés 11

Pont állományok 11

Úthálózati információk 14

Adathelyesség 20

TEIR adatbetöltések 20

# Bevezetés

A dokumentum a Hírközlési és Informatikai Tudományos Egyesület,
 mint megbízó valamint a TFSZ tanácsadó KFT. mint vállalkozó által létrejött 2013 Április 02. kötött szerződés alapján készült.

Jelen dokumentum a célja az „E-Traffic közlekedési használati előrejelző modell megalkotása” című projektben a térinformatikai adatbázisra épülő betöltő/kitároló és lekérdező adatfolyamatok implementálása.

|  |
| --- |
| * Az adatforrások csatolási felületének implementálása.
 |
| * Az adatforrások feldolgozása során a definiált ütemezési igény szerint az előre specifikált formában történik a kitárolás (ha nem valósul meg a kitárolás; eltérő a szerkezet, részleges, vagy hiányos kitárolás történik)
 |
| * Ebben a pontban kell elkészíteni a kitárolás során megjelenő, szélsőséges szcenáriókra felkészítő scripteket
 |

# Adatbetöltés általános leírása

Tekintve, hogy az adatok szerkezete, minősége változó, a stabilitás növelése érdekében olyan ellenőrzési szűrési eljárásokat implementál a rendszer, mely által eldönthető, hogy egy adat elfogadásra, javításra vagy elutasításra kerül-e.

A rendszer üzemeltetésénél elsődleges szempont volt, hogy az adatok töltése automatikusan kerüljön a rendszerbe. Tekintve, hogy az adatok szerkezete változik az idők folyamán, illetve a rendszer üzleti célú üzemeltetésénél az adatforrások elérésének módja is változik, továbbá a strukturális és minőségi diverzifikáció miatt egy célszoftver került kiválasztásra. Ennek oka, hogy egy ilyen eszközzel lényegesen könnyebben kezelhetőek a változások, legyenek azok adat szerkezetét vagy az adatforrás elérését érintő módosítások.

A rendszer adatszükséglete jelenleg három nagy részre osztható:

* Állandó, konstans adatok (területi adatok),
* Historikus, időben változó adatok (gazdasági, úthálózati információk),
* Számított adatok.

A rendszer adatbetöltésének kezelésére a Pentaho cég, Data integration termékét használjuk. Ez a termék, egy klasszikus ETL támogató eszköz, melynek ingyenes és fizetős változata egyaránt fellelhető. A választásnál további fontos szempont volt, hogy a választott rendszer rendelkezzen vásárolható támogatással, illetve stabilan megbízhatóan működjön. Elsődleges szempont volt, hogy az ezen dokumentációban részletezett adat

A Data integration eszköz alkalmas a transzformációk és job-ok létrehozására.

## Adatbetöltések ütemezése

## Transzformáció

Transzformációnak nevezzük azokat a műveletsorokat, melyekkel az adatokat átalakítja, kiegészíti.

## Job

Időzíthető, ütemezhető transzformációk.

# Állandó, konstans adatok

Területi adatok időben stabilak, nem változnak, ezért egyszeri inicializálásuk után további töltést nem igényelnek. Amennyiben mégis változás van, akkor a betöltést és ezzel a változások kezelése egy kézzel indított adatbetöltéssel is kezelhető.

## Területi információk beolvasása

### Beolvasás lépéseinek leírása

1. Adatok beolvasása
	1. település koordináta adatbázisának beolvasása
	2. hibás sorok kiszűrése (ha a település név és a koordináták nem definiáltak)
		1. hibás sorok kiírása,
	3. település név és koordináta adatbázis kiírása az stg\_telepules\_koordinata adatbázis táblába,
	4. KSH település adatforrásának beolvasása,
	5. várakozás az első lépés befejezéséig,
	6. hibás sorok kiszűrése (település név, kistérségi információ és a megye nem definiált)
		1. hibás sorok kiírása
	7. a település koordinátáinak keresése,
	8. koordináta nélküli sorok kiszűrése
		1. hibás sorok kiírása
	9. adatok kiírása a stg\_telepules táblába.



