



# Earth Observation data is now free...and big

graphics: NASA

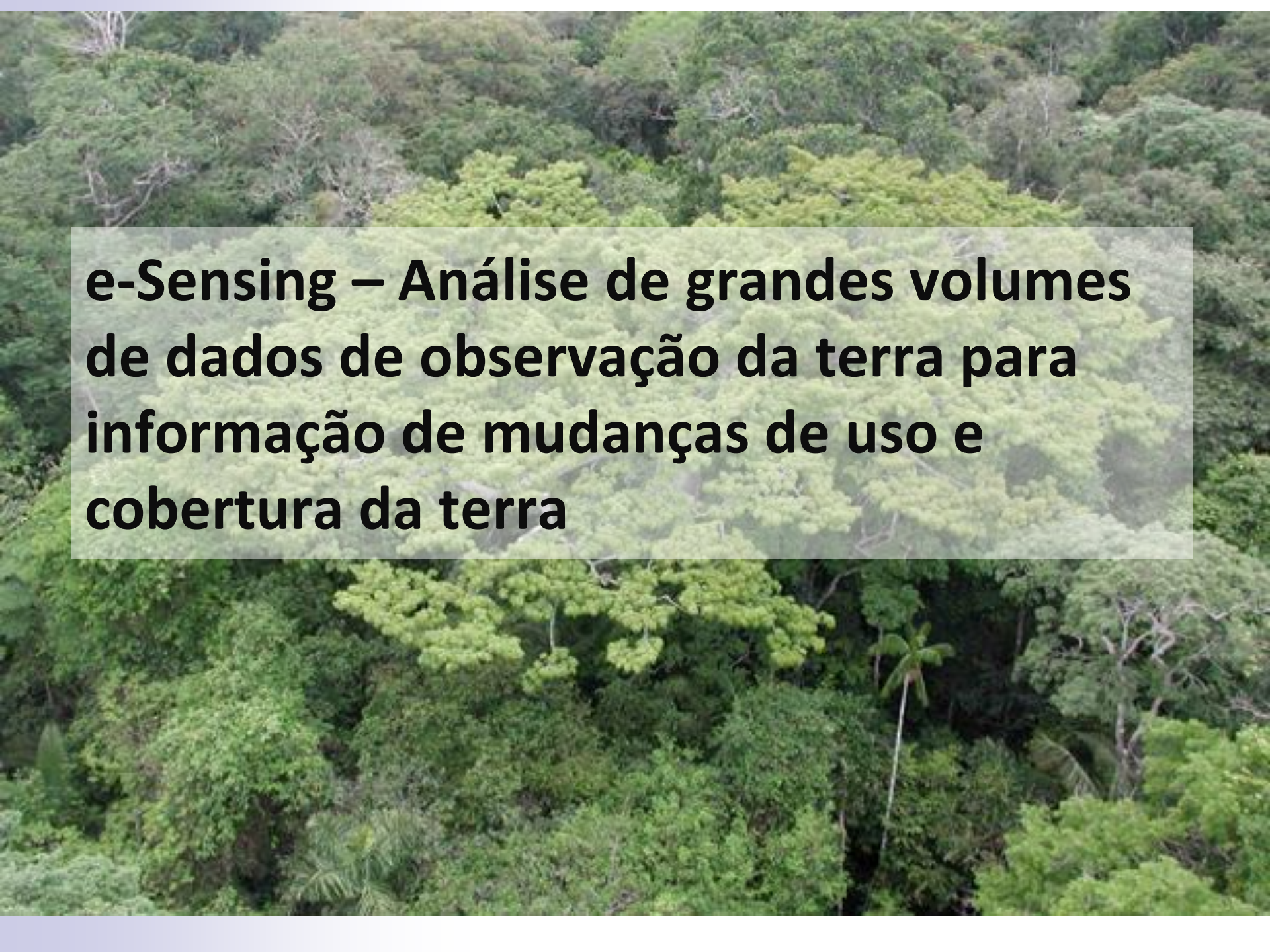


**FAPESP**

Lançamento do 2º edital e workshop do Programa  
FAPESP de Pesquisa em eScience

14/12/2015

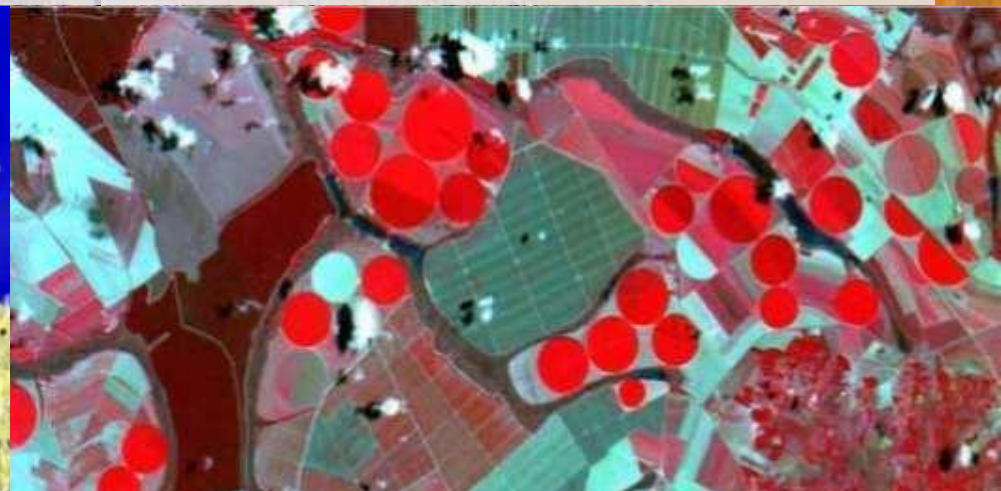
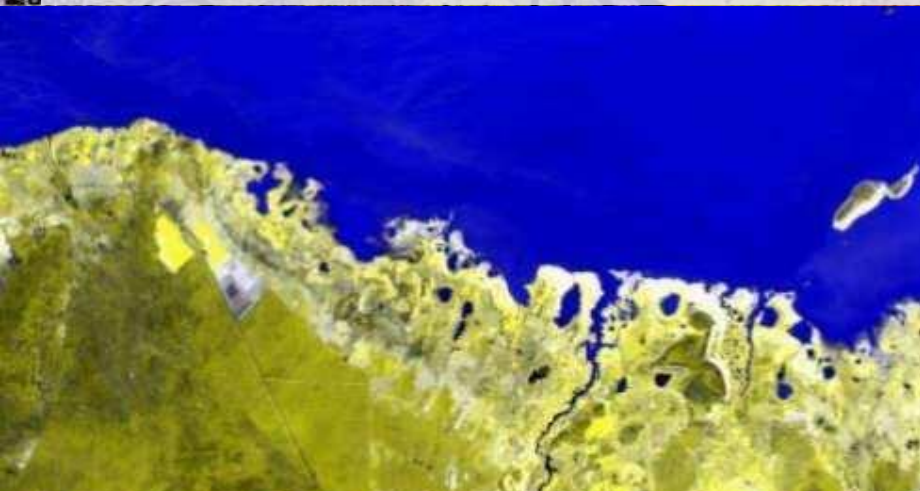
Sentinels + CBERS + LANDSAT + ...: > 10Tb/day

An aerial photograph of a dense tropical forest, showing a variety of green foliage and tree canopies. A semi-transparent white rectangular box is overlaid on the center of the image, containing text. The text is in a bold, black, sans-serif font. The background shows a mix of dark green and lighter green trees, with some taller palm trees visible in the lower right quadrant.

**e-Sensing – Análise de grandes volumes de dados de observação da terra para informação de mudanças de uso e cobertura da terra**



What are we looking for in big EO data?



*"If you don't know where you are going, you'll end up someplace else." (Yogi Berra)*



# Land use

unmanaged forest

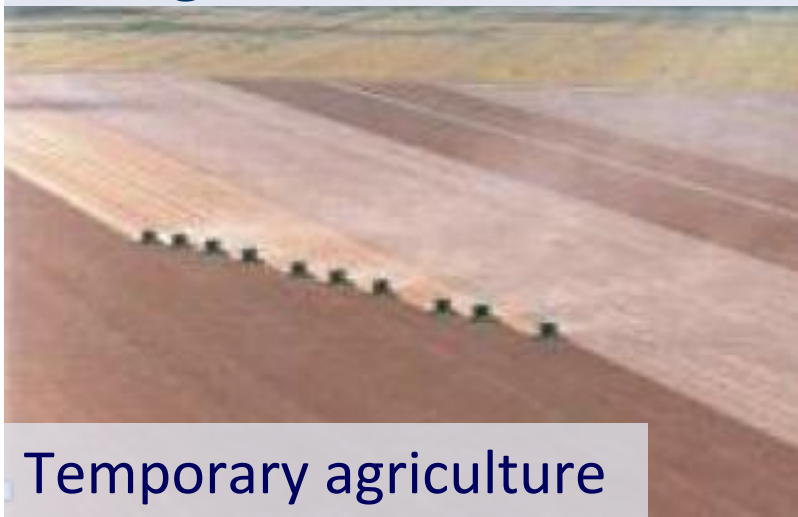


cattle production



“the arrangements, activities and inputs people undertake in a certain land cover type to produce, change or maintain it”

