




# Zachránia (big) data planétu?

Miloš Molnár, trivia data

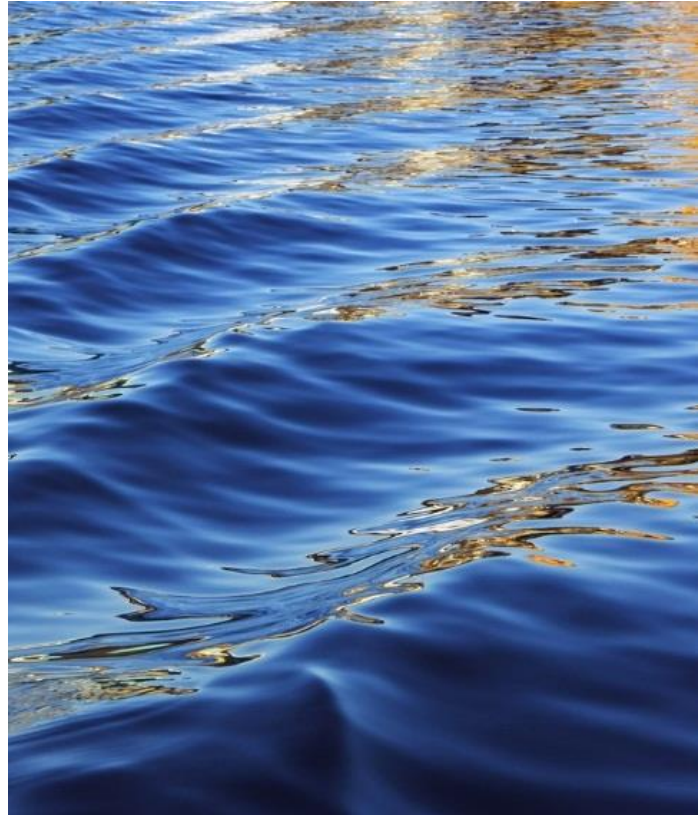


# Agenda

- Príklady prínosu dát pre životné prostredie
- Výzvy spracovania dát o životnom prostredí
- Metodický prístup - podmienka úspechu



Príklady prínosu dát  
pre životné  
prostredie



# Oblasti kde dáta, analýzy a umelá inteligencia môžu pomôcť / pomáhajú životnému prostrediu:

- Smart poľnohospodárstvo a distribúcia potravín
- Predpovede počasia a klimatických zmien
- Energetika
- Monitoring znečisťovania ovzdušia a vôd
- Manažment odpadov
- Logistika (tovary aj ľudia)

**Niekoľko zaujímavých faktov:**

69% vyťaženej vody sa používa v poľnohospodárstve

37% využiteľnej plochy zeme sa už používa na poľnohospodárske účely

1/3 vypestovaných potravín sa vyhodí alebo neskonzumuje

10% svetového HDP je priamo alebo nepriamo naviazaných na turizmus

# Príklady projektov

- Monitoring biodiverzity
  - <https://www.conservation.org/> NGO umiestnila viac ako 1000 kamier s detekciou pohybu. Pokrytie na cca 2000 km<sup>2</sup>. Umelá inteligencia pomohla rozpoznať zvieratá, ktoré prechádzali pred kamerami. Dáta obohatili o klimatické dáta, flóru, lokácie (blízkosť človeka) a pod. = veľmi dobrá údaje o vplyvoch na biodiverzitu.
- Monitoring miesta chránených živočíchov
  - Intel vyvinul zariadenie vo veľkosti kreditnej karty na sledovanie nosorožcov. Zrealizovali aj pilotný projekt a overili možnosť sledovania všetkých.
- Sledovanie čmeliakov
  - Prostredníctvom appky sledovali výskyt čmeliakov. Z fotiek získavali údaje aj o kvetoch na ktorých boli odfotené, či mieste výskytu.
- Ochrana lesov a boj proti nelegálnej ťažbe v Rumunsku a v Indonézii
  - V Indonézii vybavili staré telefóny solárnymi mini panelmi a „odpočúvali“ les. V prípade zvukov pílenia zasahovali miestne zložky.
  - V Rumunsku musel byť každý náklad dreva označený QR kódom a pomocou appky sa dala overiť legálnosť a povolenosť prepravy.



