

Adatkockától a felszínborítás-térképig

Dr. Petrik Ottó

Fény – Tér – Kép

Tihany, 2019. november 14-15



A felszínborítás térképekről...

- Miért fontos?
 - Napjaink környezeti változásainak fő előidézője az emberi tevékenység, ennek monitorozása a felszínborítási és a változás adatokon keresztül valósulhat meg
- Adatbázis szemlélet:
 - **dekompozíció**s eljárás:
domén (mezőgazdasági területek) - al-domén (állandó növényi kultúrák) - al-al-domén/osztály (szőlők) - (al-al-al-domén/osztály: kordonos művelés - magas kordon)
Az így létrejött osztályok együttese nem feltétlenül alkot kapcsolható adatmodellt!
 - **szintézis**:
osztály (szőlő kataszter) - al-domén (állandó növényi kultúrák) - domén (mezőgazdasági területek)
Nem feltétlenül alkot teljes térképet (lyukas) !

Az adatkockákról... (1)

- Elektronikus számítógépes kezdetek: katonai számítások
- 1960-es évektől: fő alkalmazás az adatfeldolgozás
 - Népszámlási előzmények: mechanikus gép (USA, 1890)
 - Adatformátumok szabványosítása
 - Általános célú adatbázis-kezelő szoftverek kifejlesztése
- **Adatbázis:** a valós világ egy részhalmazának leírására használt adatok összefüggő, rendezett halmaza
 - Implementációja az adatmodell:
 - *Egyed vagy entitás:* a valós világban létező dolog, amit tulajdonságokkal akarunk leírni
 - *Tulajdonság vagy attribútum:* az egyed (lehetőleg egyértelmű) jellemzője
 - Az egyedek között *kapcsolat* van

Az adatkockákról... (2)

Az adatbázis rendszerek felosztása:

- „hagyományos”: több konkurensen futó tranzakció
 - rendszerint adatbázis módosítási műveletet is tartalmaz
 - tranzakció-feldolgozó rendszernek (OLTP - On-Line Transaction Processing) nevezzük
- „adattárház”: lekérdezési és elemzési célra tervezett és létrehozott adatbázis
 - rendszerint nincs adatbázis módosítás
 - kevés vagy nincs párhuzamos lekérdezés
 - OLAP – On-Line Analytical Processing

Az adatkockákról... (3)

Az adattárház nem más, mint az adatok:

- *témaorientált* – az adott tárgyterület köré csoportosított
- *integrált* – a különböző forrásrendszerek adatait szabványos, egységes formátumra alakítva tároló
- *tartós* – alapvetően változatlan
- *időfüggő* – időbélyeggel ellátott

historikus gyűjteménye.

(William H. Inmon: *Building the Data Warehouse*, Wiley Publishing, 2005, 4th Edition)

Az adatkockákról... (4)

Az adattárház típusa az alkalmazott adatbázis-kezelő logikájának függvényében lehet:

- *MOLAP* – multidimensional OLAP: speciális adatbáziskezelővel multidimenzionális célstruktúrában tárolja az adatokat
- *ROLAP* – relational OLAP: a hagyományos relációs adatbázisban tároljuk az adatokat
- *HOLAP* – hybrid OLAP: relációs és multidimenzionális tárolás

Adatkockának nevezzük az adat-absztrakciós módszert, amelyben az adatokat úgy kezeljük, mintha egy *n-dimenziós kocka* pontjai lennének, ahol a dimenziók azokat a tulajdonságokat jelentik, amelyeket elemezni szeretnénk. A kocka név általános: az *n-dimenziós adathalmazt* is kockának nevezzük.

Az adatkockákról... (5)

Az adattárház séma tervezése egy 4 lépéses folyamatból áll:

- a modellezendő folyamat kiválasztása
- a felbontás (granularity) meghatározása
- dimenziók meghatározása, kidolgozása
- tematikus adatok meghatározása

Az adattárház tematikus adatait *definíció szerint* az **adott lekérdezés típusokra jól előkészítetten kell létrehozni.**

(Kimball R., Mundy J. and Thornthwaite W., *The Microsoft Data Warehouse Toolkit*, 2011, 2nd Edition, Wiley)

A tudáskinyerés folyamata

1. Alkalmazási terület felmérése, előzetes ismeretek rendszerezése
2. Céladatbázis kiválasztása, létrehozása
3. Adattisztítás, előfeldolgozás
4. Adatintegráció
5. Adattér csökkentés: cél szempontjából fontos attribútumok kiemelése
6. Adatbányászati algoritmusok kiválasztása (klaszterezés, mintakeresés, osztályozás)
7. Adatbányászati algoritmus, paraméterek előállítása
8. Algoritmus alkalmazása
9. Kinyert információ értelmezése, finomítások
10. A megszerzett tudás megerősítése, összevetése az elvárásokkal, dokumentálás