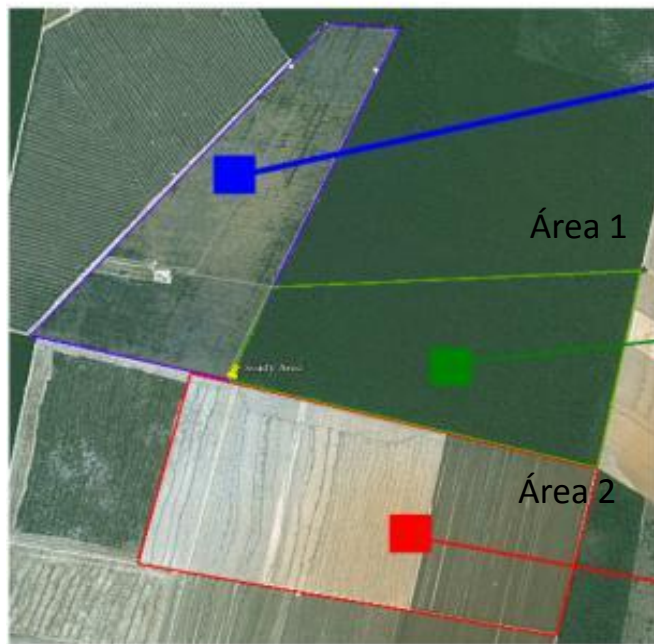
Temporary agriculture



shifting cultivation



# Land trajectories



Área 1

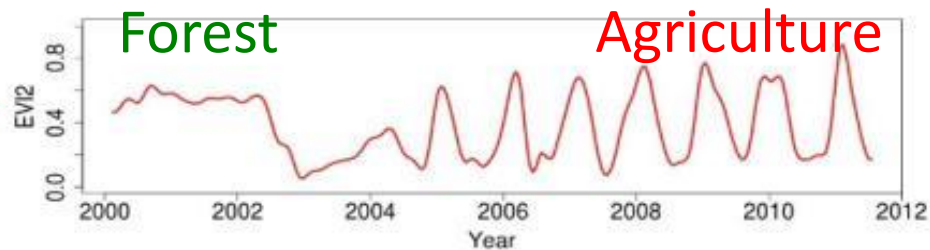
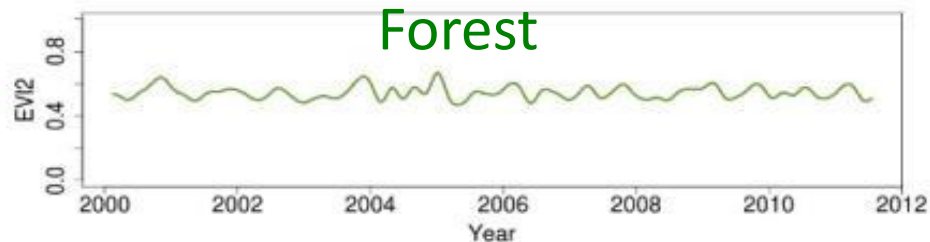
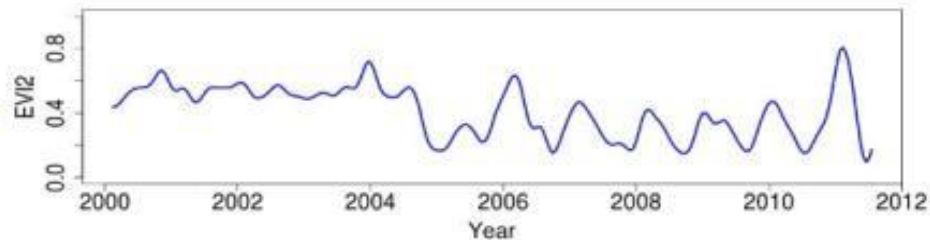
Área 2

Área 3

Forest

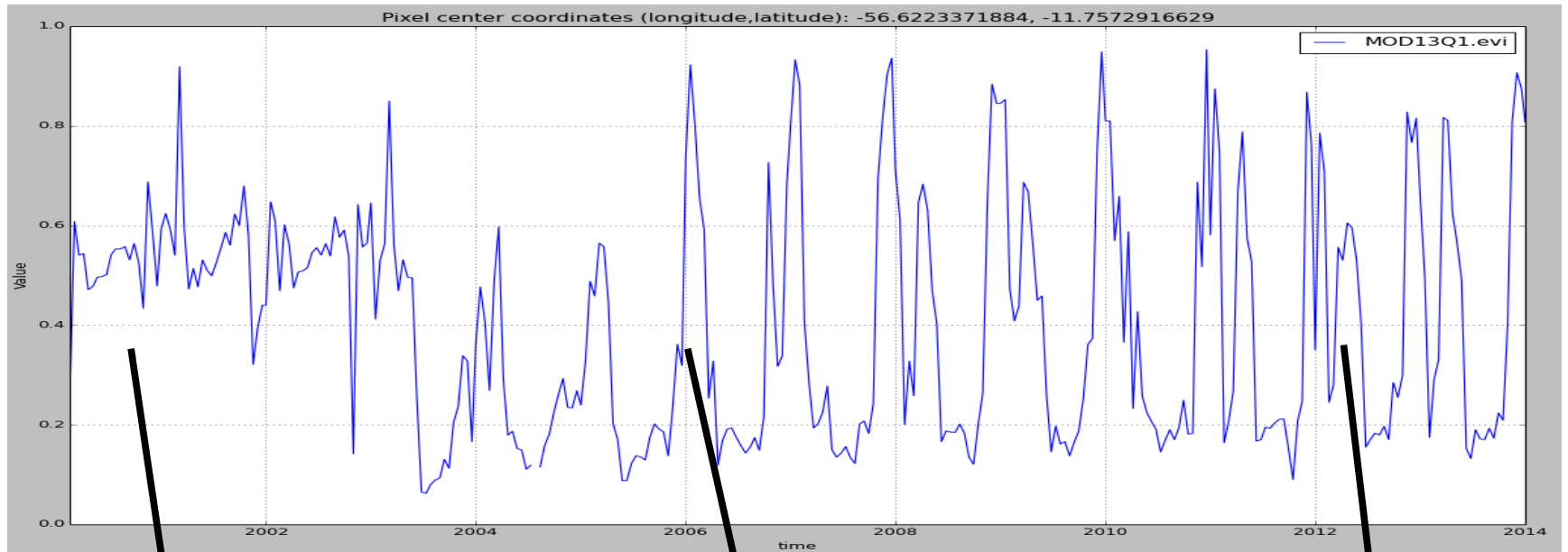
Pasture

Agric

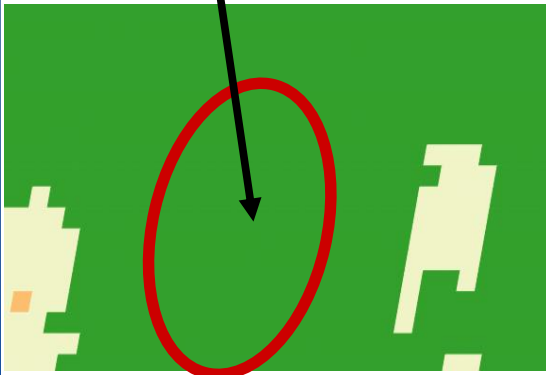


“The transformations of land cover due to actions of land use”