1. Megyék betöltése
	1. főváros értékek cseréje Budapestre,
	2. megye létezésének ellenőrzése,
	3. eddig ismeretlen megyék kiszűrése,
	4. megyeszékhelyek kiválogatása,
	5. megyék duplikációinak megszüntetése,
	6. megyék kiírása a sor\_megye táblába,
	7. megyeszékhely nélküli megyék duplikációnak megszüntetése,
	8. megyék kiírása a sor\_megye táblába.



1. Kistérségek betöltése
	1. főváros értékek cseréje Budapestre,
	2. kistérségek azonosítása,
	3. hibás sorok kiszűrése (kistérség nem beazonosítható)
		1. hibás sorok kiírása
	4. megyék azonosítása
	5. duplikációk megszüntetése,
	6. kistérségek kiírása a sor\_kisterseg táblába.
2. Települések betöltése
	1. település azonosítása,
	2. település létezésének vizsgálata,
	3. meglévő település esetén az adatok aktualizálása, egyébként
	4. megye azonosítása,
	5. kistérség azonosítása,
	6. hibás sorok kiszűrése
		1. hibás sorok kiírása,
	7. települések kiírása a sor\_telepules táblába.

### Strukturális felépítés

#### KSH által közzétett település lista

|  |  |  |
| --- | --- | --- |
| **2011. január 1.** | **2011. január 1.** | **2011. január 1.** |
| **Helység** | **Megye megnevezése**  | **Kistérség**  |
| **megnevezése** | **KSH kódja** | **jogállása** | **kódja** | **neve** | **székhelye** |
| Aba | 17376 | nagyközség | Fejér | 3708 | Abai | Aba |
| Abádszalók | 12441 | város | Jász-Nagykun-Szolnok | 4605 | Tiszafüredi | Tiszafüred |
| Abaliget | 12548 | község | Baranya | 3207 | Pécsi | Pécs |
| Abasár | 24554 | község | Heves | 4004 | Gyöngyösi | Gyöngyös |
| Abaújalpár | 15662 | község | Borsod-Abaúj-Zemplén | 3503 | Encsi | Encs |
| Abaújkér | 26718 | község | Borsod-Abaúj-Zemplén | 3503 | Encsi | Encs |
| Abaújlak | 02820 | község | Borsod-Abaúj-Zemplén | 3510 | Szikszói | Szikszó |
| Abaújszántó | 03595 | város | Borsod-Abaúj-Zemplén | 3512 | Abaúj-Hegyközi | Gönc |
| Abaújszolnok | 26338 | község | Borsod-Abaúj-Zemplén | 3510 | Szikszói | Szikszó |
| Abaújvár | 02273 | község | Borsod-Abaúj-Zemplén | 3512 | Abaúj-Hegyközi | Gönc |
| Abda | 11882 | község | Győr-Moson-Sopron | 3802 | Győri | Győr |
| Abod | 10357 | község | Borsod-Abaúj-Zemplén | 3502 | Edelényi | Edelény |
| Abony | 27872 | város | Pest | 4302 | Ceglédi | Cegléd |
| Ábrahámhegy | 04561 | község | Veszprém | 4906 | Tapolcai | Tapolca |
| Ács | 04428 | város | Komárom-Esztergom | 4104 | Komáromi | Komárom |
| Acsa | 18573 | község | Pest | 4309 | Váci | Vác |

Minta az állomány azon részéből, mely a feldolgozás szempontjából lényeges.

#### Koordináta adatbázis

|  |  |  |
| --- | --- | --- |
| Helységnév | Keleti hosszúság, fok:perc. századperc | Északi szélesség, fok:perc. századperc |
| Aba | 18:31.49 | 47:02.06 |
| Abádszalók | 20:35.94 | 47:28.48 |
| Abaliget | 18:07.09 | 46:08.63 |
| Abasár | 20:00.50 | 47:48.04 |
| Abaújalpár | 21:14.14 | 48:18.37 |
| Abaújkér | 21:12.08 | 48:18.47 |
| Abaújlak | 20:57.45 | 48:24.35 |
| Abaújszántó | 21:11.18 | 48:16.74 |
| Abaújszolnok | 20:58.63 | 48:22.35 |
| Abaújvár | 21:18.99 | 48:31.59 |
| Abda | 17:32.65 | 47:41.79 |
| Abod | 20:47.65 | 48:23.60 |
| Abony | 20:00.72 | 47:11.56 |
| Ábrahámhegy | 17:34.36 | 46:48.96 |
| Ács | 18:00.94 | 47:42.77 |
| Acsa | 19:23.30 | 47:47.71 |

Minta a koordinált adatállományból

## Hazai úthálózat adatbázisa

Ezen kitárolás állományai a Közút Kezelő Nonprofit Zrt. kezelésében lévő úthálózatot leírását tartalmazzák.

