

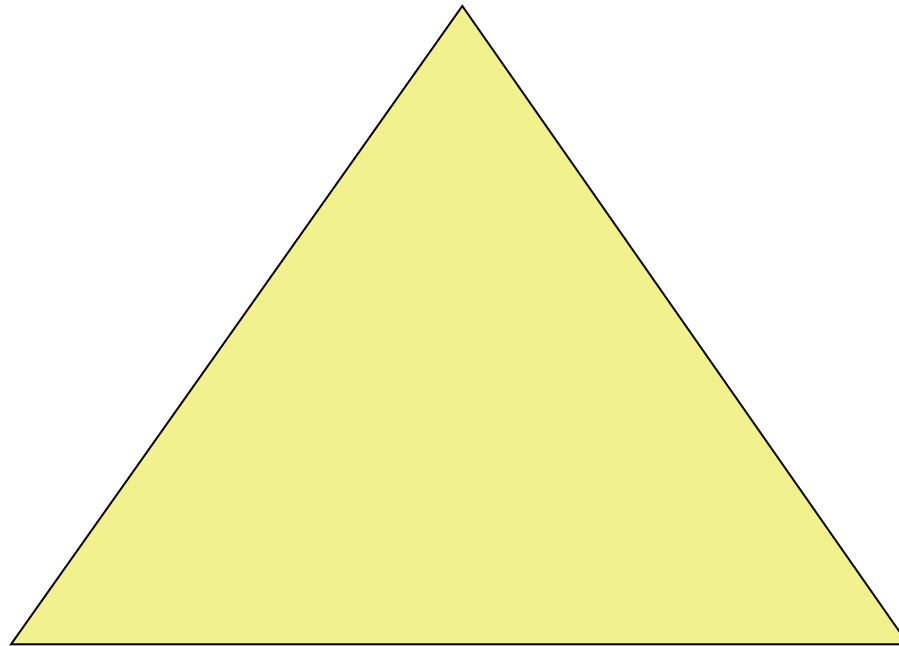
# Best Practices on Open Data: the Brazilian experience

Gilberto Câmara (INPE, Brazil)  
Belmont Forum, co-chair

# Foundations of modern democracies

Fukuyama (2011)

State (*power by technical staff*)



Rule of law  
(*limits power*)

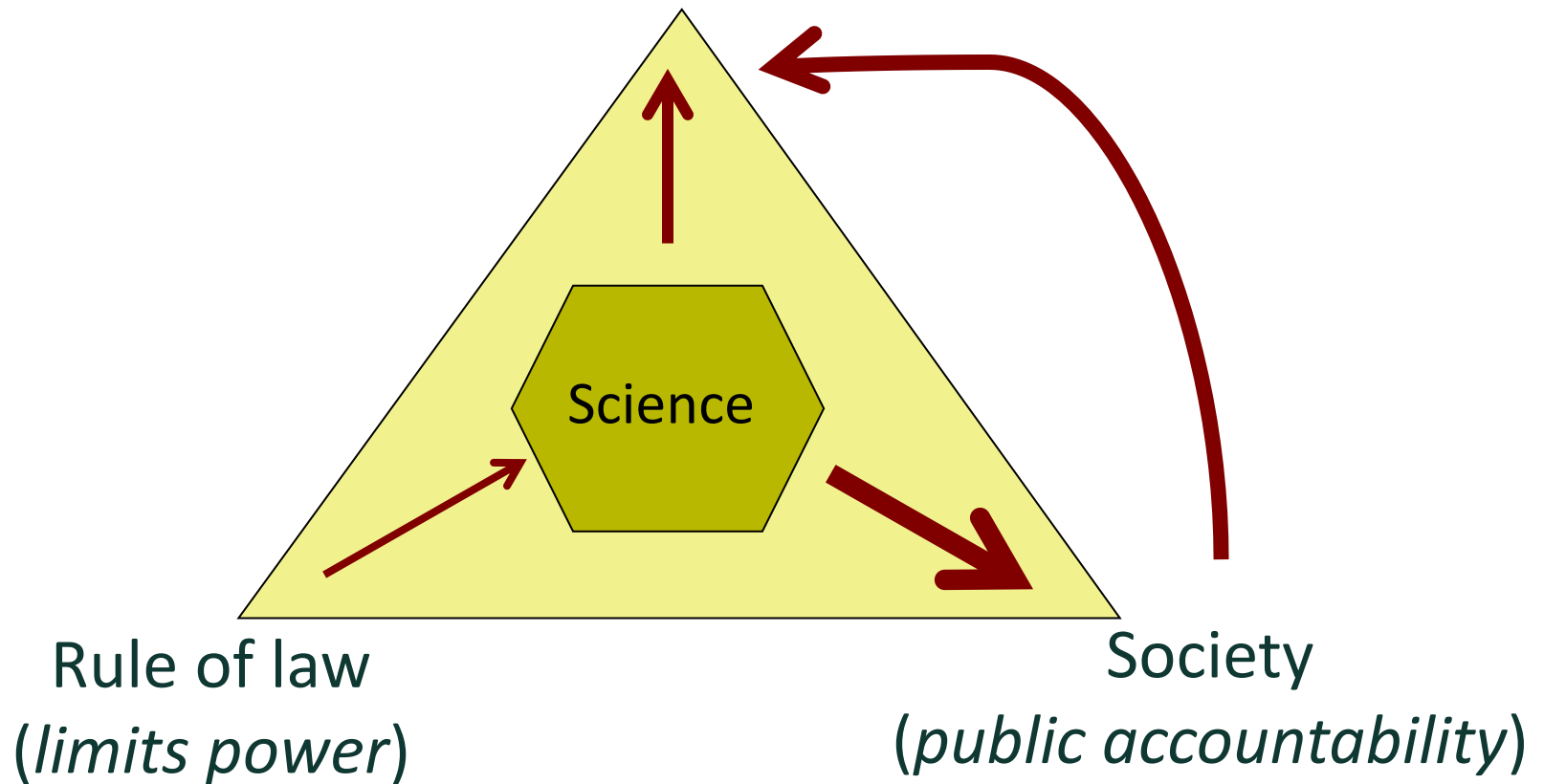
Society  
(*public accountability*)

How is the interplay between science,  
public opinion and government?