# Land trajectories

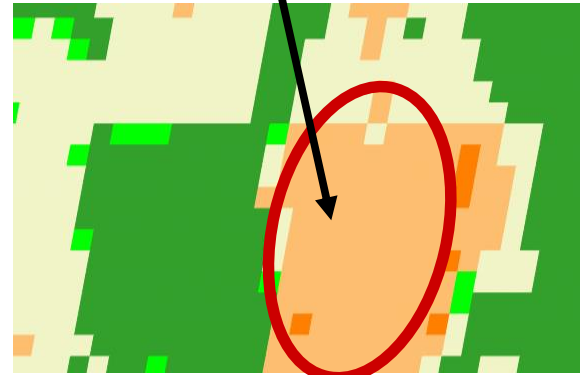


Forest



2001

Single cropping



2006

Double cropping



2013

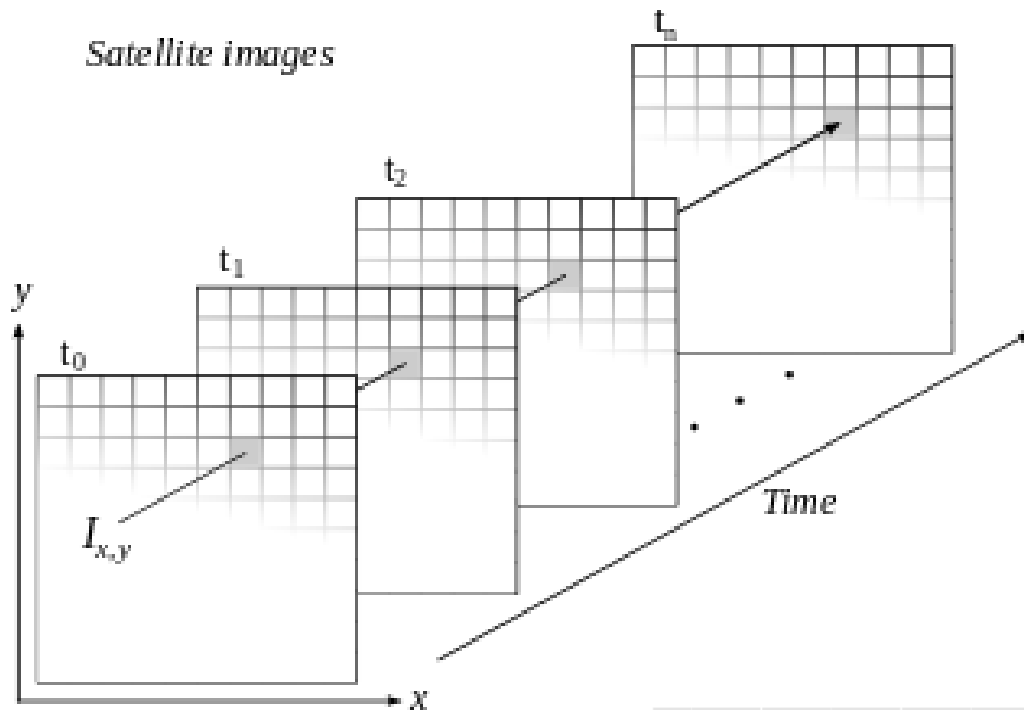


How do we find what we want in big EO data?



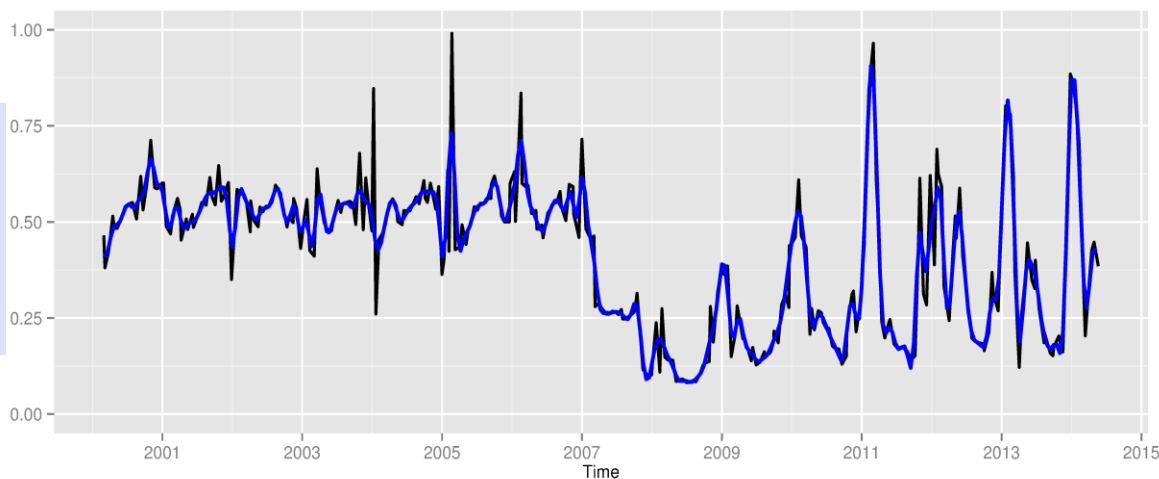
“In theory there is no difference between theory and practice. In practice there is.” (Yogi Berra)

# Space first, time later or time first, space later?



*Space first*: classify images separately  
Compare results in time

*Time first*: classify time series separately  
Join results to get maps





# Single-crop grain production: soybeans



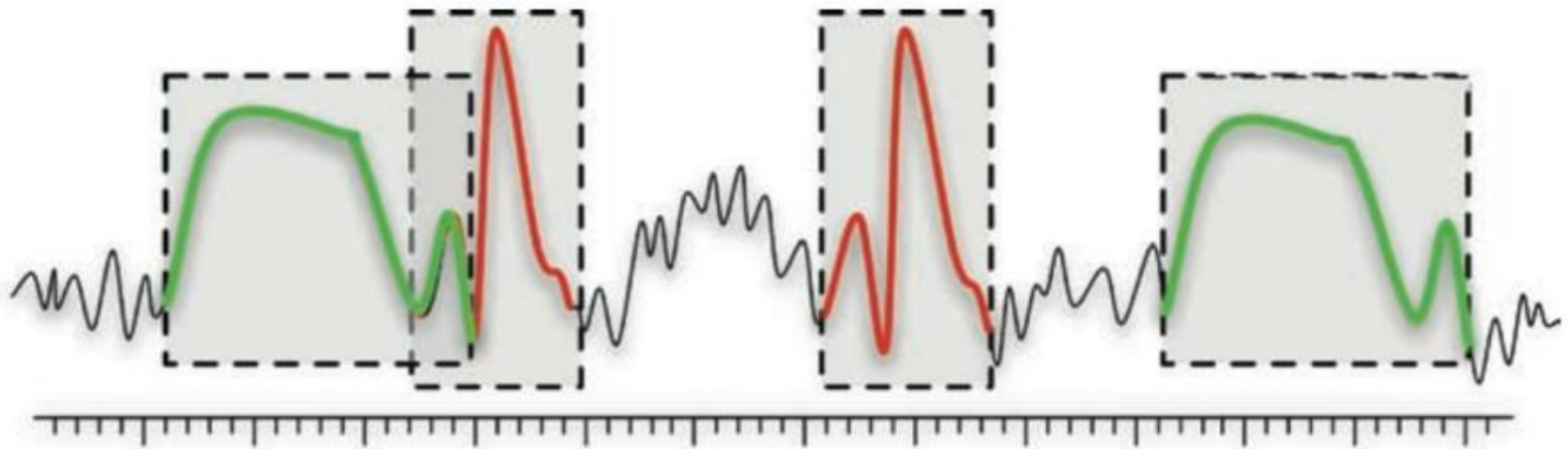
-17.8697, -45.4105  
± × -17.8710, -45.3992  
± × **-17.8710, -45.4104**

One sample per month (?)

graphics: LAF/INPE



# Time series mining: pattern matching



Finding subsequences in a time series  
High computational complexity  
Patterns are idealized, data is noisy

# What is similarity?

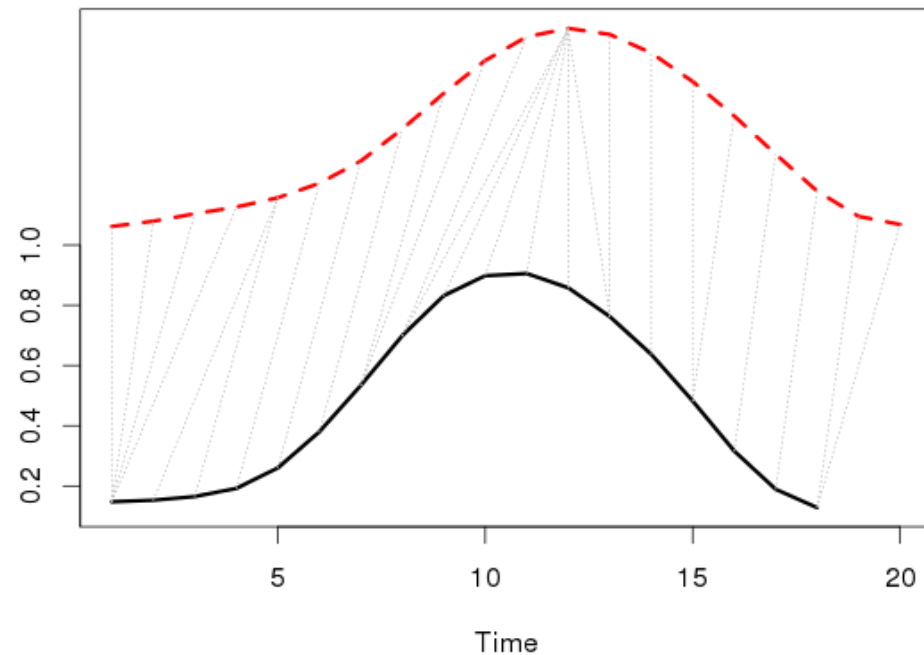
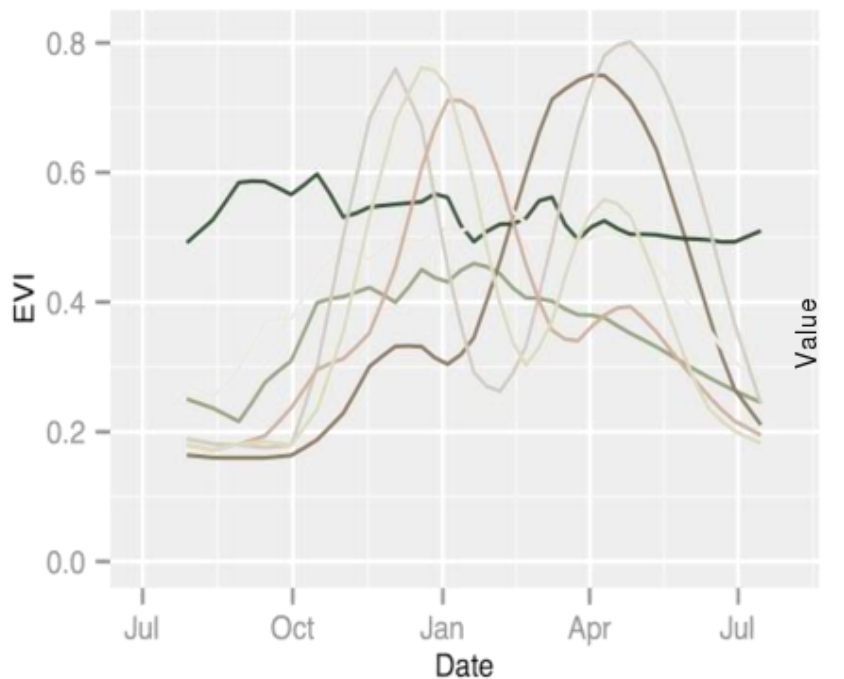


