

**Réunion GEOMEDIA**

**Paris, 22-23 octobre 2012**

# **Mesurer, évaluer, agréger l'information médiatique**

**Robin Lamarche-Perrin**

Yves Demazeau

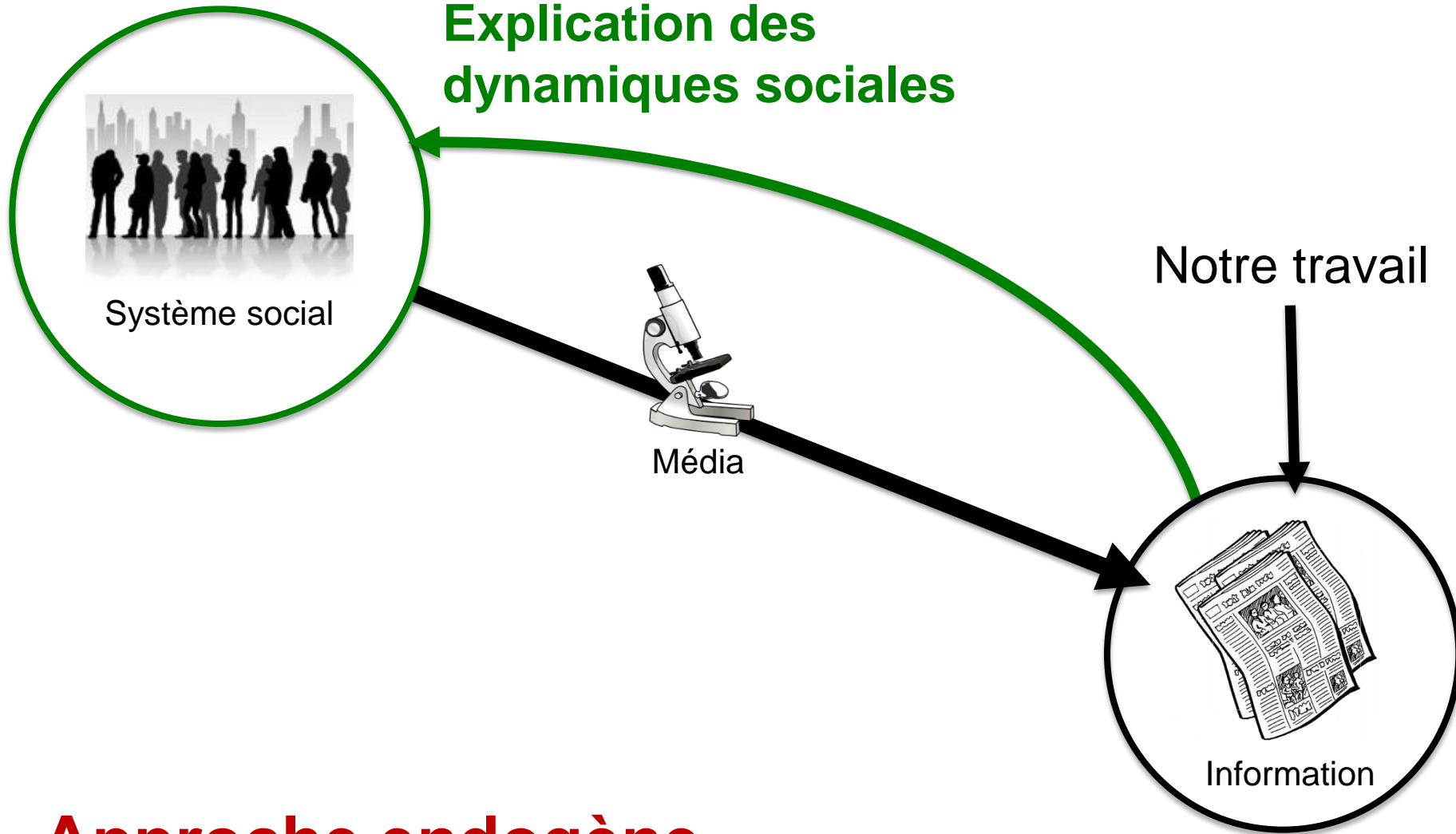
Jean-Marc Vincent

LIG-MAGMA

LIG-MESCAL

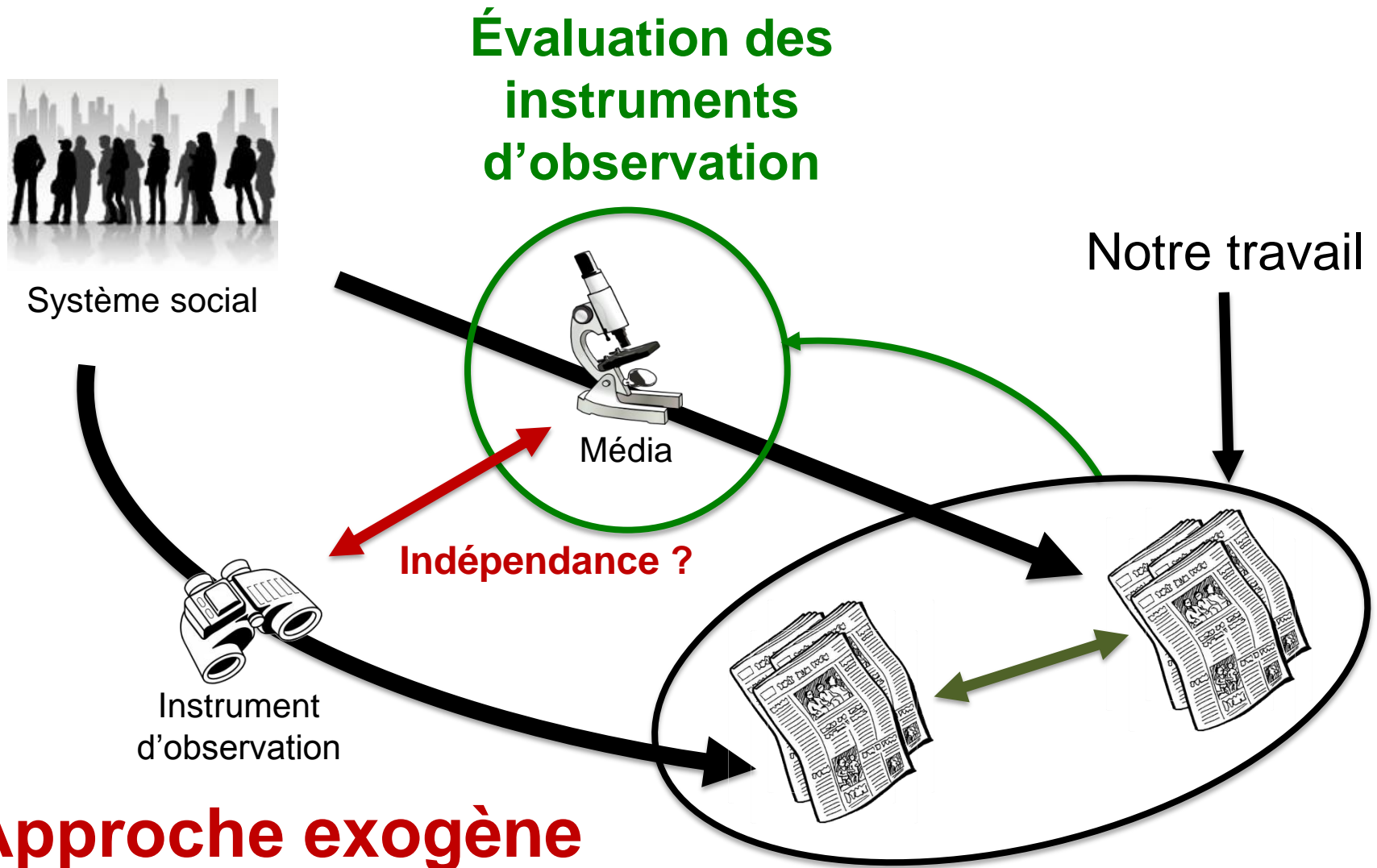
# Objectifs de l'analyse de l'information