# Foundations of modern democracies

Fukuyama (2011)

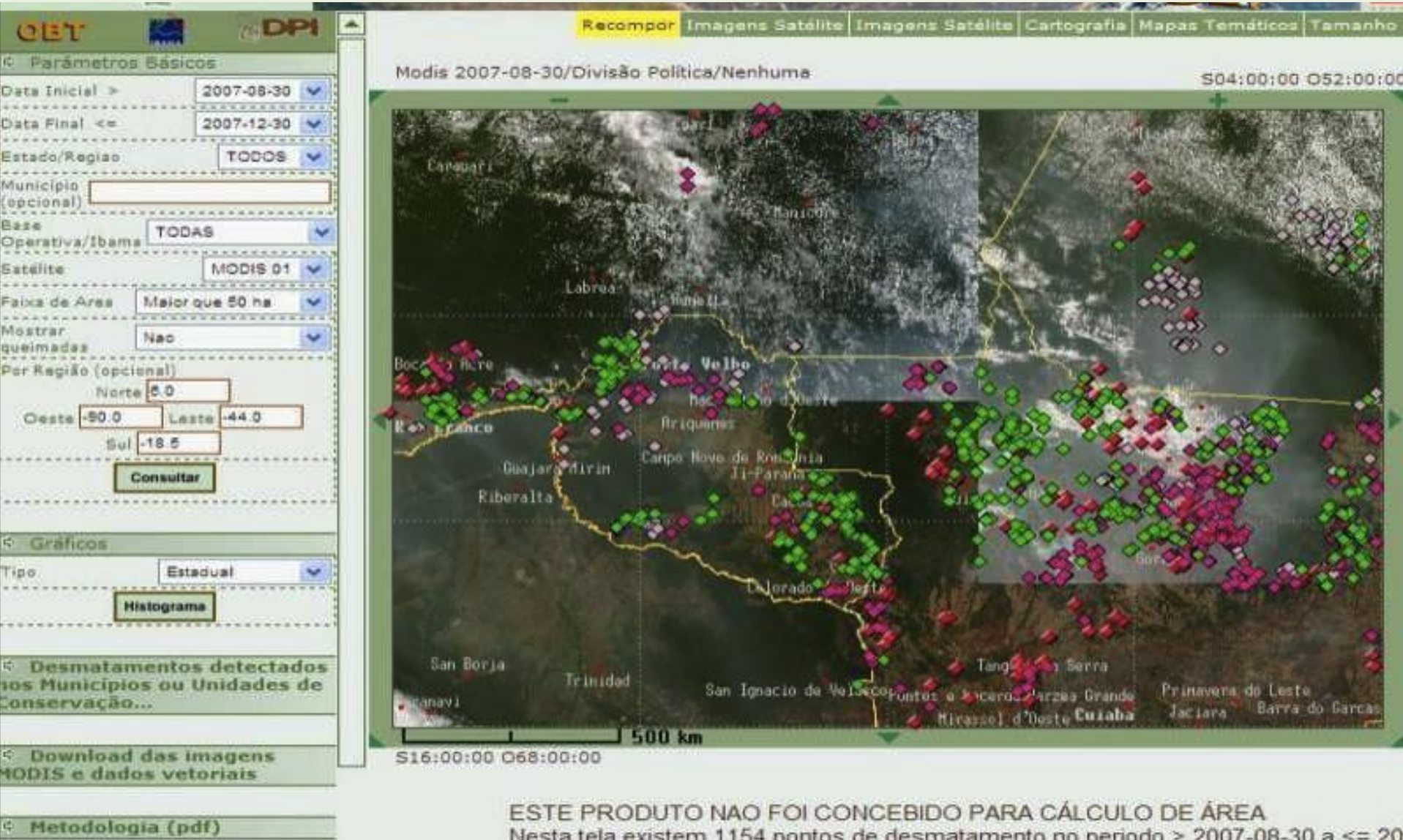
State (*power by technical staff*)



How is the interplay between science,  
public opinion and government?

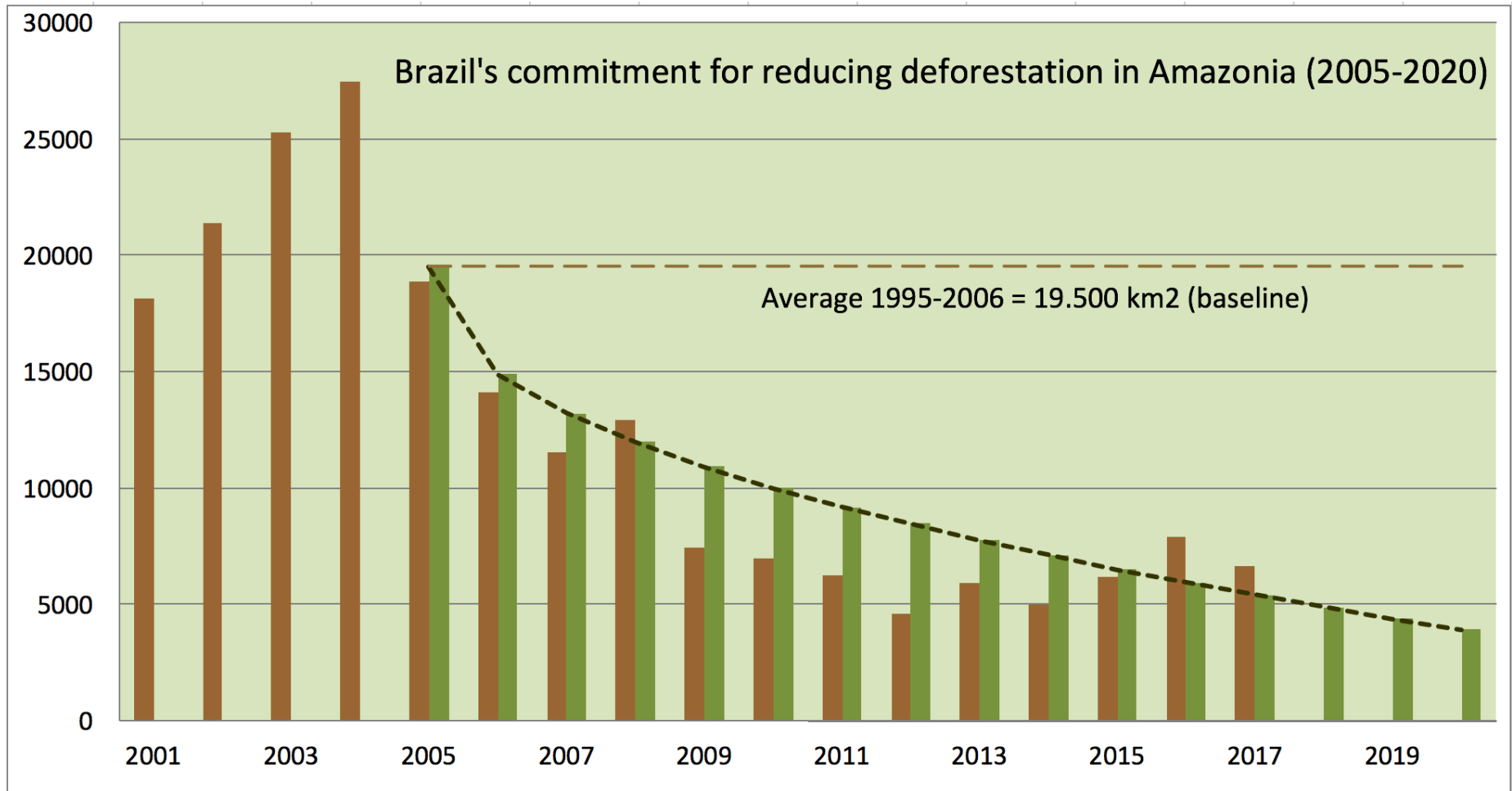
“A few satellites can cover the entire globe, but there needs to be a system in place to ensure their images are readily available to everyone who needs them. Brazil has set an important precedent by making its Earth-observation data available, and the rest of the world should follow suit.”