Az állományt a QGIS-ben MMQGIS modul exportjaként olvassa be. A rendszer számára az adatállományok szerkezete kötött. Amennyiben ettől eltérést tapasztal, az adtok beolvasását a rendszer megtagadja.

### Beolvasás lépéseinek leírása

1. Adatállományok beolvasása
	1. a táblához kapcsolódó automatikus számláló nullázása,
	2. a csomópontokat tartalmazó shape állomány betöltése az stg\_nodes táblába,
	3. a csomóponthoz tartozó adatok betöltése az stg\_node\_attribs táblába.
2. Csomópontok (node) beazonosítása
	1. kezdőpontok leválogatása,
	2. végpontok leválogatása,
	3. csomópontok kiírása a stg\_real\_nodes táblába.
3. Új csomópontok (node) beazonosítása, kiírása
	1. csomópontok beolvasása,
	2. létezésének ellenőrzése az adatbázisban,
	3. új csomópontok kiszűrése,
	4. az új csomópontok kiírása a sor\_node táblába,
	5. az új csomópontok azonosítójának kigyűjtése a későbbi adatfeldolgozási folyamatokhoz
4. Új útszakaszok (link) beazonosítása
	1. új útszakaszok lekérdezése,
	2. új útszakaszok kiírása a sor\_link táblába,
	3. az új linkek azonosítójának kigyűjtése a későbbi feldolgozási folyamatokhoz.
5. Új ábrázolási pontok (point) leválogatása, kiírása
	1. új ábrázolási pontok kigyűjtése,
	2. új ábrázolási pontok kiírása a sor\_pont táblába.

SQL

Csomópontok (node) beazonosítása

Kezdőpontok

SELECT o.\*, na.kkod as OKA\_NODE FROM

(

 SELECT min(number) as mnumber FROM etraffic.stg\_nodes group by shape\_id

) as t

left JOIN etraffic.stg\_nodes as o

ON t.mnumber = o.number

left join etraffic.stg\_node\_attribs as na on o.shape\_id = na.shape\_id

Végpontok

SELECT o.\*, na.vvkod as OKA\_NODE FROM

(

 SELECT max(number) as mnumber FROM etraffic.stg\_nodes group by shape\_id

) as t

left JOIN etraffic.stg\_nodes as o

ON t.mnumber = o.number

left join etraffic.stg\_node\_attribs as na on o.shape\_id = na.shape\_id

Új útszakaszok (link) beazonosítása

SELECT nmin.sor\_node\_id as from\_node ,nmax.sor\_node\_id as to\_node,

na.kszam, na.RSHOSSZ as szho, null as DIJ, null as ktype, null as LAN, NULL as KUTKA\_ID , null as mkod