Explication des dynamiques sociales

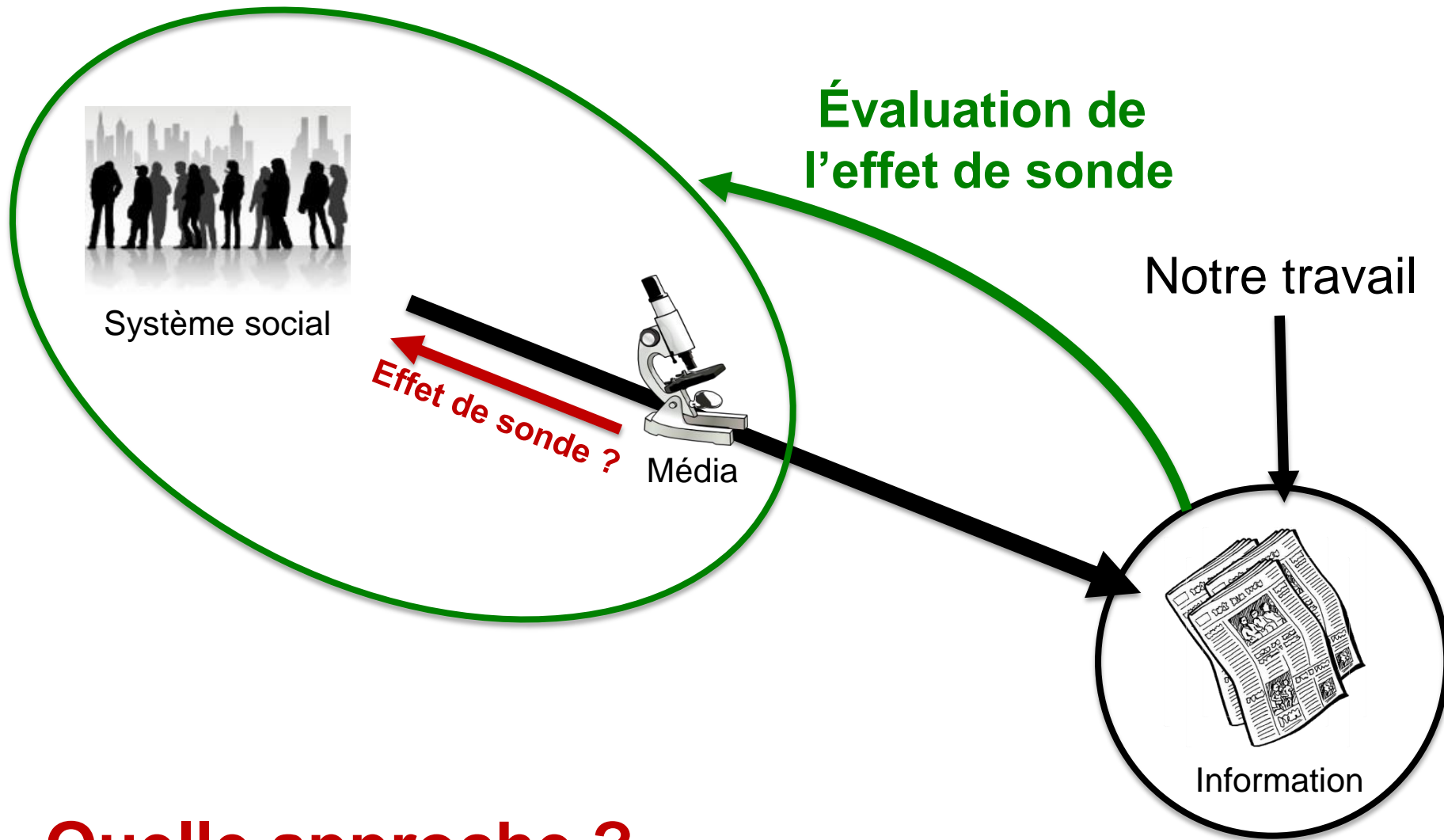


→ **Approche endogène**