Térinformatikai sajátosságok (1)

A térinformatika valós földrajzi objektumokról alkot modellt, ennek legfontosabb eleme a geometria, amit 2 nagy csoportba szokás sorolni:

- Szabályos geometriai elemek: pl. a négyzetes elemekből álló grid – a raszteres adatmodell
- Szabálytalan geometria elemek: pont, vonal, poligon – a vektoros adatmodell

„Vegyük észre”, hogy az eddigiek alapján az adatkockát csak raszteres adatokkal lehet feltölteni!

- Granularity: (a kocka felbontásán) azt a legkisebb adategységet értjük, mely egységekben az adatot még elérhetővé szeretnénk tenni
- Ha a kockában több különböző adat is szerepel, akkor mindegyik granularitása meg kell egyezzen!

Térinformatikai sajátosságok (2)

Vektoros adat: raszterizálni kell:

- Adatvesztés – a pixelméret (granularity) határozza meg az információ csökkenés mértékét
- Egyértelmű megfeleltetés (?): több vektoros elem esik a pixelre
 - „grid-approach” – denzitások
 - „csak egy maradhat”: lekérdezési sorrend, prioritási tábla, pixelközép, legnagyobb területű poligon

Raszteres adat: (eltérés esetén) át kell mintázni az adatkocka felbontására:

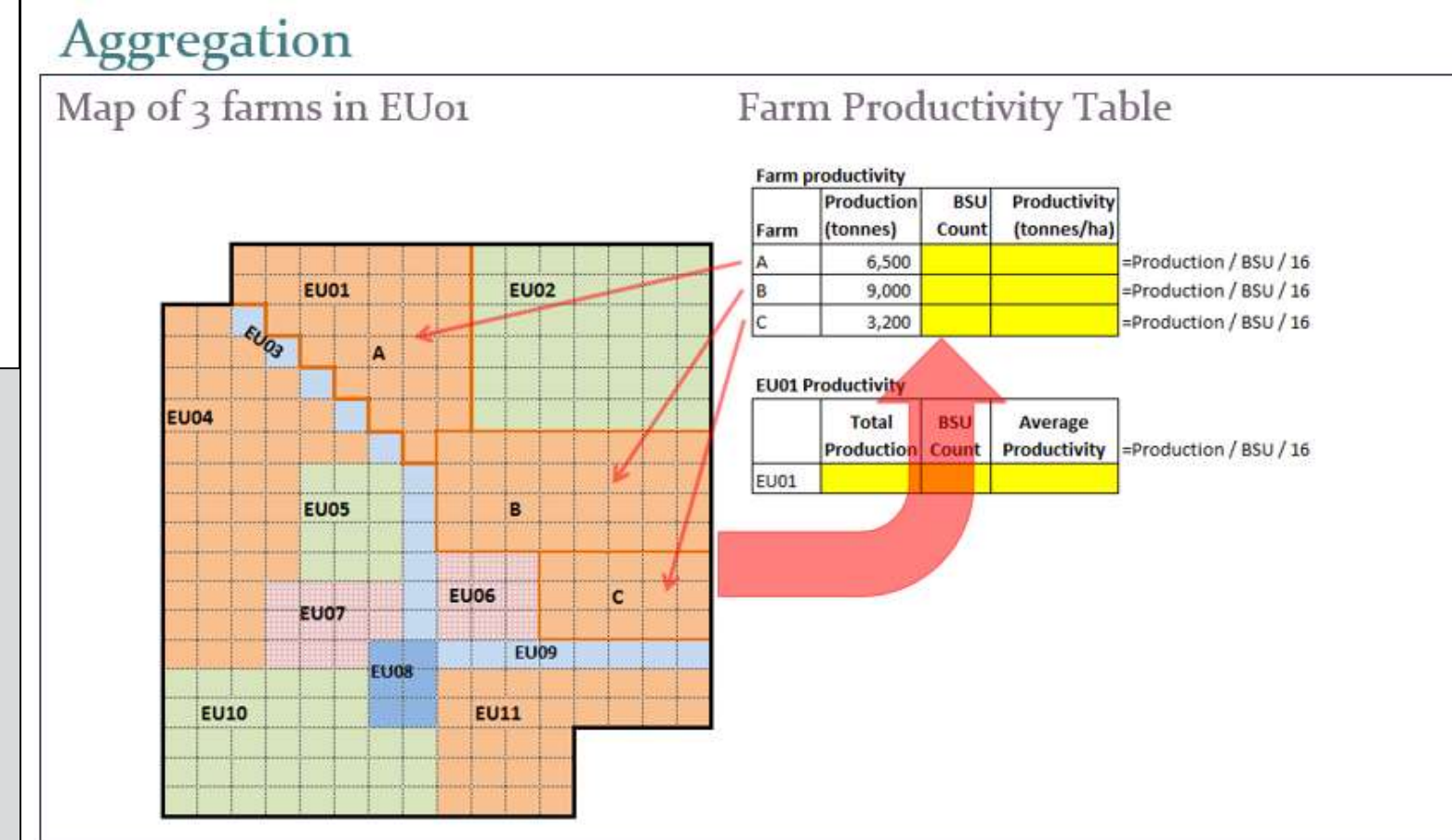
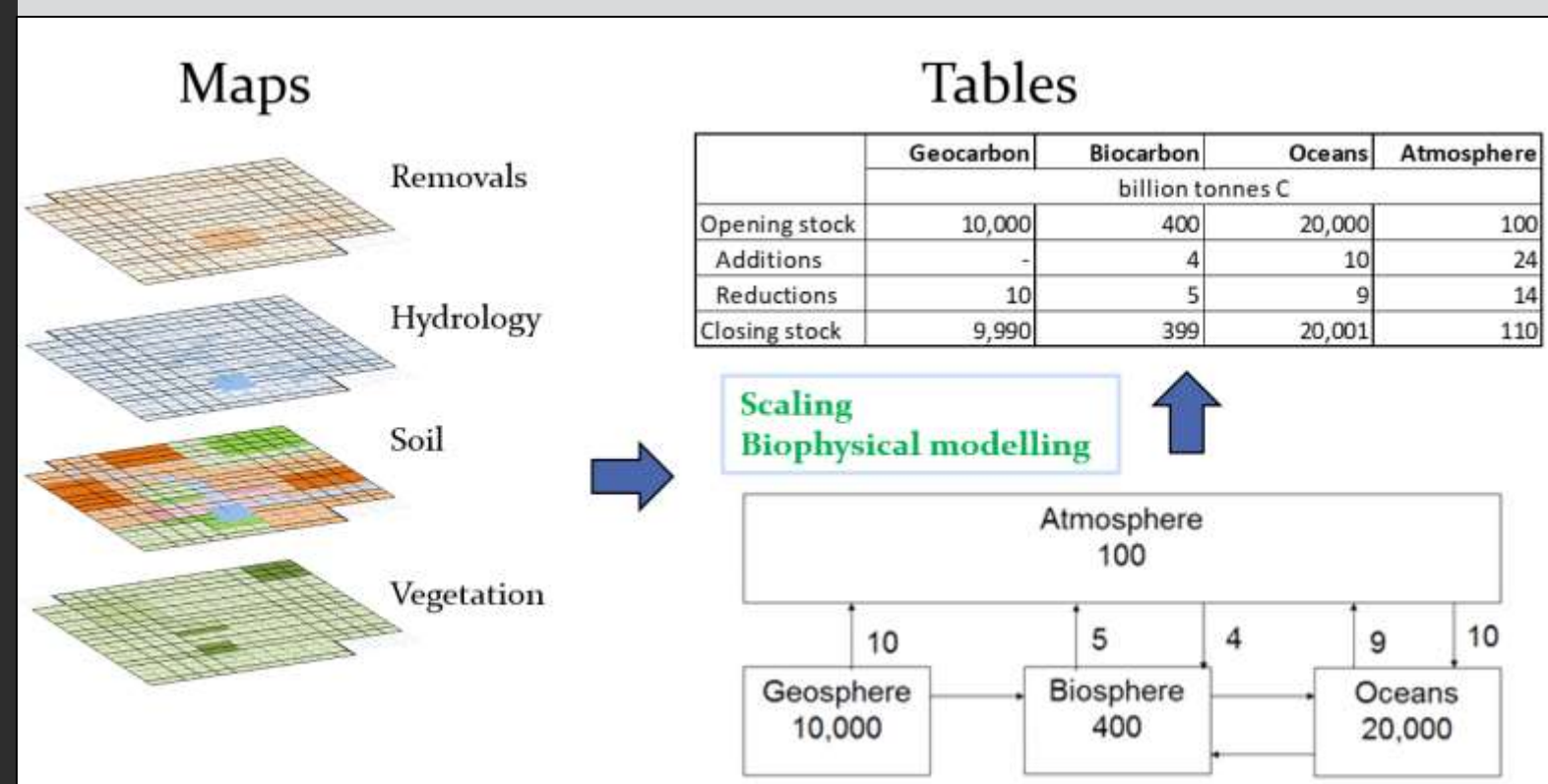
- tematikus vs folytonos: többségi (majority), priorizálás
- interpoláció: NN vs CC/BL/IDW