# Real-time Deforestation Monitoring



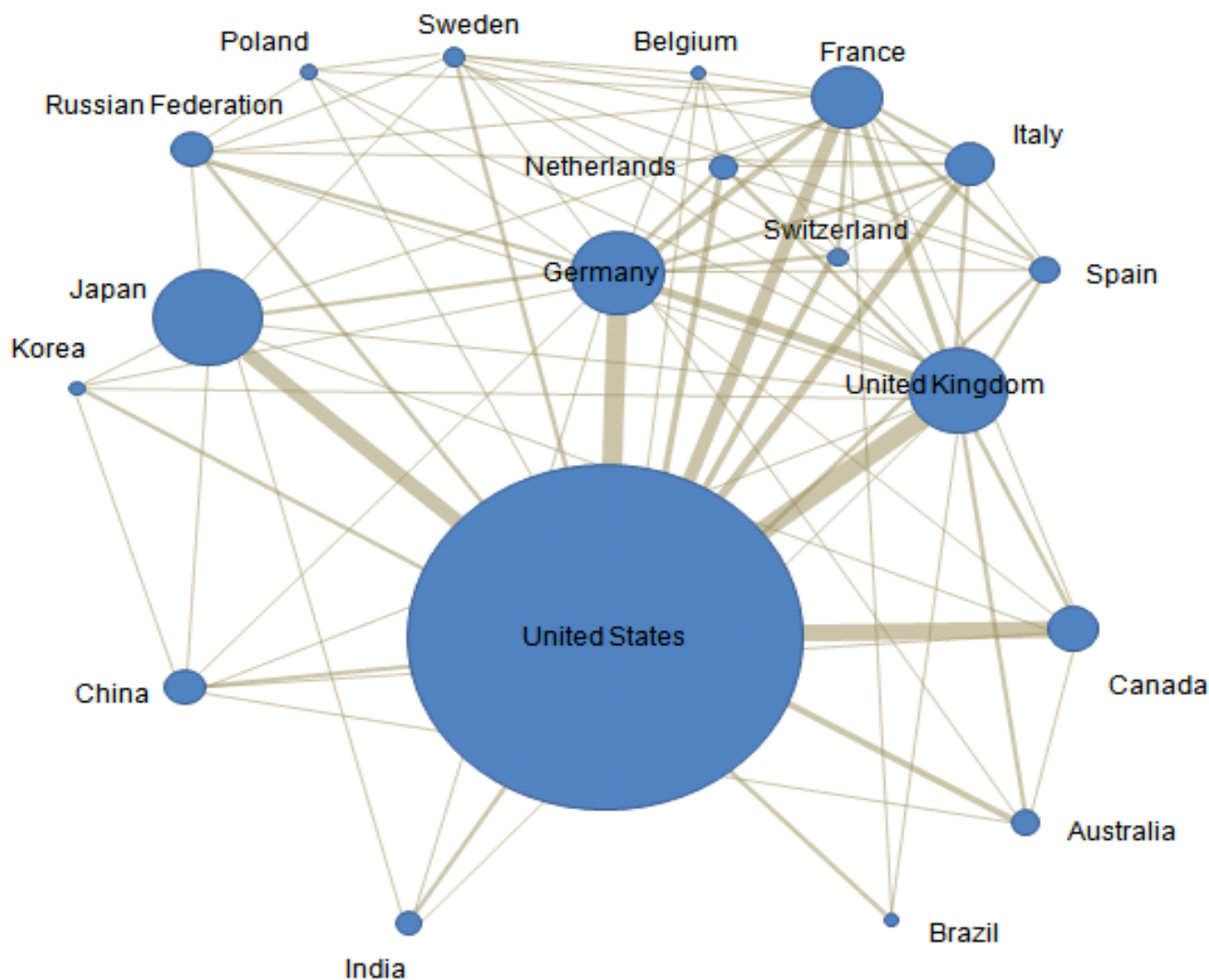
Daily warnings of newly deforested large areas