# Objectifs de l'analyse de l'information



# Objectifs de l'analyse de l'information

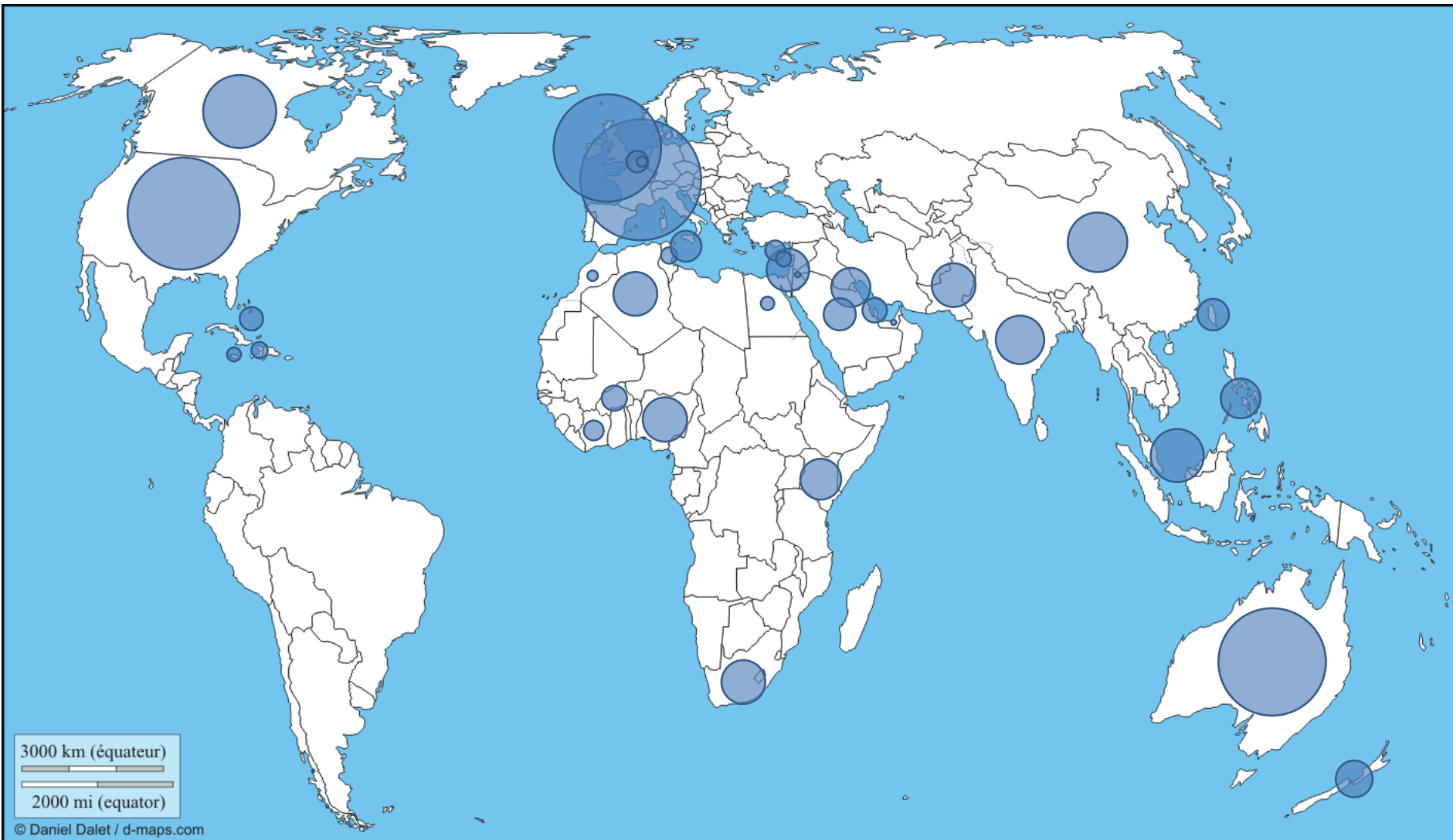