resemblance, likeness, sameness, comparability, correspondence, analogy, parallel, equivalence;



# Dynamic Time Warping: pattern matching

Arvor et al (2012), Eamon Keogh

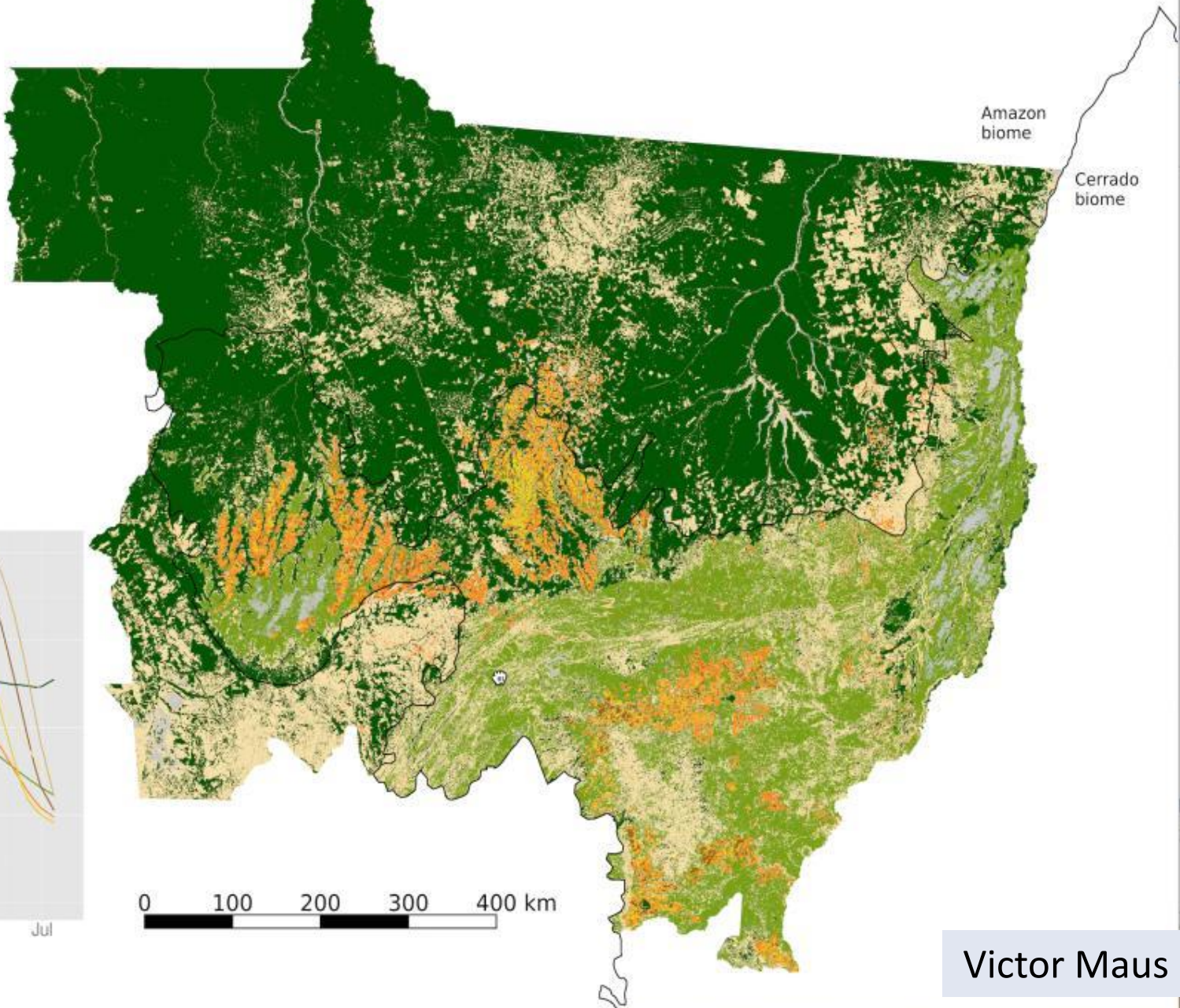


DTW “warps” the time axis: nonlinear matching

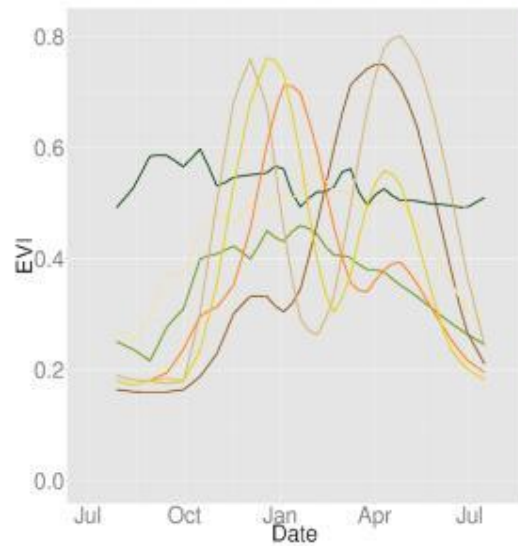
# Mato Grosso, Brazil (Jul/2000 - Jun/2001)

## Land cover/use

- Forest
- Cerrado
- Pasture
- Cotton
- Soybean-cotton
- Soybean
- Soybean-maize
- No-class