# Transparency builds governance!



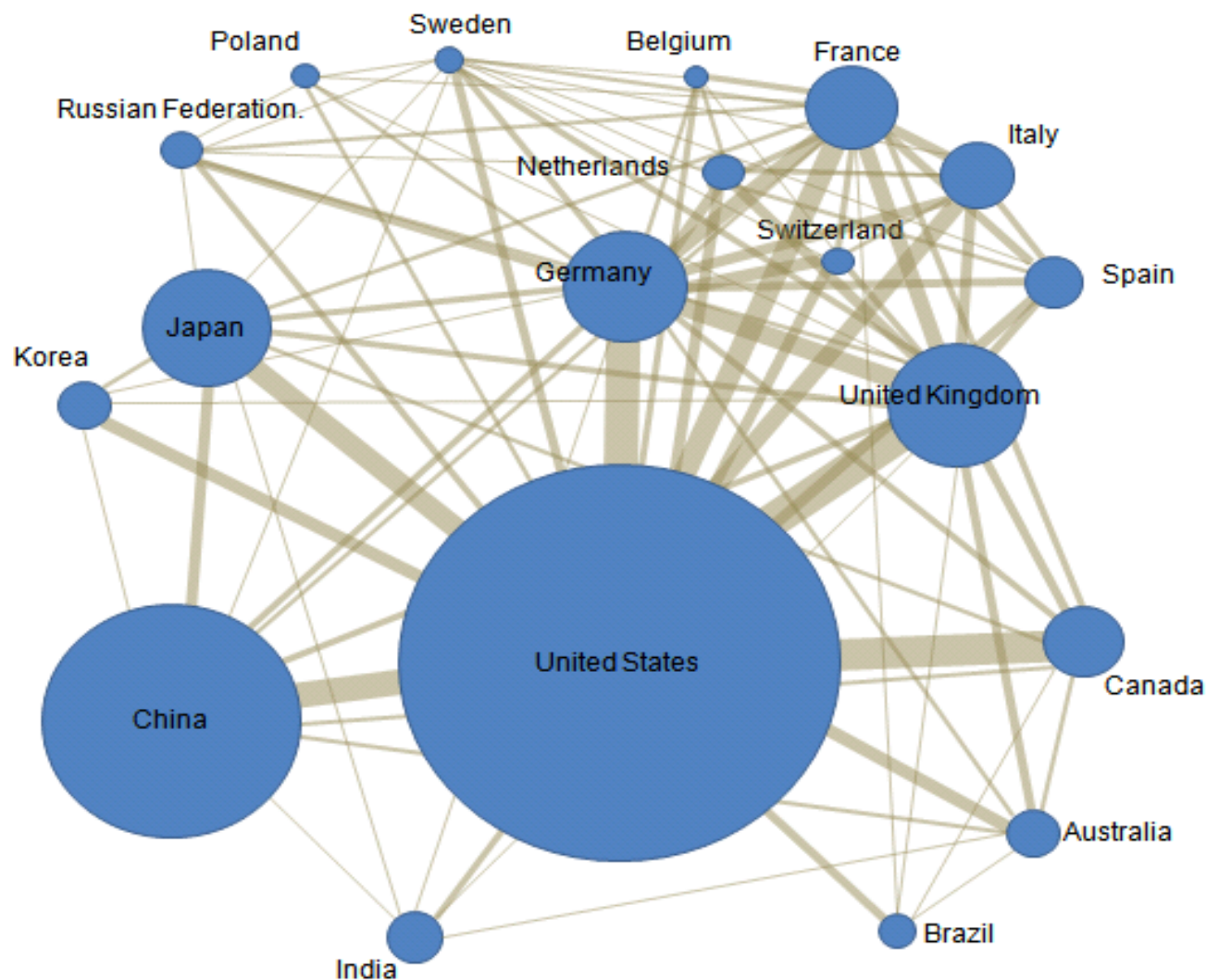
Deforestation in Amazonia (2001-2017): 27,000 km<sup>2</sup> to 6,200 km<sup>2</sup>

# World Science (1998)



OECD

# World Science (2008)



OECD



# SCIENTIFIC DATA



Altmetric: 439

Views: 27,909

Citations: 26

[More detail >>](#)

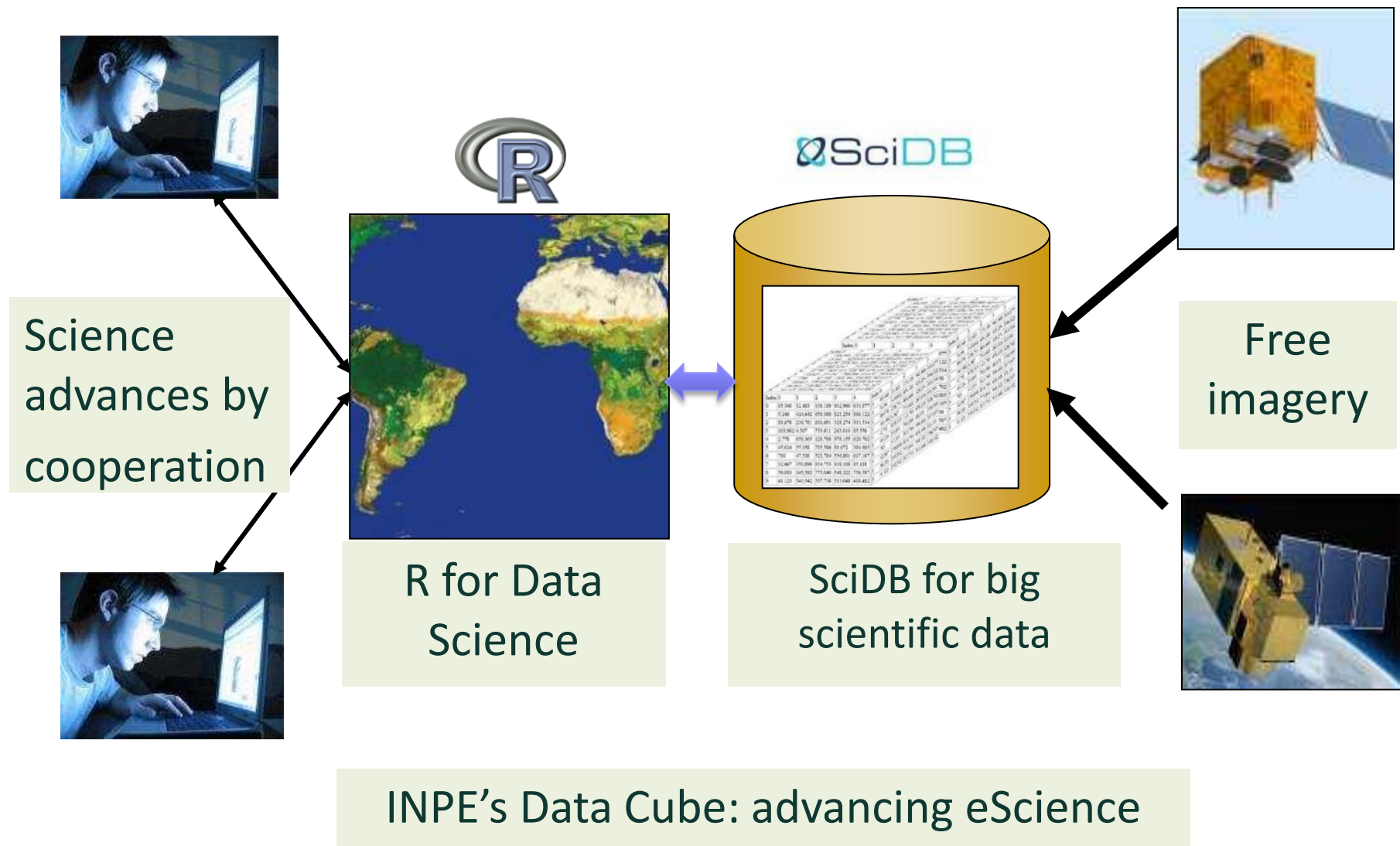
Comment | [OPEN](#)