→ **Quelle approche ?**

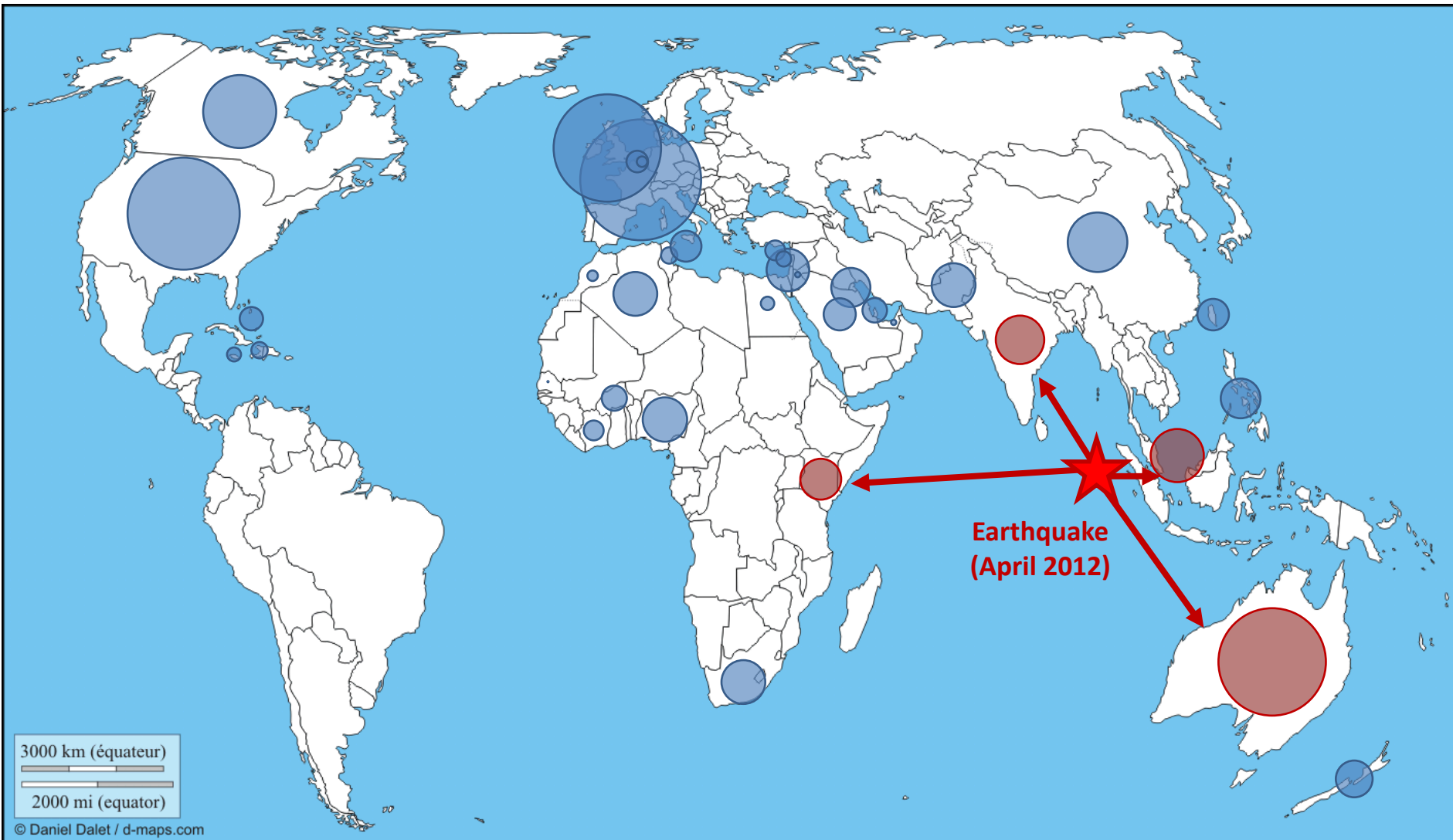
# **COUVERTURE ET REDONDANCE DU RÉSEAU DE CAPTEURS (APPROCHE EXOGÈNE)**

# Taille des flux RSS

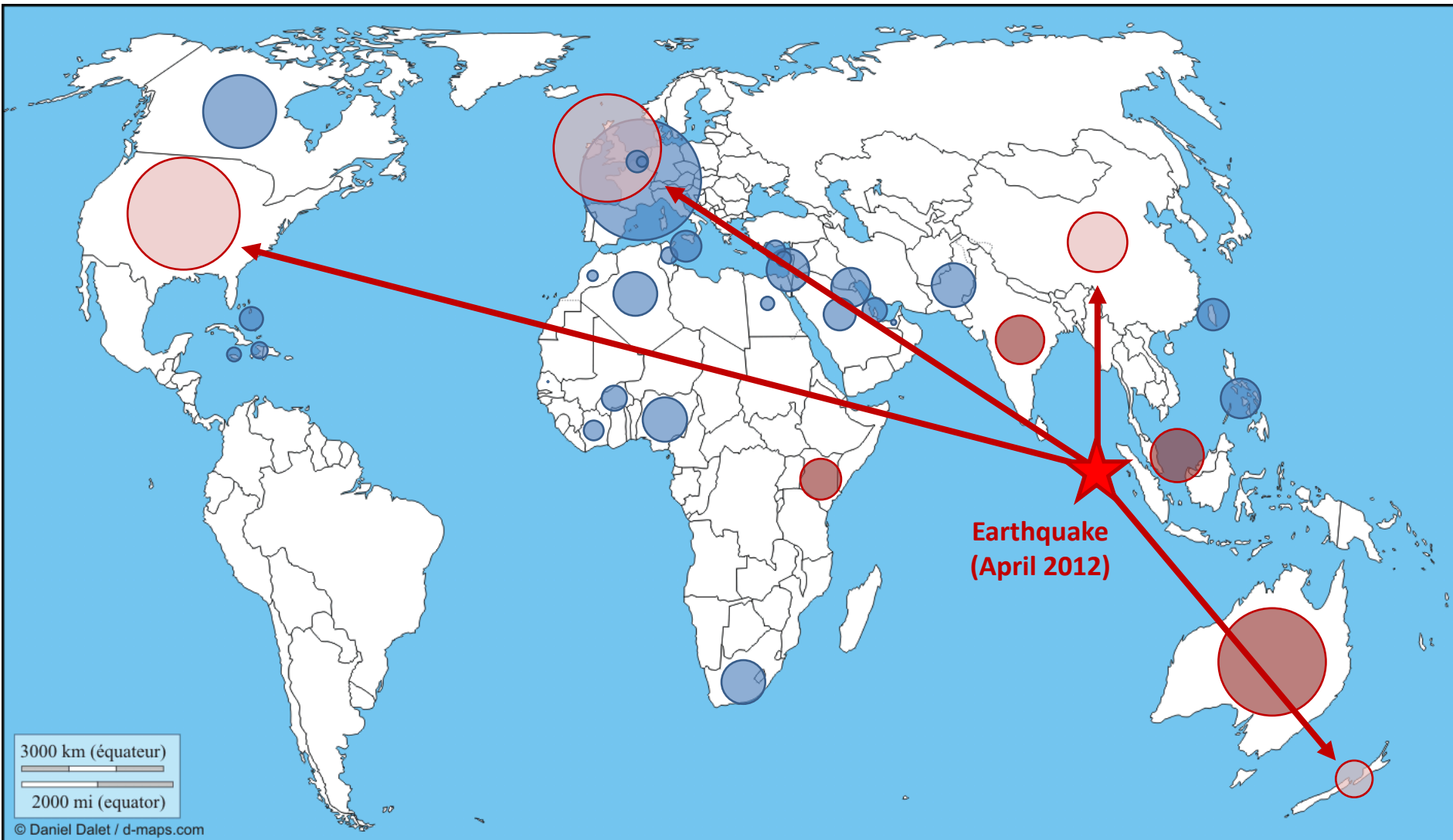
(mai 2011 – sept. 2012)