## Temporal patterns

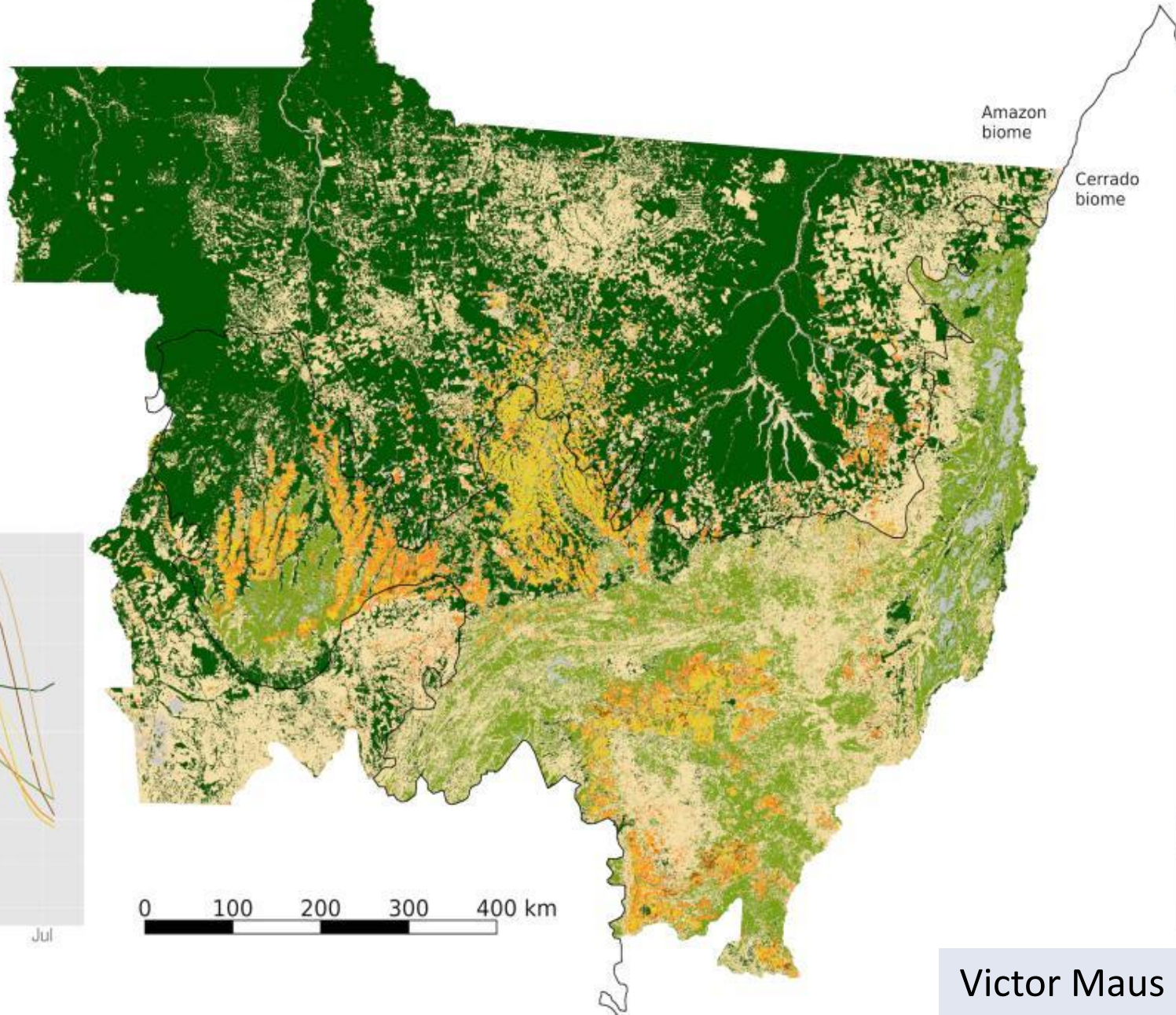


0 100 200 300 400 km

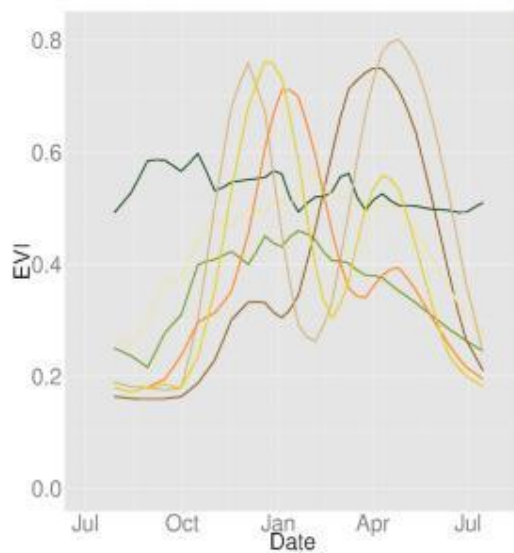
# Mato Grosso, Brazil (Jul/2005 - Jun/2006)

## Land cover/use

- Forest
- Cerrado
- Pasture
- Cotton
- Soybean-cotton
- Soybean
- Soybean-maize
- No-class



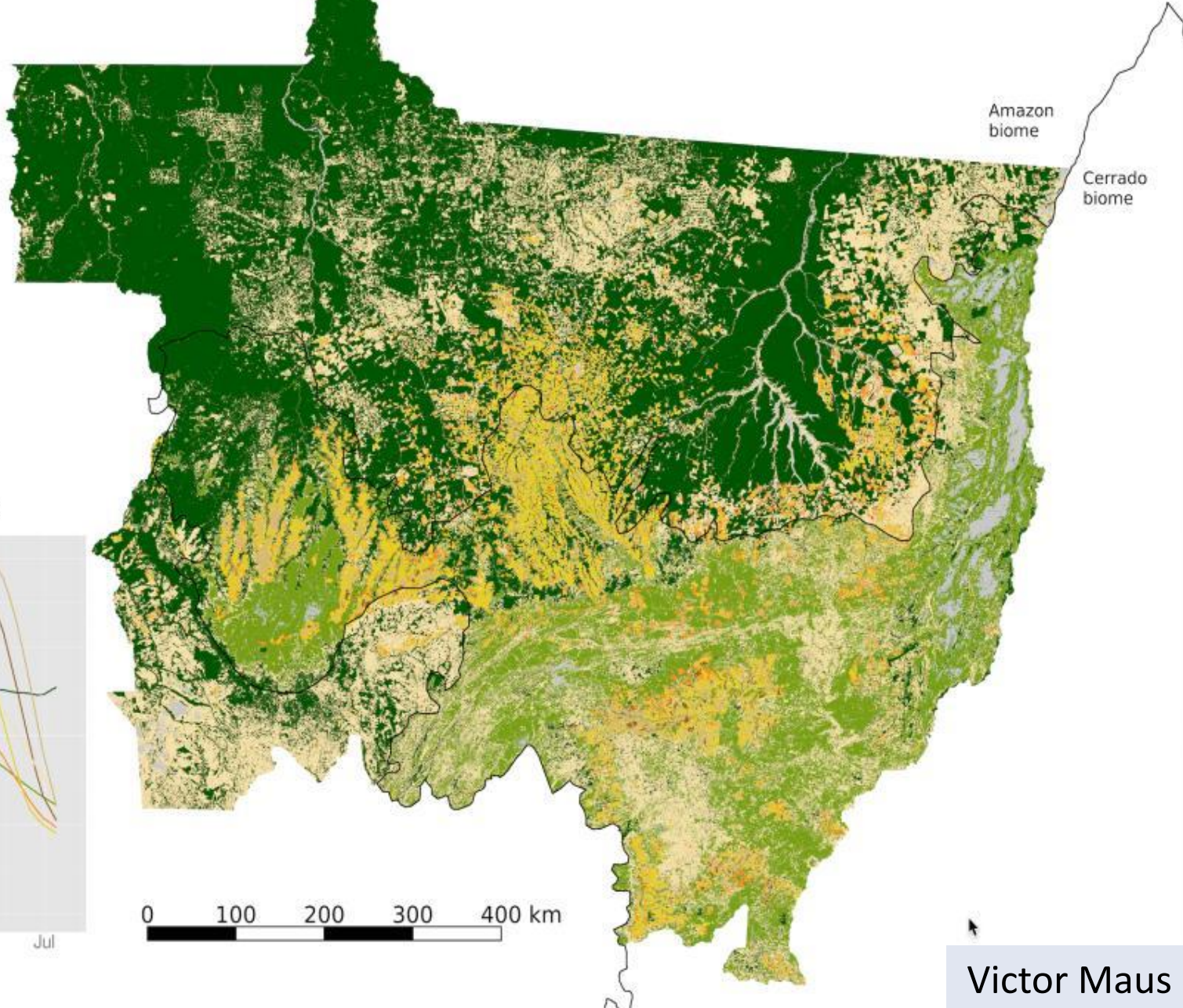
## Temporal patterns



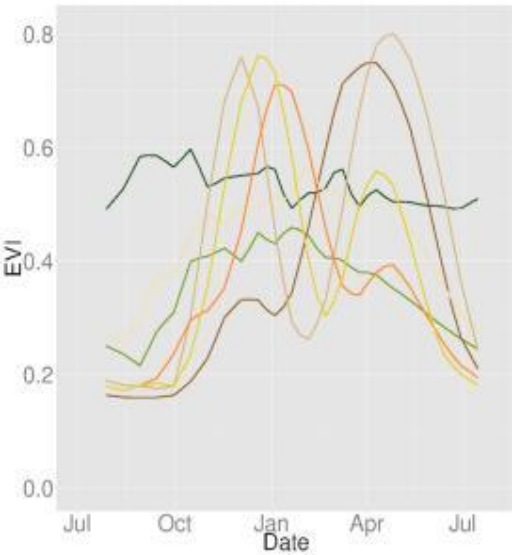
# Mato Grosso, Brazil (Jul/2013 - Jun/2014)

## Land cover/use

- Forest
- Cerrado
- Pasture
- Cotton
- Soybean-cotton
- Soybean
- Soybean-maize
- No-class