# Projekt Global Forest Watch

- [globalforestwatch.org](https://globalforestwatch.org)
- na základe dát zo satelitu sú sledované lesy a zalesnenie
- za 5 rokov 700 tisíc obrázkov, cca 20 terra-pixelov
- dáta sú dostupné ako open data



-  FORES CHANG
-  FOREST CHANGE
-  LAND CC
-  LAND COVER
-  LAND U
-  LAND USE
-  CLIMA
-  CLIMATE
-  BIODIVER
-  BIODIVERSITY

**LEGEND** | **ANALYSIS**


Tree cover gain - 2001-2012

- Tree cover gain

Tree cover loss - 2001-2020

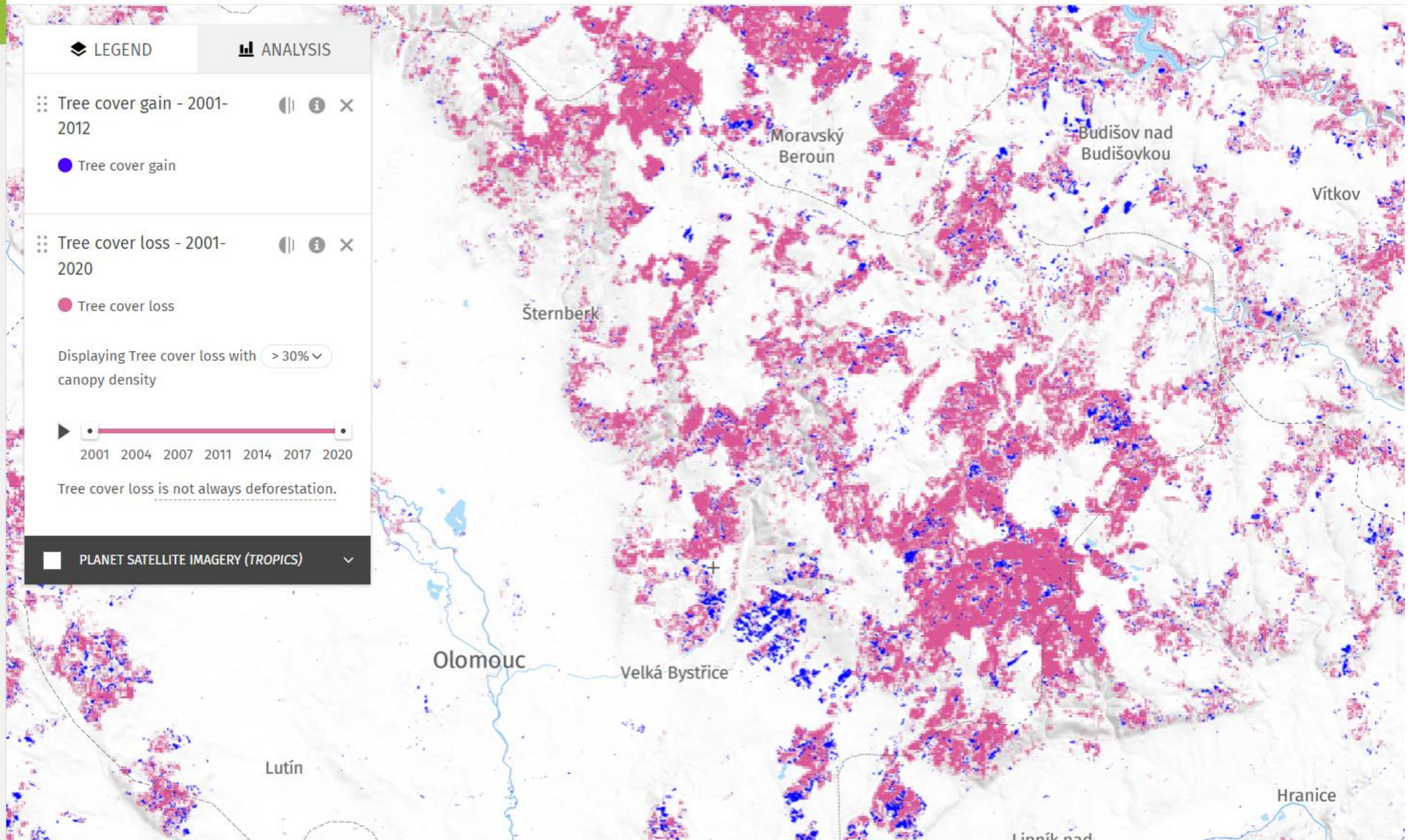
- Tree cover loss


Displaying Tree cover loss with canopy density > 30% ▾




Tree cover loss is not always deforestation.

PLANET SATELLITE IMAGERY (TROPICS) ▾





Výzvy spracovania  
dát o životnom  
prostredí





# Aké sú (a budú) dáta o životnom prostredí

## Vlastnosti dát

- heterogénne
- rýchle
- veľké
- s krátkou životnosťou
- neštandardizované
- „nekvalitné“
- premenlivé

## Problémy / výzvy

- zber
- spracovanie
- uloženie
- analýza
- vyťažovanie hodnoty

# Dáta zo senzorov - časť prvá

Senzory niekedy  
klamú

Aj keď senzory  
neklamú,  
nemusia hovoriť  
celú pravdu.

Extrakcia  
užitočných  
informácií nie je  
jednoduchá.

Senzory typicky  
nemerajú to čo  
nás zaujíma  
priamo.

Senzory samé  
osebe nevedia  
konať.

Nepredpokladajte  
, že dáta zo  
senzorov sú  
presné,  
kompletné a  
