

Journée des doctorants du LIG
Grenoble, 14 juin 2012

Agrégation de données pour l'analyse de systèmes complexes

Robin Lamarche-Perrin

LIG-MAGMA+MESCAL, UdG

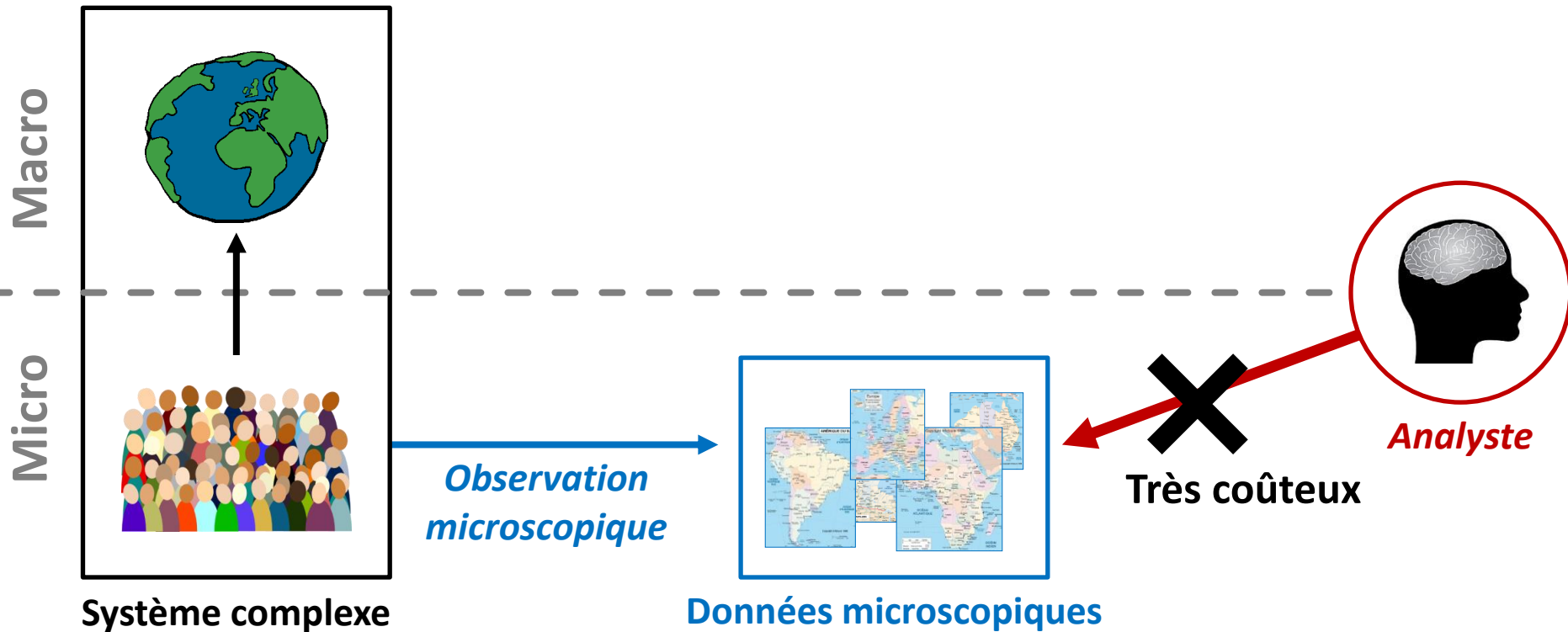
Yves Demazeau

LIG-MAGMA, CNRS

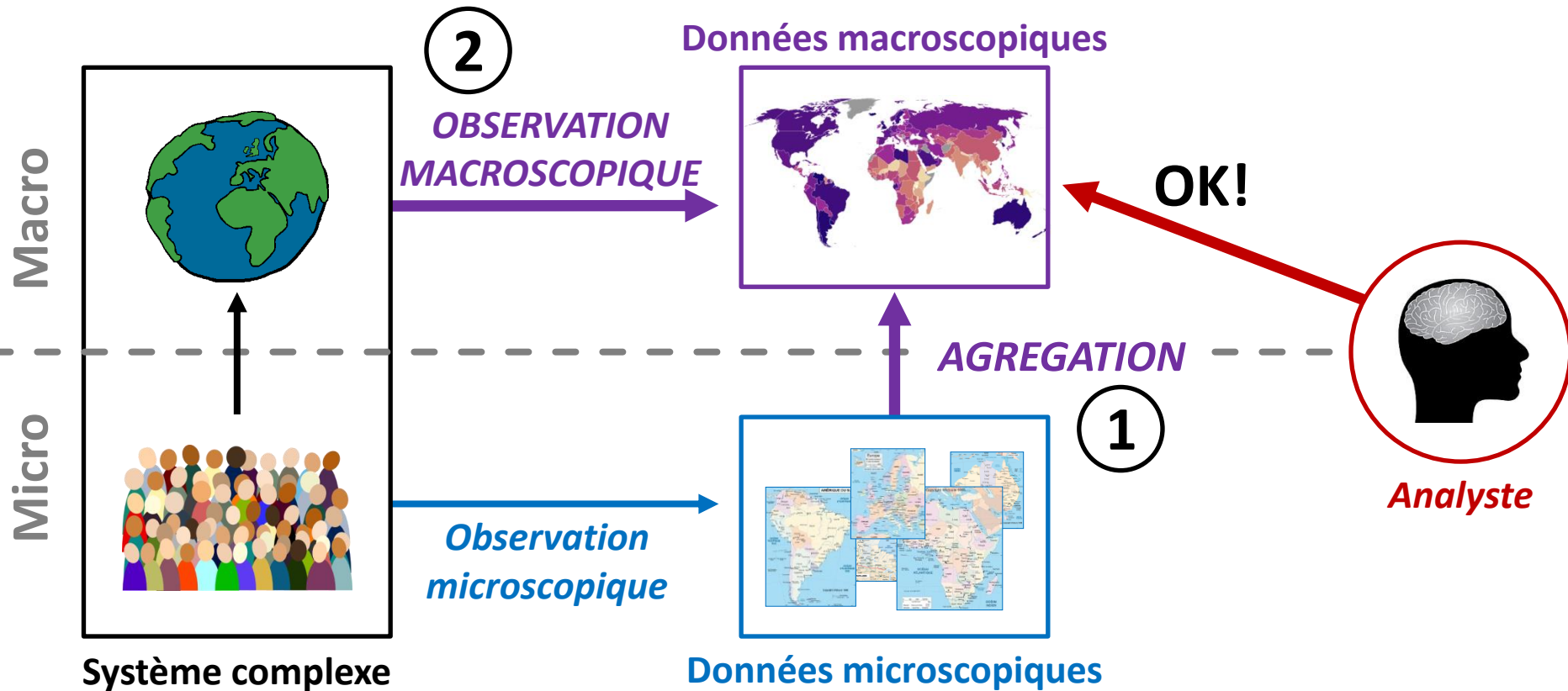