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons

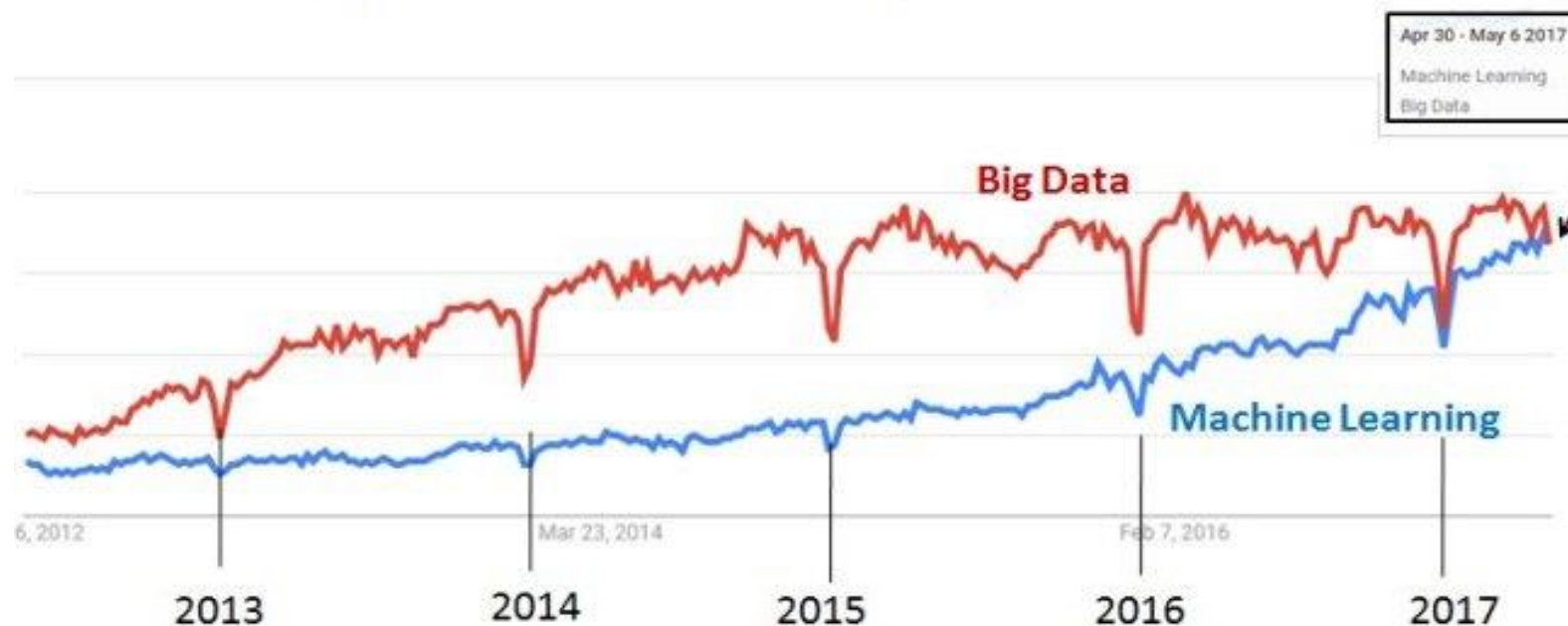
Findable, Accessible, Interoperable, Reproducible

# Hard problems need Science!



# Machine learning with big data

Google Trends, May 2012 - April 2017  
Big Data vs Machine Learning search terms



Combine SVM, Random forest, Deep learning, Neural nets with satellite image time series

# From User Requirements to Data Architectures

## Needs of research community

*Analytical scaling*: algorithms developed at the desktop run on big databases

*Collaborative work*: share results with the scientific community.

*Replication*: research teams can build their own infrastructure.