Adatkocka megvalósítás (1)

SEEA-EEA (JEDI):

Az Európai Környezetvédelmi Ügynökség (EEA) által készített integrált statisztikai keretrendszer a biofizikai adatok rendezésére, az ökoszisztéma-szolgáltatások mérésére, az ökoszisztéma-eszközök változásának követésére valamint ezen információk összekapcsolása a gazdasági és emberi tevékenységekkel

(SEEA: Shared Environmental Ecosystem Accounting – JEDI: Joint Environmental Data Infrastructure)



Recommended three levels: hierarchical and mutually exclusive:

1. Basic Spatial Unit (BSU)
 - Pixel or grid cell
2. Ecosystem Unit (EU)
 - Homogenous according to criteria (cover, slope, drainage area, elevation...)
 - Consolidate for tables by EU type
3. Ecosystem Reporting Area (ERA)
 - For reporting (sub-drainage area, administrative area...)

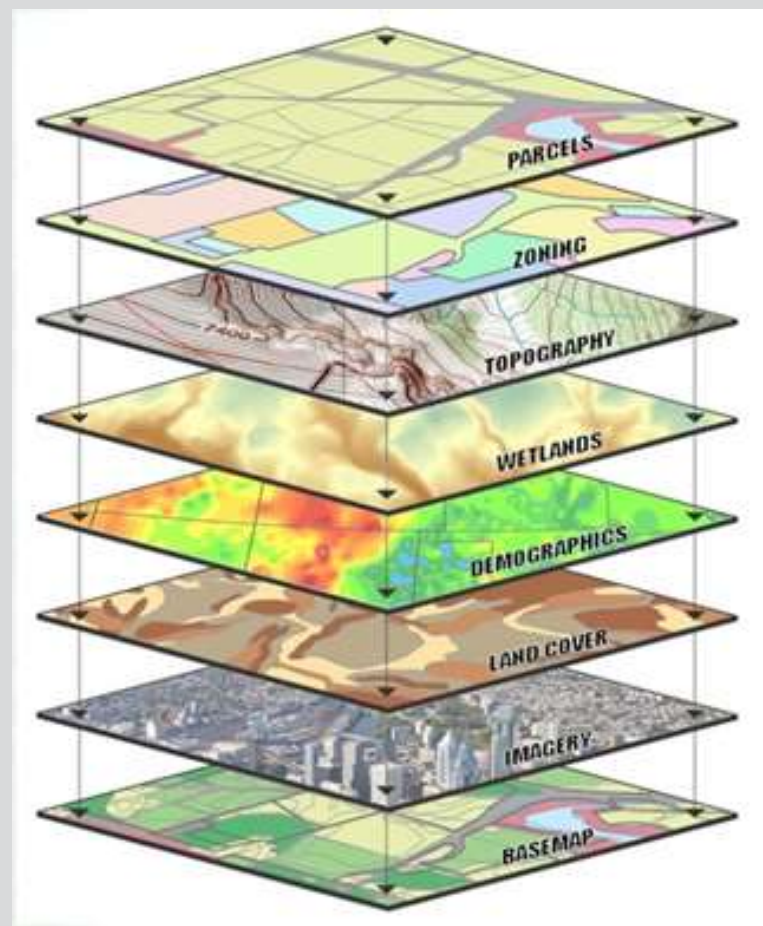
→ Establishes **Ecosystem Extent Account**

Ez egy „valódi” adatkocka SQL alapú lekérdezéssel

„Adatkocka” megvalósítás (2)