Jean-Marc Vincent

LIG-MESCAL, UJF

Changer le niveau de l'analyse

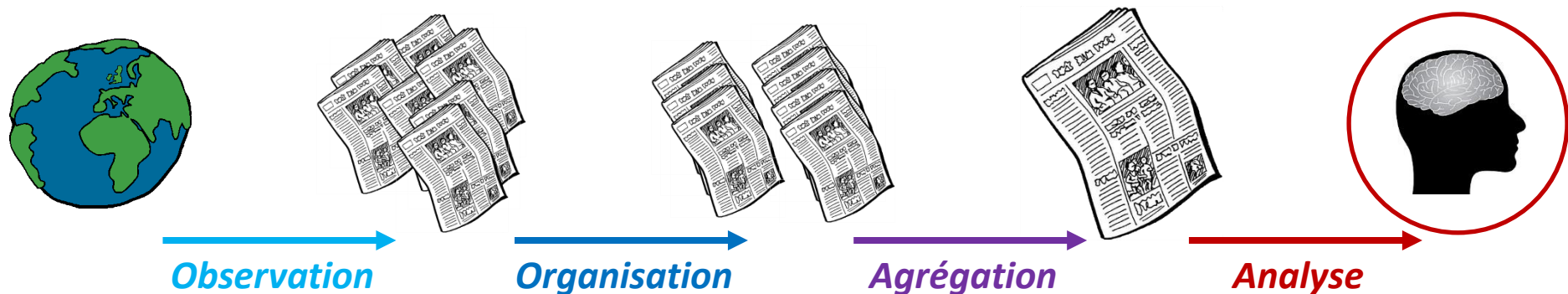


Changer le niveau de l'analyse



Projet GEOMEDIA

Agrégations spatiales de données médiatiques



Quelles agrégations sont les plus pertinentes ?