, case

 when na.pkod = 0 then 2

 when na.pkod = 1 then 1

 when na.pkod = 2 then -1

end as direction,

YEAR(CURDATE())-1 as FROM\_YEAR, null as TO\_YEAR

, nmax.shape\_id

FROM

(

SELECT min(number) as min\_number, max(number) as max\_number

FROM `etraffic`.`stg\_real\_nodes`

WHERE sor\_node\_id is not null

group by shape\_id

) as n

LEFT JOIN `etraffic`.`stg\_real\_nodes` as nmin ON n.min\_number = nmin.number

LEFT JOIN `etraffic`.`stg\_real\_nodes` as nmax ON n.max\_number = nmax.number

left join etraffic.stg\_node\_attribs as na on nmax.shape\_id = na.shape\_id

Új ábrázolási pontok (point) leválogatása

SELECT n.\* , sl.id as link\_id

FROM etraffic.stg\_nodes as n

LEFT join etraffic.stg\_real\_nodes as rn on n.number = rn.number

LEFT JOIN etraffic.stg\_link as sl ON sl.shape\_id = n.shape\_id

WHERE rn.shape\_id is null

### Strukturális felépítés

#### Pont állományok

Ezen állományok formátuma az alábbi

|  |  |  |  |
| --- | --- | --- | --- |
| Oszlop neve | Értéke | Megkötés | Megjegyzés |
| shapeid | Numerikus | Kötelező | 0-tól sorszámozódik |
| x | Numerikus, tizedes | Kötelező |  |
| y | Numerikus, tizedes | Kötelező |  |

Az oszlopsorrend kötött, minden változást elutasítással kezel a rendszer.

|  |  |  |
| --- | --- | --- |
| shapeid | x | y |
| 0 | 637255.799902 | 235282.299988 |
| 0 | 637255.199838 | 235251.300436 |
| 0 | 637256.899678 | 235210.89954 |
| 0 | 637260.999774 | 235174.400084 |
| 0 | 637262.399582 | 235164.699732 |
| 0 | 637263.899742 | 235154.700372 |
| 0 | 637265.599582 | 235144.80034 |
| 0 | 637267.500126 | 235134.999636 |
| 0 | 637269.399646 | 235125.399636 |
| 0 | 637271.30019 | 235116.00034 |
| 0 | 637273.400414 | 235106.499668 |
| 0 | 637275.599966 | 235097.100372 |
| 0 | 637277.799518 | 235087.800404 |
| 0 | 637280.100446 | 235078.599764 |
| 0 | 637282.60003 | 235069.5005 |
| 0 | 637285.199966 | 235060.29986 |
| 0 | 637287.999582 | 235050.999892 |
| 0 | 637290.999902 | 235041.699924 |
| 0 | 637294.199902 | 235032.200276 |