## System architecture choices

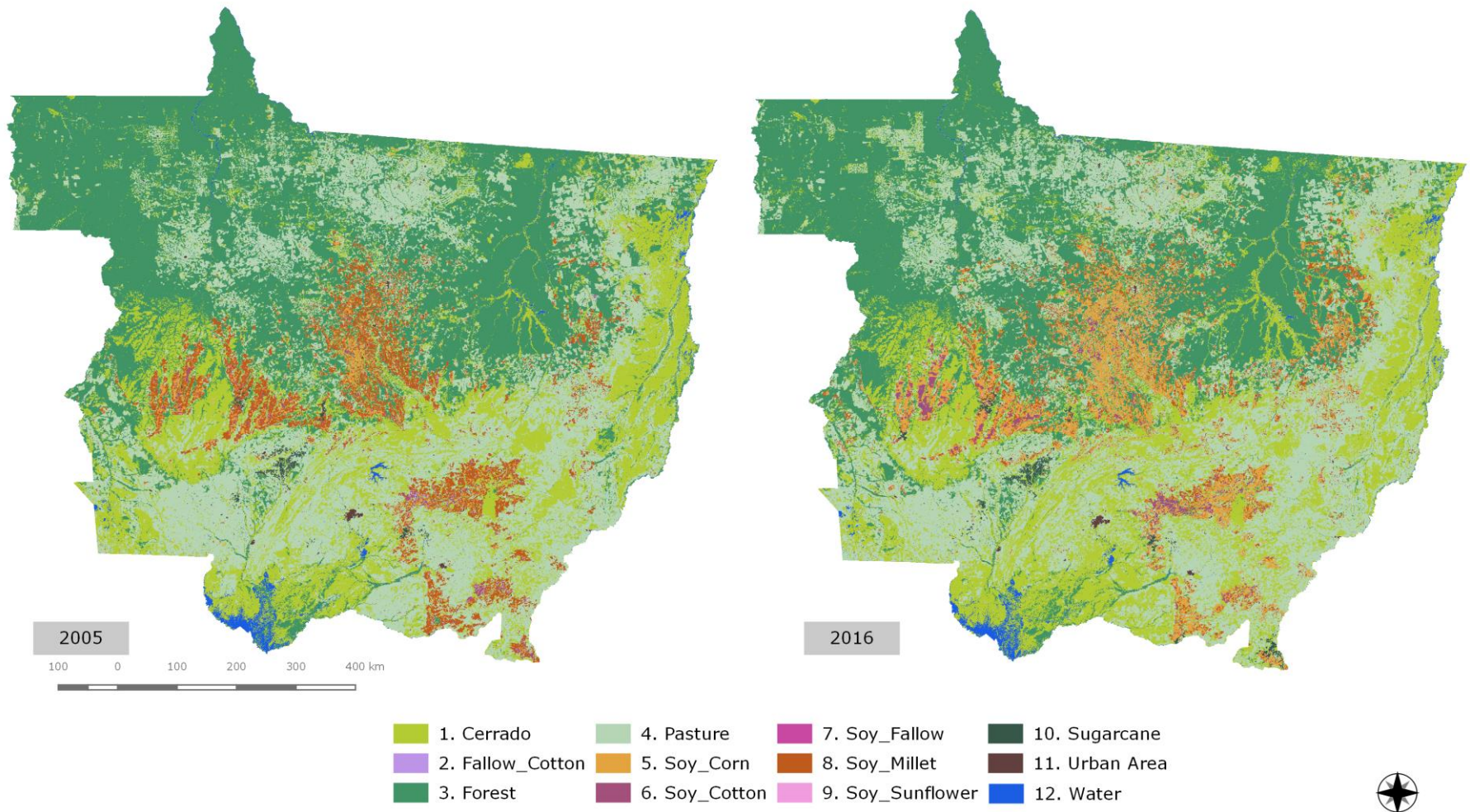


State-of-the-art statistical tools for Data Science



Data management for scientific data

# Mato Grosso – Brazil's agricultural frontier

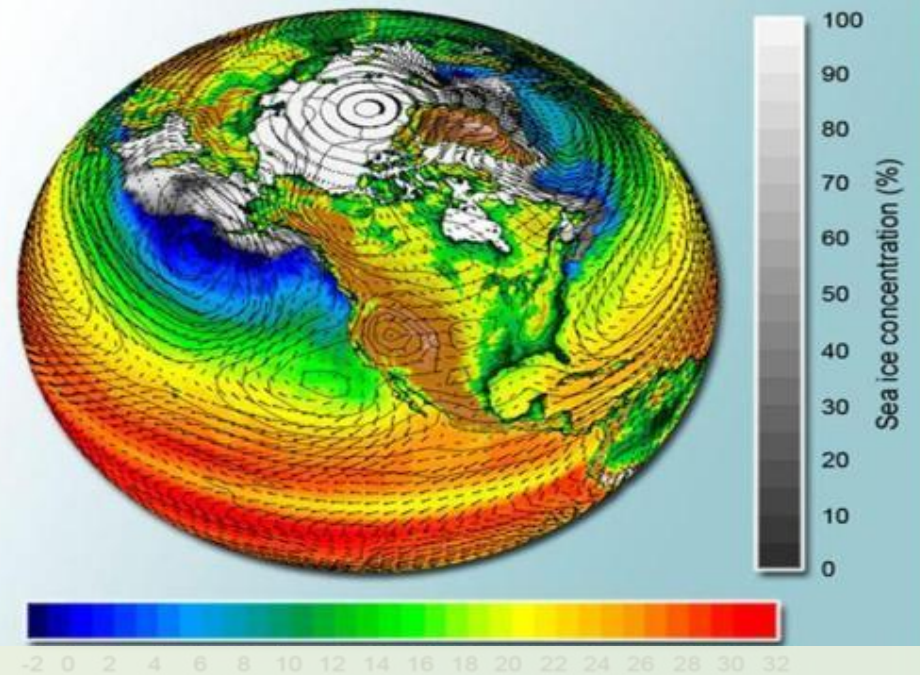
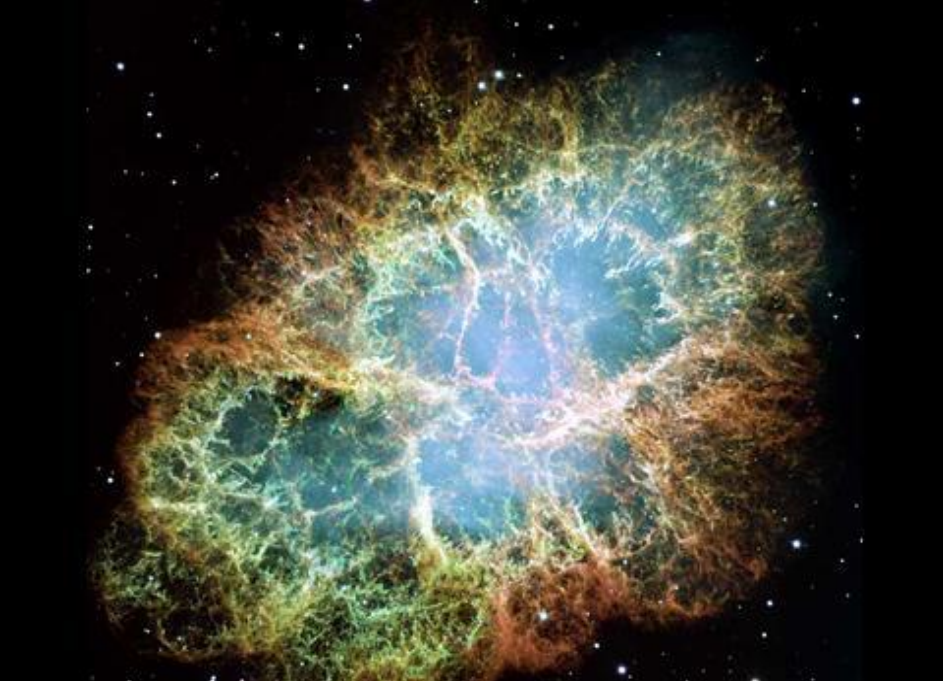


Land change dynamics (2001-2017)

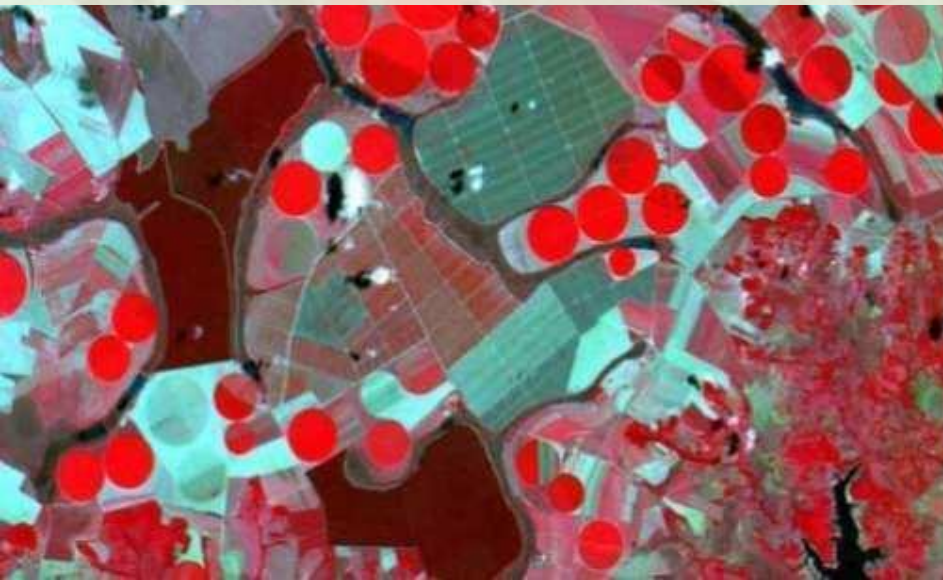
# High accuracy for mapping land use change