## Temporal patterns





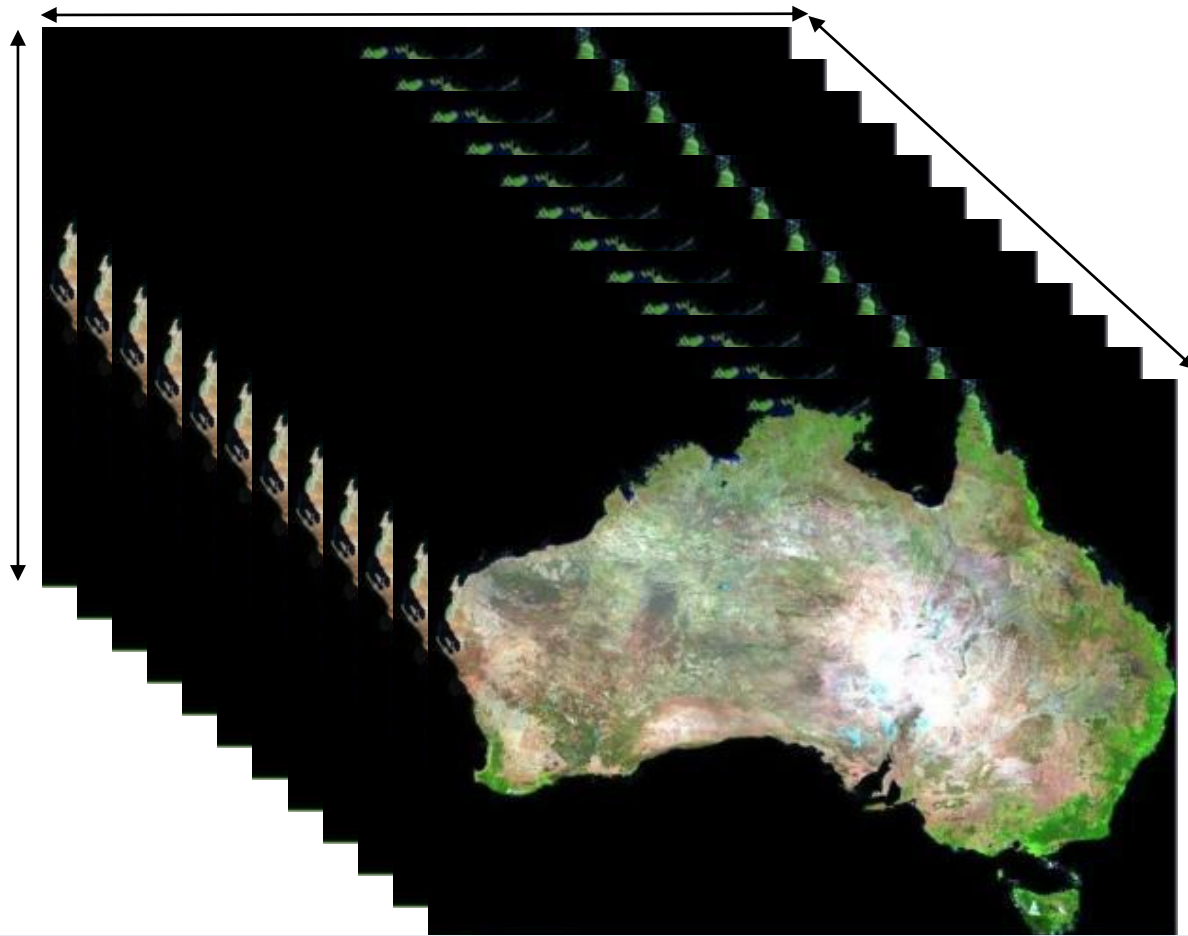
How do we share what we have found  
with big EO data?



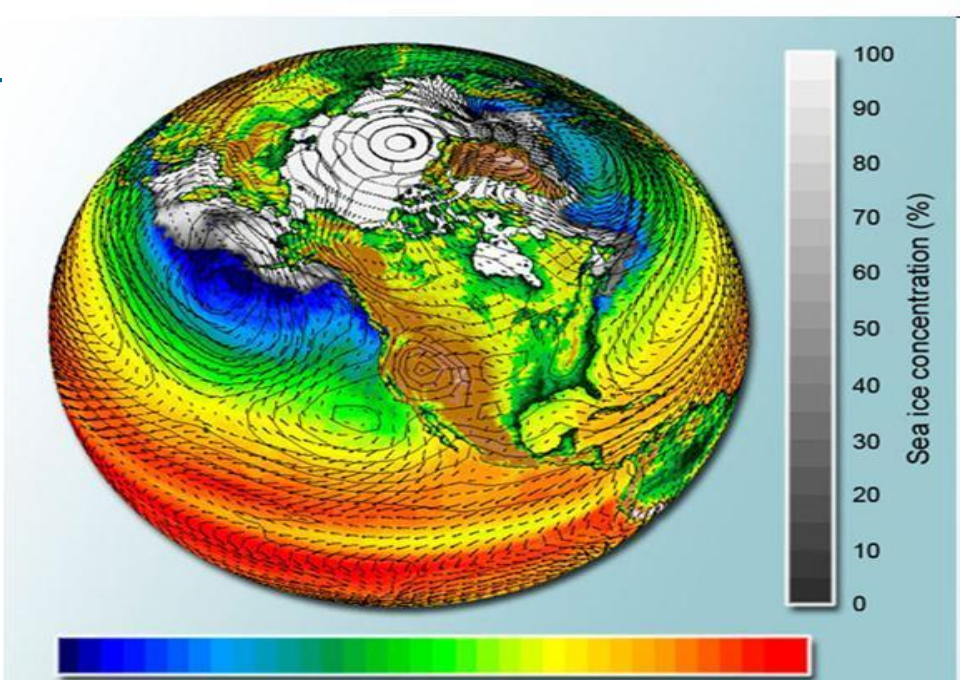
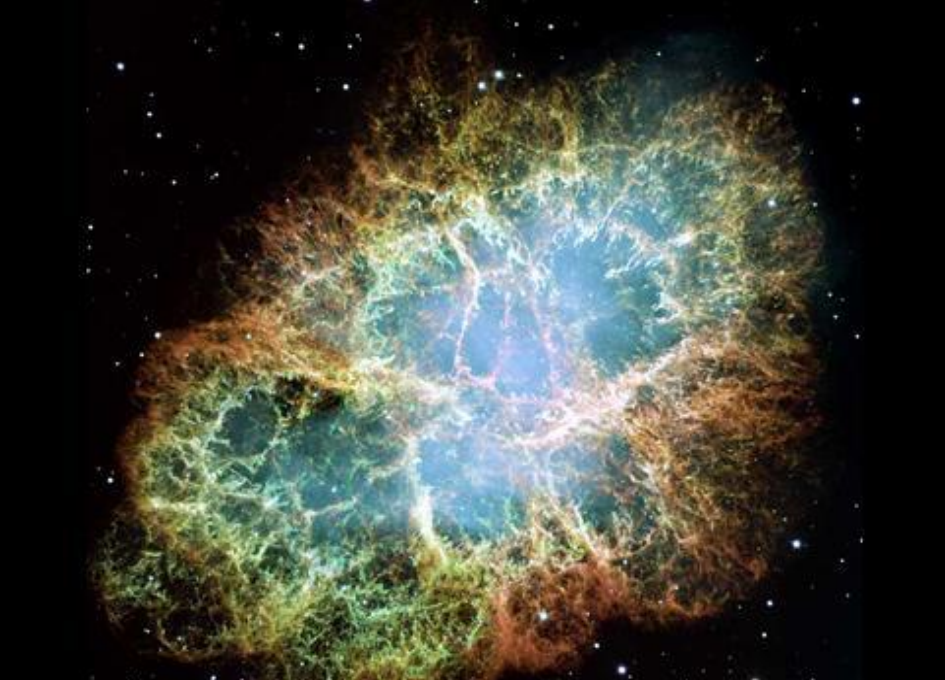
“You have to go to other people’s funerals.  
Otherwise, they won’t go to yours” (Yogi Berra)



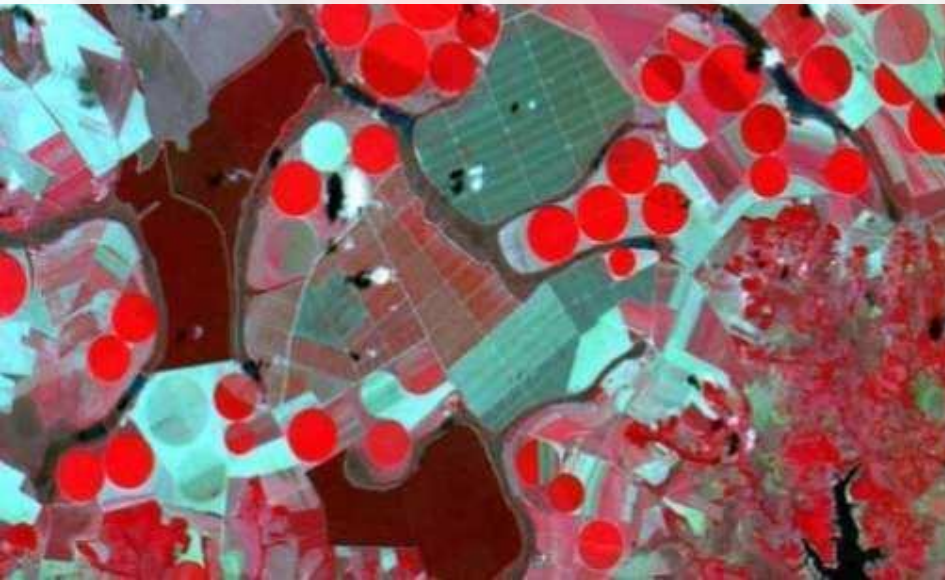
# Big data requires new conceptual views