konzistentné.



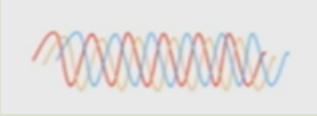

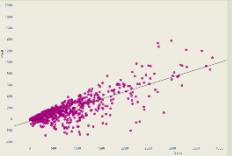
Najlepšie je zberať  
"hrubé" dáta, a aj  
porozumieť ako  
dáta zberajú.


Potrebné je  
aplikovať rôzne  
metódy pri čistení  
dát, hľadani  
súvislostí,  
odchýlok a pod. a  
často doplnkové  
dáta

Indície a "zdravý  
rozum" nás môžu  
oklamať. Len  
matematika je  
neomylná.


Získané  
informácie je  
potrebné začleniť  
do ekosystému, až  
potom môžu byť  
materializované  
na hodnotu.

# Dáta zo senzorov - časť druhá

	„hrubé“ dáta zo senzorov	očistené dáta	detekcia udalosti	označené dáta	hľadanie asociácií / predikcie
					
Proces	zber najmenejších dát umožní špecifickú detekciu udalosti	Interpolácia chýbajúcich hodnôt, korekcia extrémnych hodnôt, rekalibrácia, a pod.	Identifikácia zmien stavu, porovnanie so signatúrou z minulosti	Porovnanie a korelácia s pozorovaním alebo meraním človekom	Porovnanie a korelácia s inými zdrojmi dát
Dáta	„hrubé“ dáta	„hrubé“ dáta očistené, referenčné hodnoty, master dáta	dáta o udalostiach, celoplošné dáta zariadenia/í, dáta o prostredí	prevádzkové a údržbové dáta	historické dáta celého prostredia a všetkých zariadení / senzorov
Analýza	real-time detekcia hodnôt	interpolácie, vyhládanie, a pod.	časové rady, paterny, podobnosť, korelácie	komparácie, prekryvanie	strojové učenie