	1	2	3	4	5	6	7	8	9	UA
1 Cerrado	393	0	0	12	0	0	0	0	0	0.97
2 Fallow-Cotton	0	33	0	0	1	2	0	0	0	0.92
3 Forest	1	0	136	0	0	0	0	0	0	0.99
4 Pasture	6	0	1	357	3	1	0	5	0	0.96
5 Soy-Corn	0	1	1	1	352	18	0	26	4	0.87
6 Soy-Cotton	0	0	0	0	13	376	0	4	0	0.96
7 Soy-Fallow	0	0	0	0	0	0	88	0	0	1.00
8 Soy-Millet	0	0	0	0	25	2	0	199	2	0.87
9 Soy-Sunflower	0	0	0	0	4	0	0	1	47	0.90
PA	0.98	0.97	0.99	0.96	0.88	0.94	1.00	0.85	0.89	

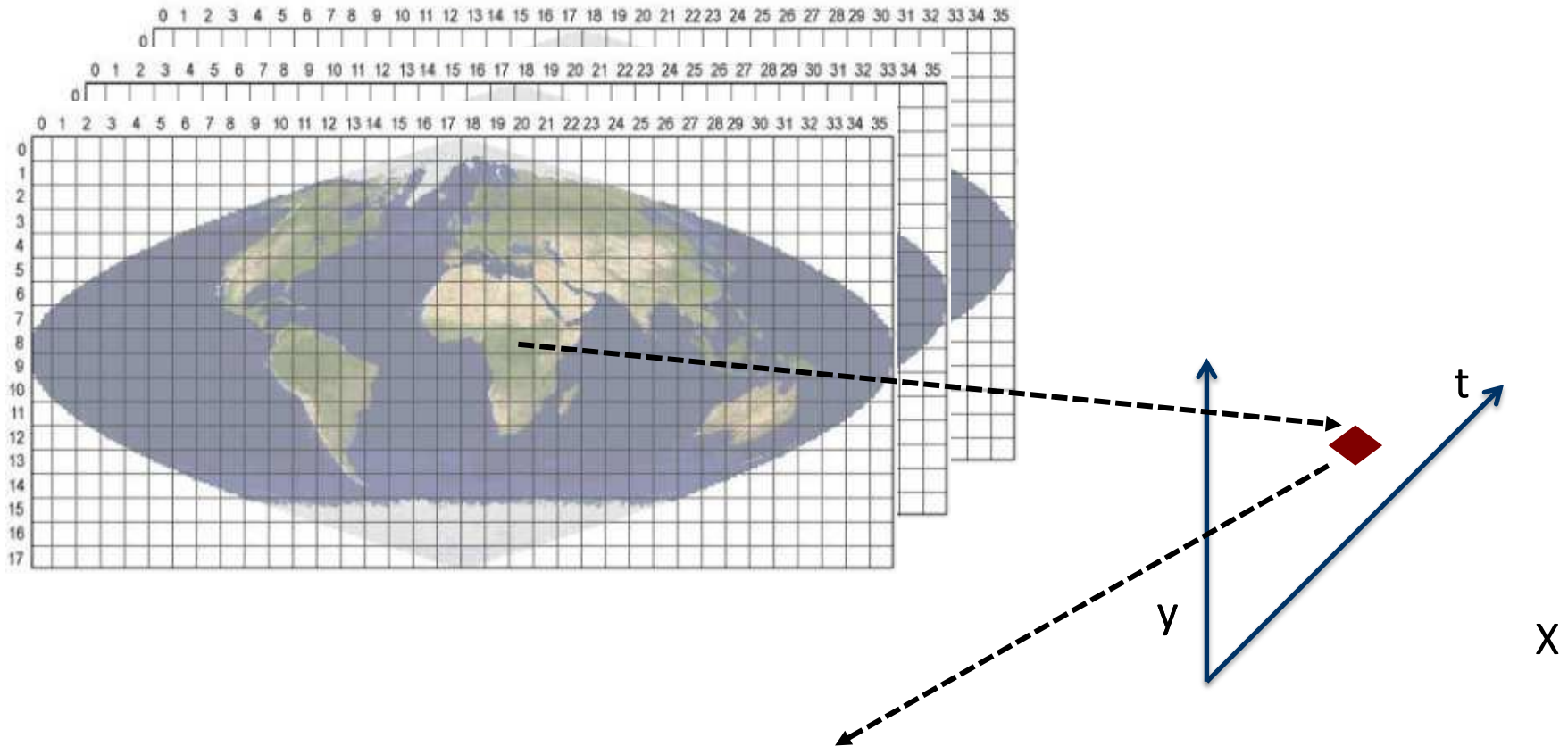
Mato Grosso – Brazil's agricultural frontier  
Overall accuracy of 94%