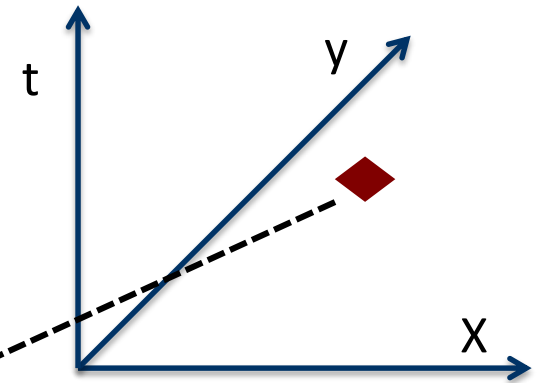
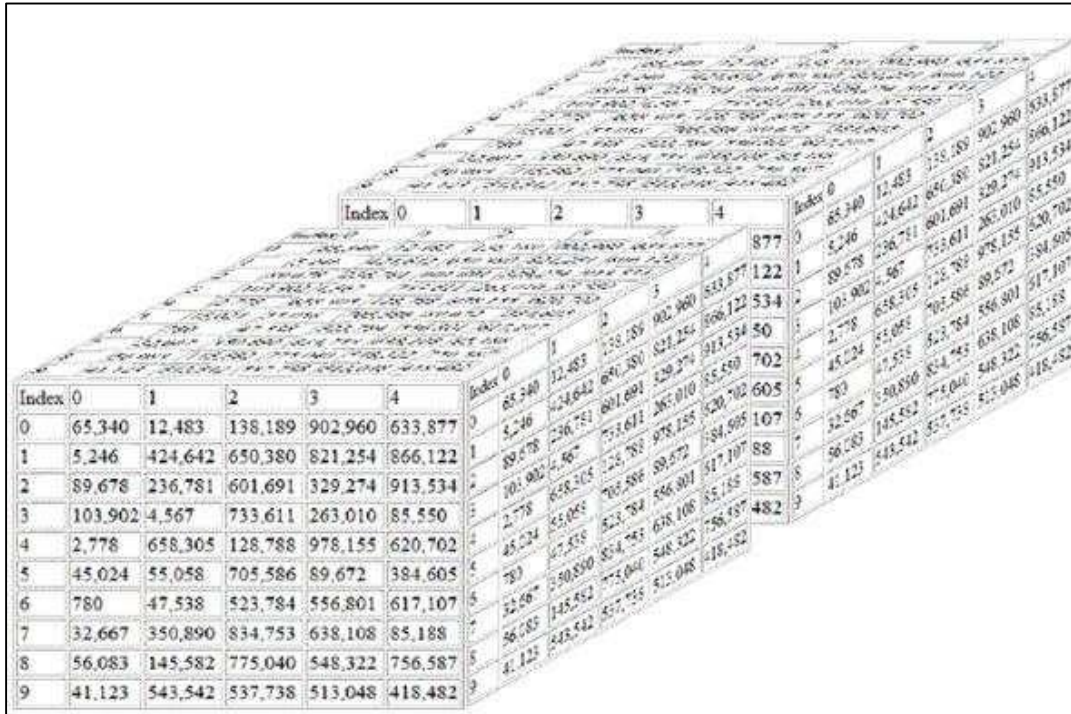
How can we best use the information provided by big data sources?



What do these data have in common?



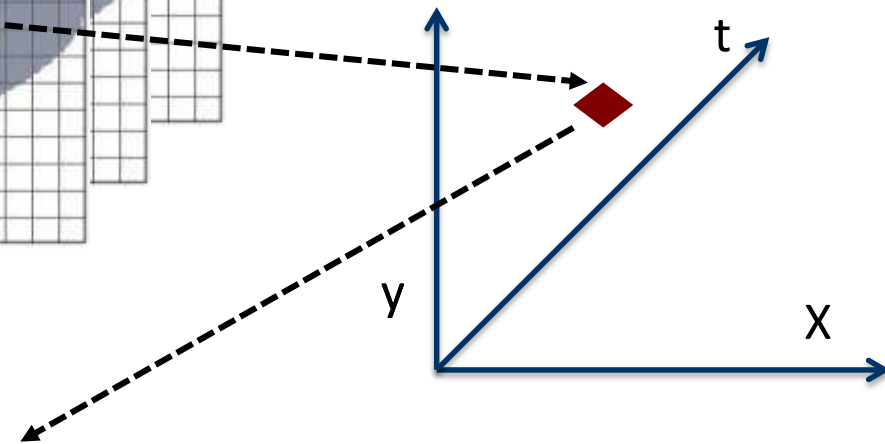
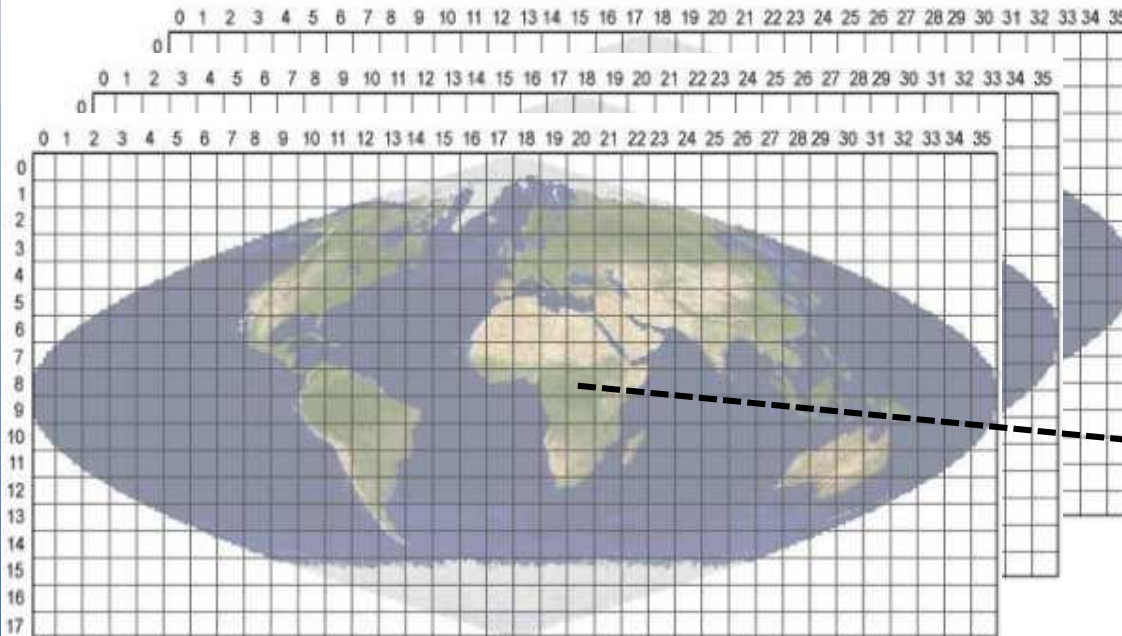
# Scientific data: multidimensional arrays



$$g = f(\langle x, y, z \rangle [a_1, \dots, a_n])$$



# Array databases: all data from a sensor put together in a single array



result = analysis\_function (points in space-time )

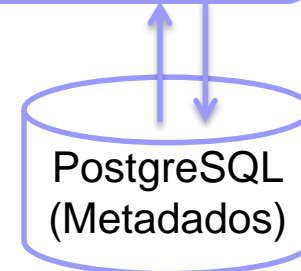
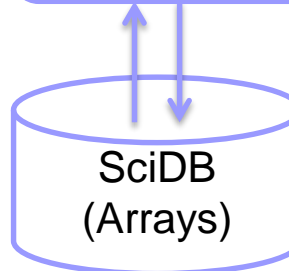
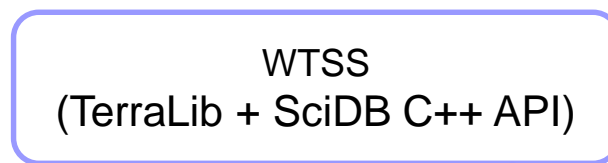


WTSS Client

http://www.dpi.inpe.br/wtss/time\_series?  
coverage=MOD09Q1,attributes=red,nir&  
longitude=-54,latitude=-12&start=2000-02-18&end=2000-03-05



JSON Document



```
{ "result": {
  "attributes": [ { "name": "red",
    "values": [ 1004, 1160, 241 ]
  },
    { "name": "quality",
    "values": [ 4842, 3102, 2116 ]
  }
  ],
  "timeline": [ "2000-02-18", "2000-02-26", "2000-03-05" ],
  "center_coordinates": { "latitude": -11.99, "longitude": -53.99
  }
},
"query": {
  "coverage": "MOD09Q1",
  "attributes": [ "red", "quality" ],
  "latitude": -12,
  "longitude": -54,
  "start": "2000-02-18",
  "end": "2000-03-05"
}
}
```

**WTSS – Web Time Series Service: a lightweight service for serving remote sensing imagery as time series**

Source: Gilberto Queiroz, INPE

# Projeto TerraClass Cerrado - Validador

Parametros da Consulta
   
 Orbita do Landsat: Todas
   
 Ponto do Landsat: Todos
   

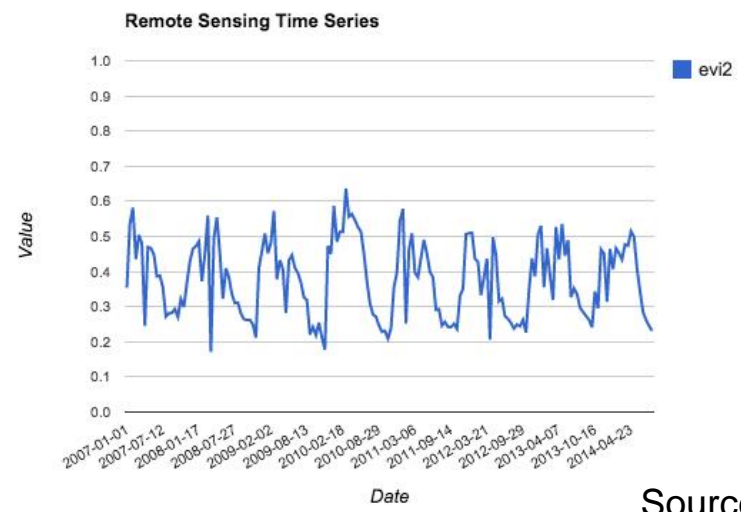
  
 Ajuda...