	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...	Total
USA	x	25	19	18	24	24	24	...	423
France	25	x	36	10	15	15	30	...	248
Libye	19	36	x	0	7	0	2	...	308
Israël	18	10	0	x	4	62	0	...	153
Syrie	24	15	7	4	x	1	0	...	260
Palestine	24	15	0	62	1	x	0	...	126
Afghanistan	24	30	2	0	0	0	x	...	131
...
Total	423	248	308	153	260	126	131	...	3520

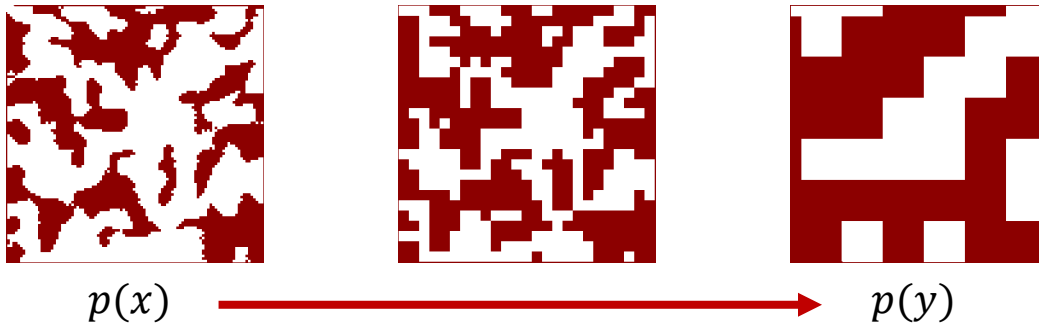


UNEP



WUTS2

Évaluer l'agrégation des données



Que gagne-t-on en simplicité ?
Que perd-on en information ?

Théorie de l'information

Entropie de Shannon

$$H = - \sum_x p(x) \log_2(p(x))$$

Perte d'information

$$L = \sum_y p(y) \log_2 |y|$$

Divergence de Kullback-Leibler

$$D = - \sum_x p(x) \log_2 \left(\frac{p(y)}{p(x)|y|} \right)$$

Merci pour votre attention

Robin.Lamarche-Perrin@imag.fr

Information-Theoretic Measures of Aggregation for the Analysis of Complex Systems
 Robin Lamarche-Perrin¹, Yves Demazeau², and Jean-Marc Vincent³
 Laboratoire d'Informatique de Grenoble
¹Université de Grenoble, ²CNRS, ³Université Joseph Fourier

General Problem
Context: Analysis of complex systems in order to describe, explain and predict their dynamics
Problem: Microscopic data are unusable in practice (size, heterogeneity, limitation to local semantics)
The analyst needs a macroscopic point of view

Notion of Aggregation
 Aggregation (or generalization) consists in losing information in order to generate macroscopic points of view

How to measure and compare aggregations?

What do we Gain?
Shannon Entropy
 Measures the quantity of information needed to encode a set of data

$$H = - \sum_x p(x) \log_2(p(x))$$

Entropy Gain
 Measures the quantity of information reduced by the aggregation

$$G = H_{\text{micro}} - H_{\text{macro}}$$

What do we Lose?
Information Loss
 Measures the quantity of information needed to disaggregate a set of data

$$L = \sum_y p(y) \log_2|y|$$

K.-L. Divergence
 Measures the quantity of information that differs between two sets of data

$$D = - \sum_x p(x) \log_2 \left(\frac{p(y)}{p(x)|y|} \right)$$

Which Trade-Off?
Information Criterion
 Measures a trade-off between what we gain and what we lose

$$C = G - D$$

Parametric Criterion

$$C_p = pG - (1 - p)D$$

$$p = 0 \rightarrow \text{No aggregation at all}$$

$$p = \frac{1}{2} \rightarrow \text{Aggregation iff } G > D$$

$$p = 1 \rightarrow \text{Maximal aggregation}$$

Territorial Aggregations
The GEOMEDIA Project (in collaboration with the CIST, Paris)
 Project: building a platform for the global analysis of media information
 Data: newspaper articles, demographic data, economic data
 Aggregations: by states, by dates, by actors, by topics, etc.
 What is the best geographical level for a given analysis?
 Which aggregates are the most meaningful for a given analysis?

Processes Aggregations
The TRIVA Software
 Project: building tools for the analysis of large-scale distributed systems
 Data: data flows, communications, computing powers, internal states
 Aggregations: in space and time
 Which parts of the hierarchy should be aggregated?

Lamarche-Perrin, Vincent, Demazeau. Informational Measures of Aggregation for Complex Systems Analysis. Research Report RR-LIG-026, LIG, France, May 2012.

Laboratoire d'Informatique de Grenoble

Lamarche-Perrin, Vincent, Demazeau. Informational Measures of Aggregation for Complex Systems Analysis. *LIG Report RR-LIG-026*, mai 2012.

Lamarche-Perrin, Demazeau, Vincent. Organisation, agrégation et visualisation d'informations médiatiques. *GIS-CIST « Fonder les sciences du territoire »*, p. 240-246, Paris, nov. 2011.

Lamarche-Perrin, Demazeau, Vincent. Observation macroscopique et émergence dans les SMA de très grande taille. *JFSMA'11*:53-62, Valenciennes, oct. 2011.