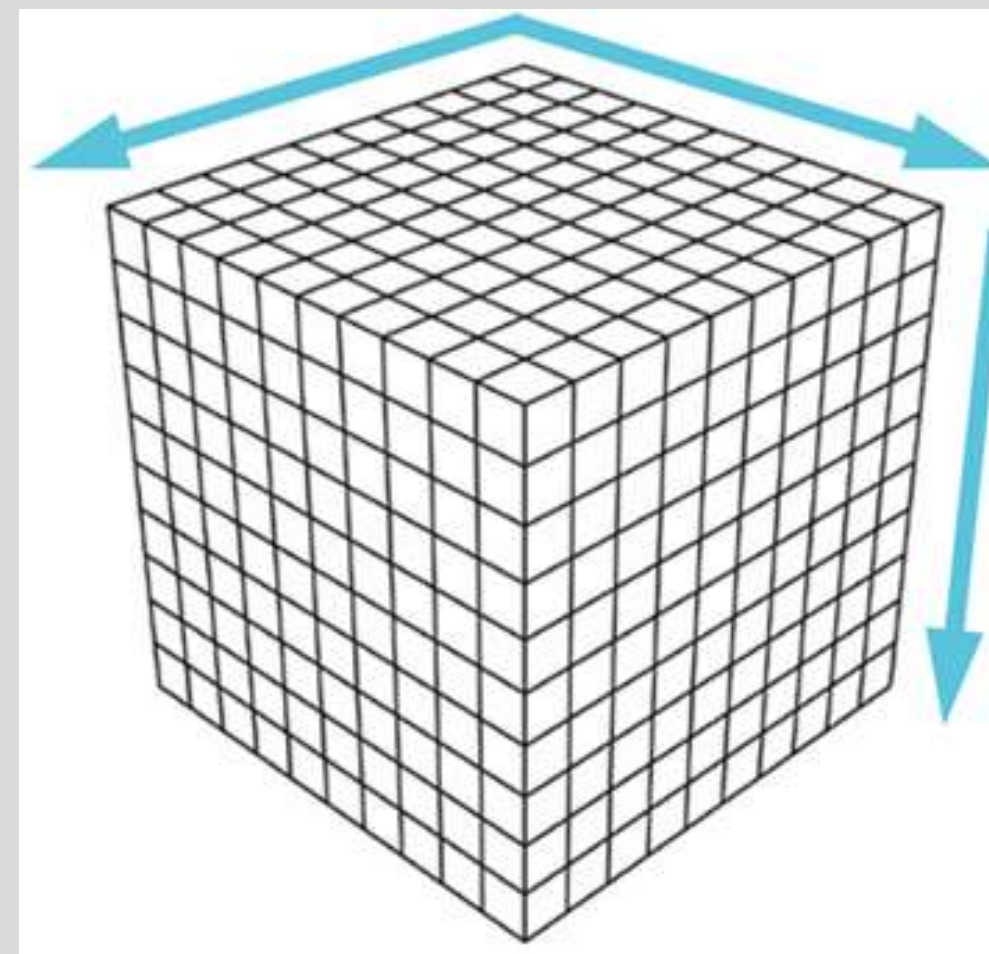
Az alulról építkező térképezés modellje a Nemzeti ökoszisztéma szolgáltatás-térképezés projektben, melynek célja felmérni, térképezni a természetes ökoszisztémákat. (KEHOP-4.3.0-VEKOP-15-2016-00001)

Bemenő adatok



- Meglevő térinformatikai adatbázisok
- Távérzékeléssel gyűjtött adatok
- Új térképezés
- Referencia adatok

„Adatkocka”



- Egységes geometria (10 / 20m raszter)
- Egységes vetület (EOV / ETRS-LAEA)
- INSPIRE kompatibilis metaadatok
- Cella alapú lekérdezések
- Egyéb lehetséges „adatbányászat”

Felszínborítás alaptérkép



- 110 - Mesterséges felszín
- 121 - Települési zöld felület fákkal
- 122 - Települési zöld felület fák nélkül
- 131 - Szilárd burkolatú út, vasút
- 132 - Földút
- 210 - Szántó
- 220 - Állandó kultúra
- 230 - Komplex területek
- 300 - Gyepterületek és egyéb lágyszárú növényzet
- 314 - Padkás szikesek
- 330 - Máshová nem besorolható lágyszárú állomány
- 400 - Erdők és fás területek
- 440 - Faültetvény
- 450 - Máshová nem besorolható fásszárú állomány
- 510 - Vízben álló mocsári/lápi növényzet
- 520 - Láp és mocsárrétek
- 530 - Fás lápok
- 620 - Folyó

- 20m raszter, európai vetület
- MAES alapú kategóriarendszer
 - ÁNÉR-ből optimalizálva
 - Bővíthető a rendelkezésre álló bemenő adatok szabta korlátokkal

Ez egy „virtuális” adatkocka (a számítógép memóriájában készül el)
a lekérdezés Python scriptben fut

Köszönöm a figyelmet!

Dr. Petrik Ottó



1149 Budapest,
Bosnyák tér 5.

TELEFON: +36 (1) 460-4178



www.lechnerkozpont.hu

EMAIL:

OTTO.PETRIK@LECHNERKOZPONT.HU