Lista dos pontos dessa consulta:  
(2400)

| Nr | Lat    | Lon    | UF | Cena L8 |
|----|--------|--------|----|---------|
| 1  | -11.20 | -44.65 | BA | 220/68  |
| 2  | -17.29 | -44.58 | MG | 219/72  |
| 3  | -20.67 | -54.60 | MS | 225/74  |
| 4  | -19.13 | -52.73 | MS | 224/73  |
| 5  | -17.68 | -53.55 | MT | 224/72  |
| 6  | -16.47 | -48.25 | GO | 221/71  |
| 7  | -10.51 | -50.58 | MT | 223/67  |
| 8  | -4.35  | -44.80 | MA | 221/63  |
| 9  | -15.90 | -43.27 | MG | 218/71  |
| 10 | -19.39 | -46.96 | MG | 220/73  |
| 11 | -9.83  | -44.16 | PI | 219/67  |
| 12 | -11.48 | -46.89 | TO | 221/68  |
| 13 | -16.86 | -48.05 | GO | 221/72  |
| 14 | -19.06 | -54.56 | MS | 225/73  |
| 15 | -6.26  | -46.84 | MA | 222/64  |
| 16 | -17.94 | -54.37 | MS | 225/72  |
| 17 | -16.68 | -50.02 | GO | 222/72  |
| 18 | -19.05 | -54.77 | MS | 225/73  |
| 19 | -15.63 | -55.21 | MT | 226/71  |
| 20 | -16.98 | -52.79 | GO | 224/72  |
| 21 | -16.90 | -45.15 | MG | 219/72  |
| 22 | -10.03 | -50.30 | MT | 223/67  |
| 23 | -13.36 | -57.53 | MT | 228/69  |
| 24 | -3.69  | -42.46 | PI | 219/63  |
| 25 | -15.55 | -51.80 | MT | 224/71  |
| 26 | -17.18 | -43.17 | MG | 218/72  |
| 27 | -13.21 | -55.08 | MT | 226/69  |
| 28 | -16.19 | -48.79 | GO | 222/72  |

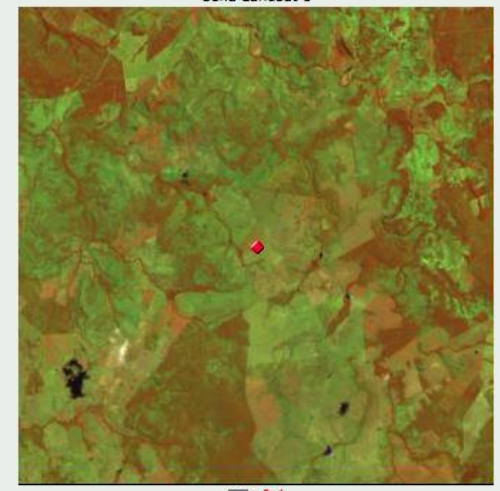


Modis Times Series

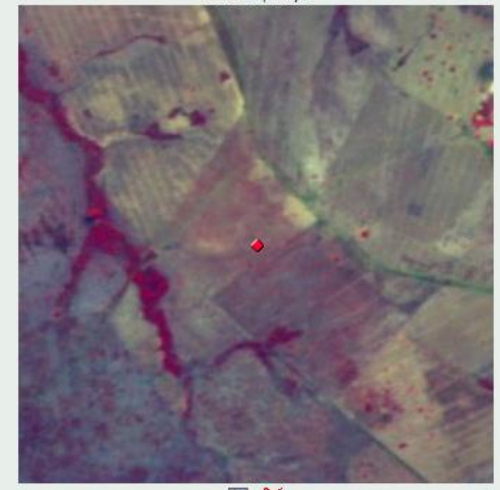


## Classifica Ponto (15)

Cena Landsat 8

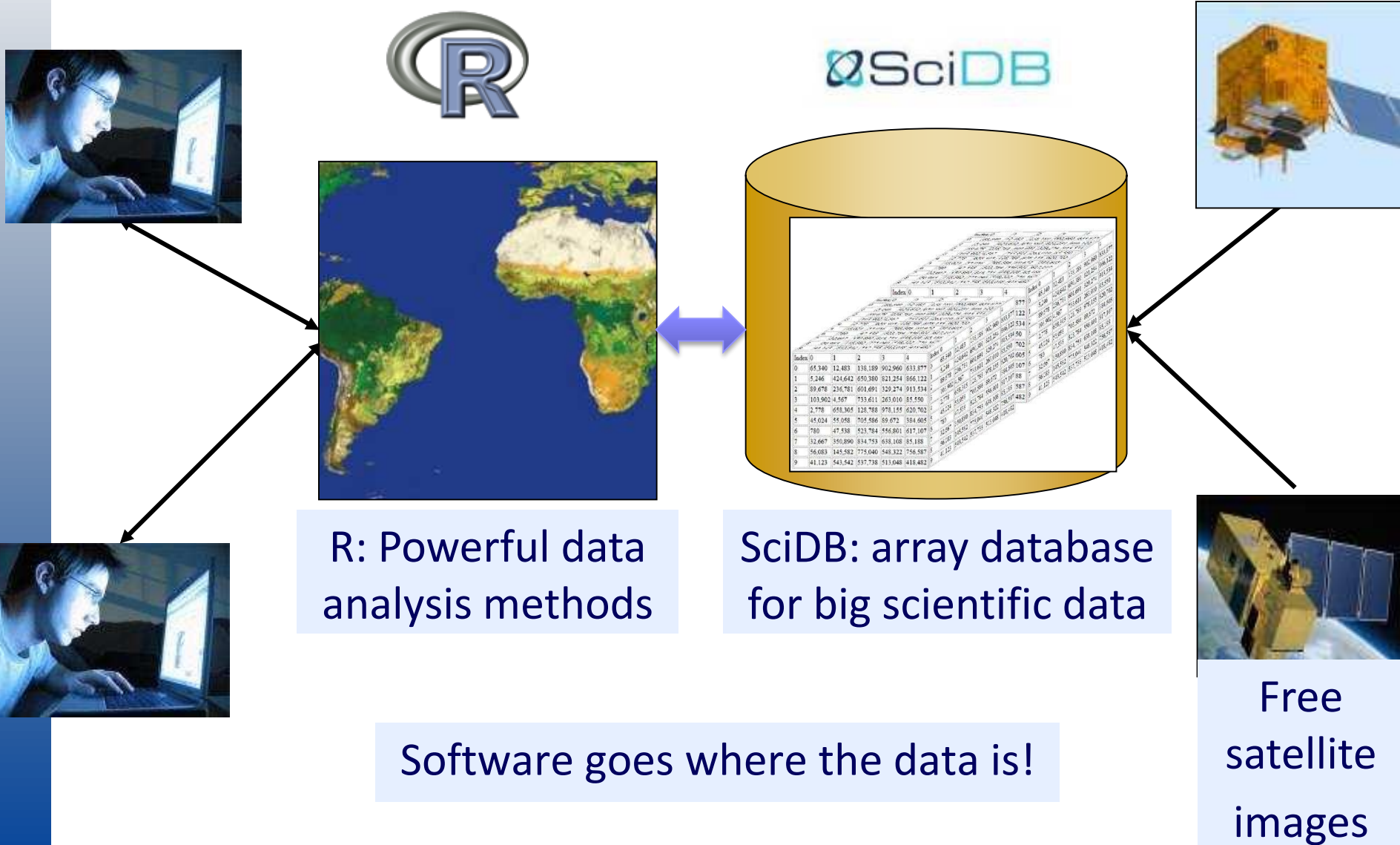


Cena RapidEye



Ponto nr 15 (-6.2552,-46.8354)  
 Area ha 17.18  
 Municipio/UF Porto Franco/MA  
 Florestal

# Global Land Observatory: describing change in a connected world





# Status do projeto

- Infraestrutura – Cluster 96 TB (5 servidores)
- Pessoal
  - 2 PD (em contratação)
  - 2 DR (selecionados)
  - 2 TT4 (1 contratado, 1 início 2016)
- Artigos científicos
  - 5 artigos em conferência apresentados
  - 1 artigo aceito (IEEE JSTARS)
  - 1 submetido (Journal of Statistical Software)





## Contrapartida institucional

- 2 PI (Gilberto Câmara, Leila Fonseca)
- 7 Doutores ~ 20% tempo
- 1 CLT contratado Fundo Amazônia
- 1 Contratado Univ Münster