**What do these data have in common?**

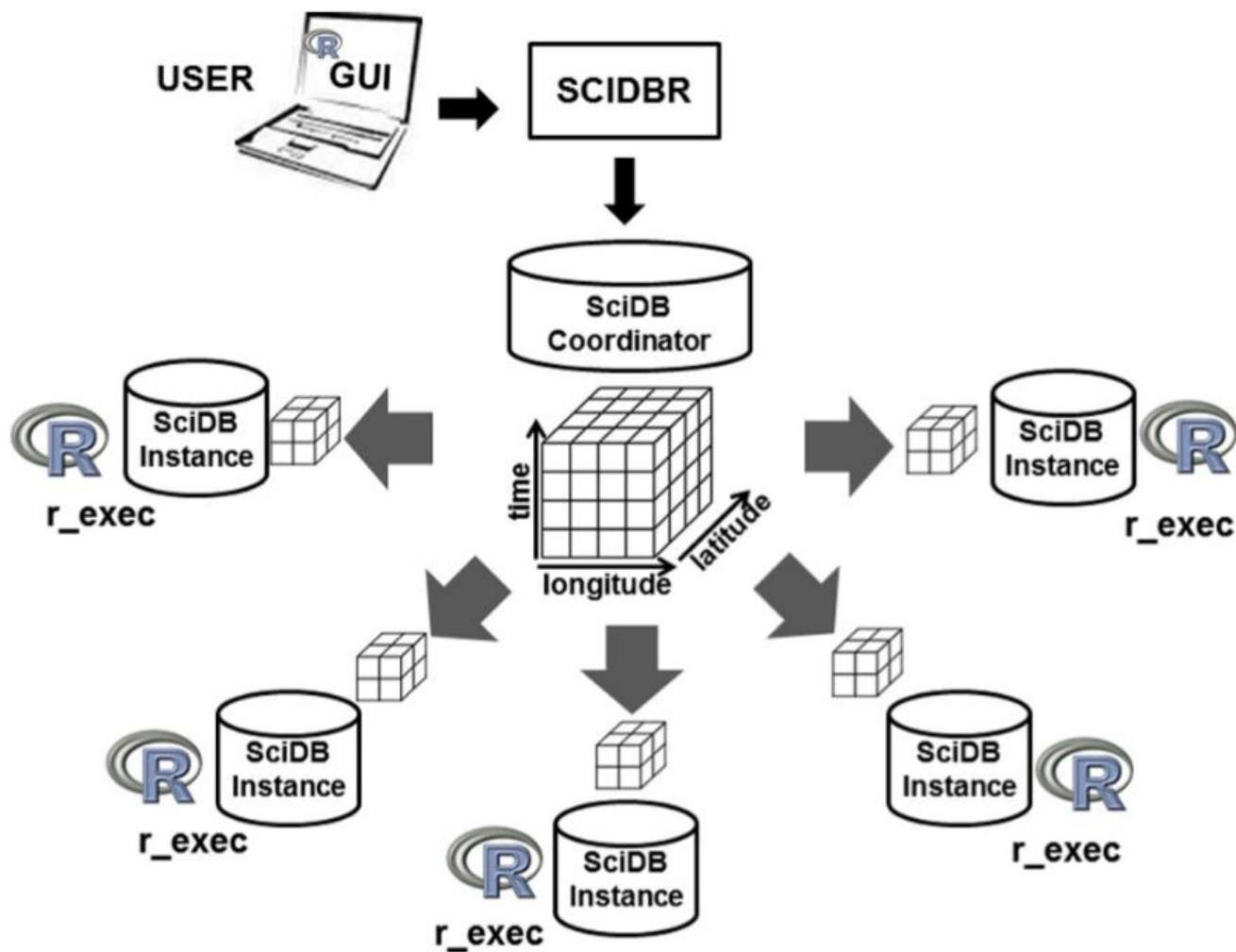


# Array databases: all data in a single array

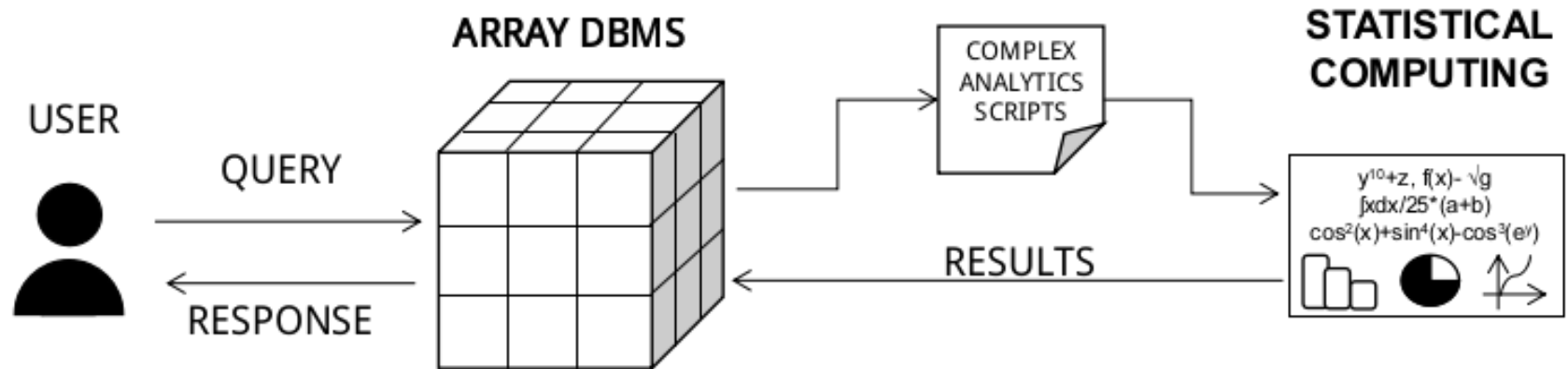


result = analysis\_function (points in space-time )

# SciDB: efficient processing of R scripts

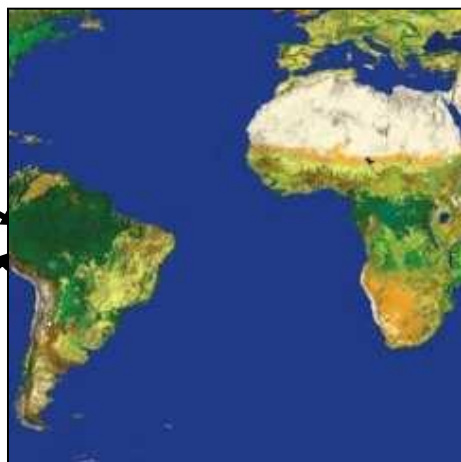


# Flexibility with performance

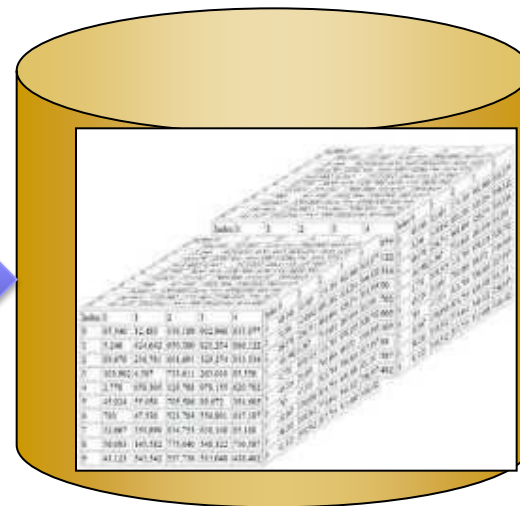


Case Study	Area (km <sup>2</sup> )	Decision dimensions	Measures (GB)	Proc time (hours)
Mato Grosso	900,000	92	135	6
Cerrado	2,050,000	92	308	13

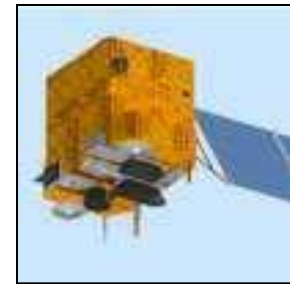
# Global Land Observatory



Methods for land change for forestry and agriculture uses



40 years of LANDSAT + 12 years of MODIS + SENTINELs + CBERS



Building knowledge and data about global land change