# Význam metodického prístupu



# Spracovanie a analýza dát – metodiky a štandardy

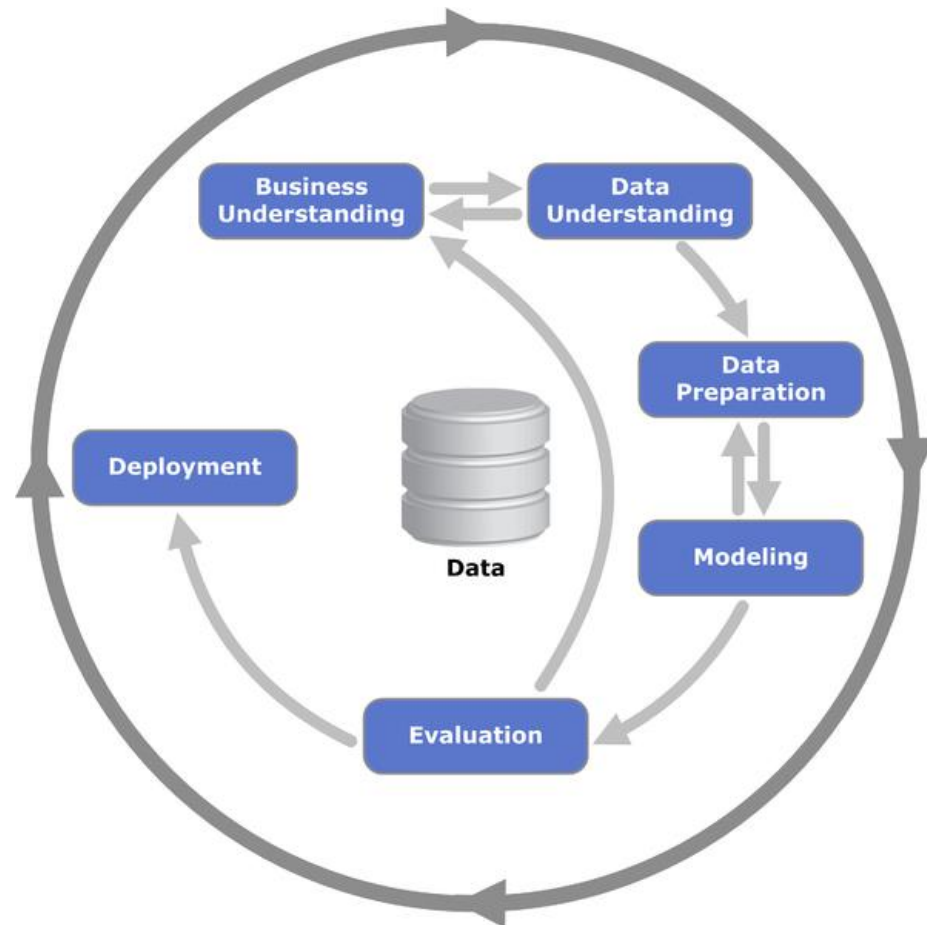
## Metodiky a štandardy

- **Cross Industry Standard Process for Data Mining** - (CRISP-DM) - vyvinutý a adoptovaný DaimlerChrysler, SPSS (teraz IBM) a NCR (teraz Teradata)
- **SEMMA** - vyvinutá spoločnosťou SAS Institute. Je to skôr logická organizácia činností a funkčných nástrojov pre Data Mining.
- **Foundational Methodology for Data Science** - redefiníciu a rozšírenie metodiky CRISP-DM od IBM.

## Benefity metodického prístupu

- Manažovateľnosť
- Dlhodobá udržateľnosť výsledkov
- Flexibilita
- Opakovateľnosť
- Efektivita – funkčné vzory

# Spracovanie a analýza dát – metodiky a štandardy



- Metodika vs. procesný štandard
- „doplnkové“ štandardy – dôležitejšie ako základný štandard
  - Názvoslovie,
  - Popisy,
  - Postupy dokumentovania
  - Postupy verziovania
  - Správa kódu
  - Riadenie zmien
  - ...
- Najdôležitejšia / najnáročnejšia časť jej uplatnenia je „príprava údajov“ a „modelovanie“
- **Akákoľvek metodika je lepšia ako žiadna!**

Ďakujem za pozornosť!

Miloš Molnár  
[molnar@triviadata.com](mailto:molnar@triviadata.com)  
+421903554330