#### Úthálózati információk

|  |  |  |  |
| --- | --- | --- | --- |
| Oszlop neve | Értéke | Megkötés | Megjegyzés |
| shapeid | numerikus | Kötelező | Pontállomány shapeid-hez rendelt érték |
| KSZAM | Alfanumerikus | Kötelező |  |
| PKOD | Numerikus |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| shapeid | KSZAM | PKOD | KKOD | VVKOD | KSZELV | VSZELV | RSHOSSZ | ANF | ANET | MOF | ONGJ | OJ | OMOT |
| 0 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 1 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 2 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 3 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 4 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 5 | M0 | 0 | R130881K | Y130883 | 0 + 000 | 2 + 742 | 2846.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 6 | M0 | 1 | Y130883 | R130883B | 2 + 742 | 4 + 322 | 799.0 | 36663.0 | 8441.0 | 2884.0 | 6464.0 | 25806.0 | 25806.0 |
| 7 | M0 | 1 | R130883B | R130887A | 4 + 322 | 8 + 641 | 2166.0 | 53467.0 | 17074.0 | 4206.0 | 11285.0 | 34810.0 | 34810.0 |
| 8 | M0 | 1 | R130883B | R130887A | 4 + 322 | 8 + 641 | 2166.0 | 53467.0 | 17074.0 | 4206.0 | 11285.0 | 34810.0 | 34810.0 |
| 9 | M0 | 1 | R130883B | R130887A | 4 + 322 | 8 + 641 | 2166.0 | 53467.0 | 17074.0 | 4206.0 | 11285.0 | 34810.0 | 34810.0 |
| 10 | M0 | 1 | R130885B | Y131579 | 8 + 641 | 12 + 140 | 1750.0 | 73852.0 | 20147.0 | 5810.0 | 16051.0 | 48051.0 | 48051.0 |
| 11 | M0 | 1 | R130885B | Y131579 | 8 + 641 | 12 + 140 | 1750.0 | 73852.0 | 20147.0 | 5810.0 | 16051.0 | 48051.0 | 48051.0 |
| 12 | M0 | 1 | R130885B | Y131579 | 8 + 641 | 12 + 140 | 1750.0 | 73852.0 | 20147.0 | 5810.0 | 16051.0 | 48051.0 | 48051.0 |
| 13 | M0 | 1 | R130885B | Y131579 | 8 + 641 | 12 + 140 | 1750.0 | 73852.0 | 20147.0 | 5810.0 | 16051.0 | 48051.0 | 48051.0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| EV | ASZ | BUSZCS | BUSZE | OBUSZ | NYSZER | POTKTGK | KTGK | NTGK | KNTGK | OTGK |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3485.0 | 16.0 | 190.0 | 206.0 | 3634.0 | 458.0 | 3727.0 | 2149.0 | 798.0 | 7056.0 |
| 2012.0 | 3480.0 | 41.0 | 191.0 | 232.0 | 8651.0 | 686.0 | 4161.0 | 1696.0 | 1178.0 | 12231.0 |
| 2012.0 | 3480.0 | 41.0 | 191.0 | 232.0 | 8651.0 | 686.0 | 4161.0 | 1696.0 | 1178.0 | 12231.0 |
| 2012.0 | 3480.0 | 41.0 | 191.0 | 232.0 | 8651.0 | 686.0 | 4161.0 | 1696.0 | 1178.0 | 12231.0 |
| 2012.0 | 2531.0 | 22.0 | 390.0 | 412.0 | 9950.0 | 1576.0 | 12244.0 | 4046.0 | 1166.0 | 16805.0 |
| 2012.0 | 2531.0 | 22.0 | 390.0 | 412.0 | 9950.0 | 1576.0 | 12244.0 | 4046.0 | 1166.0 | 16805.0 |
| 2012.0 | 2531.0 | 22.0 | 390.0 | 412.0 | 9950.0 | 1576.0 | 12244.0 | 4046.0 | 1166.0 | 16805.0 |
| 2012.0 | 2531.0 | 22.0 | 390.0 | 412.0 | 9950.0 | 1576.0 | 12244.0 | 4046.0 | 1166.0 | 16805.0 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SZGK | OSZGK | MKP | KPF | LASSU | SPEC | FMEGB | ADATFORR | SZAMLNAP | JELLEG1 | JELLEG2 | FMEGJ |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 14641.0 | 18368.0 | 176.0 | 0.0 | 0.0 | 17.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 17998.0 | 22159.0 | 188.0 | 0.0 | 0.0 | 20.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 17998.0 | 22159.0 | 188.0 | 0.0 | 0.0 | 20.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 17998.0 | 22159.0 | 188.0 | 0.0 | 0.0 | 20.0 | 10.0 | E | 0.0 | B | 1 | NULL |
| 18461.0 | 30705.0 | 129.0 | 0.0 | 0.0 | 67.0 | 2.988423782 | M | 285.0 | B | 1 | NULL |
| 18461.0 | 30705.0 | 129.0 | 0.0 | 0.0 | 67.0 | 2.988423782 | M | 285.0 | B | 1 | NULL |
| 18461.0 | 30705.0 | 129.0 | 0.0 | 0.0 | 67.0 | 2.988423782 | M | 285.0 | B | 1 | NULL |
| 18461.0 | 30705.0 | 129.0 | 0.0 | 0.0 | 67.0 | 2.988423782 | M | 285.0 | B | 1 | NULL |

Adatminta a forgalmi atatok adatforrásból. A táblázat hozzsa meghaladta a lapszélességet, eztért az első 15 sor adatait 3 részre darabolva egymás alá helyeztük.

### Adathelyesség

Az adathelyesség ellenőrzésére több SQL scriptet készítettünk, de a tapasztalat az, hogy az adatforrások minősége évenként változik, ezért az alapfeltévés az, hogy az exportok minden esetben adathelyesek.

## TEIR adatbetöltések

A teírből származó adatformátuma egységes, az adatok staging beolvasását és adattáblába másolását, a beolvasás egyszerűsége miatt, ezen dokumentáció nem részletezi. Ezen adatok struktúrális renezése csak az elemzői schema előállításának érdekében szükésges, amennyiben ezen indikátorok véglegesednek, a modellezésben kialakul a végleges súlyozásuk az adatbetöltő folyamban is megvalósításra kerülnek -az ott kidolgozott eredmények alapján- a transzformációs folyamatok.