# Observation d'un évènement

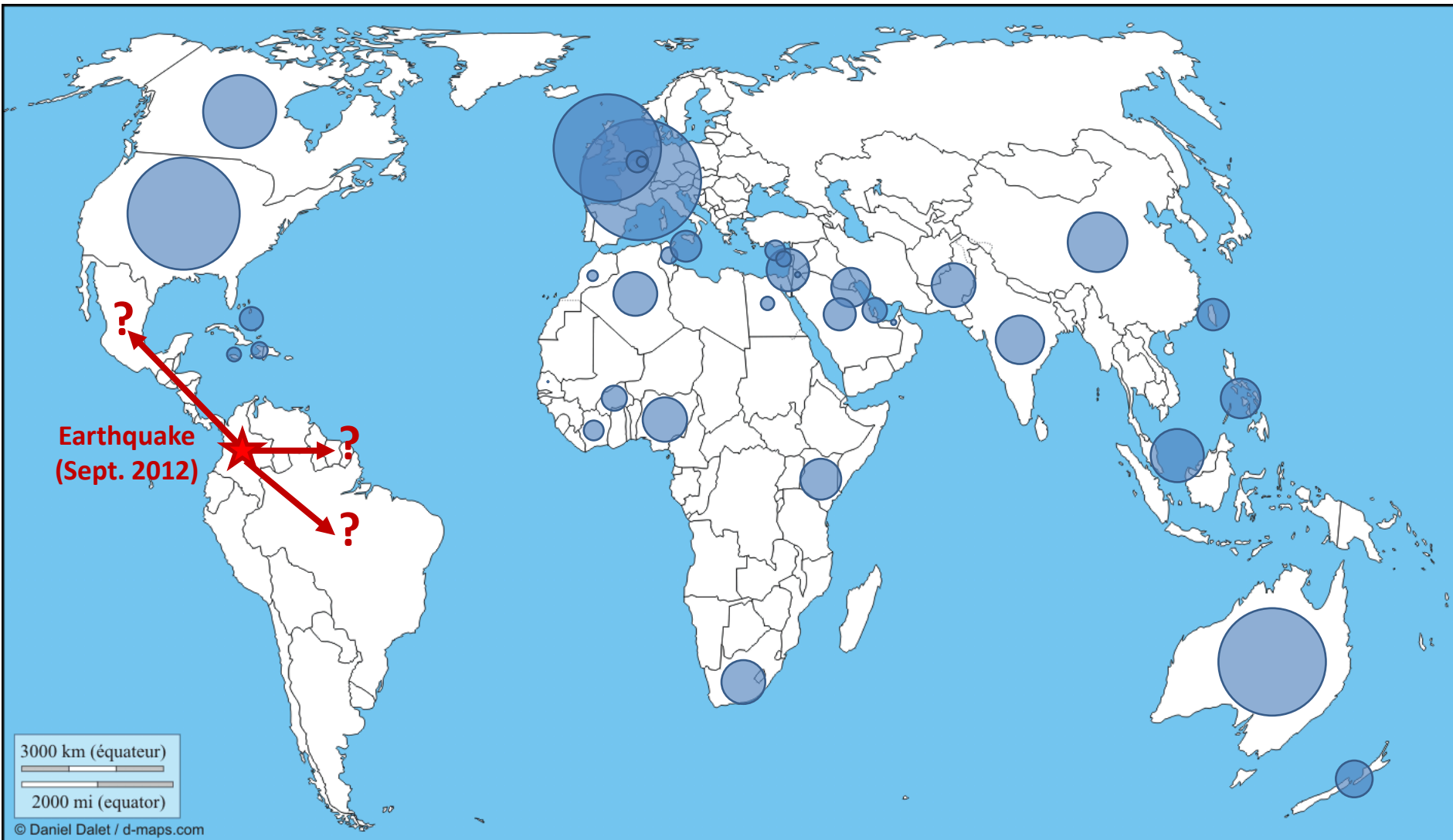


# Observation d'un évènement



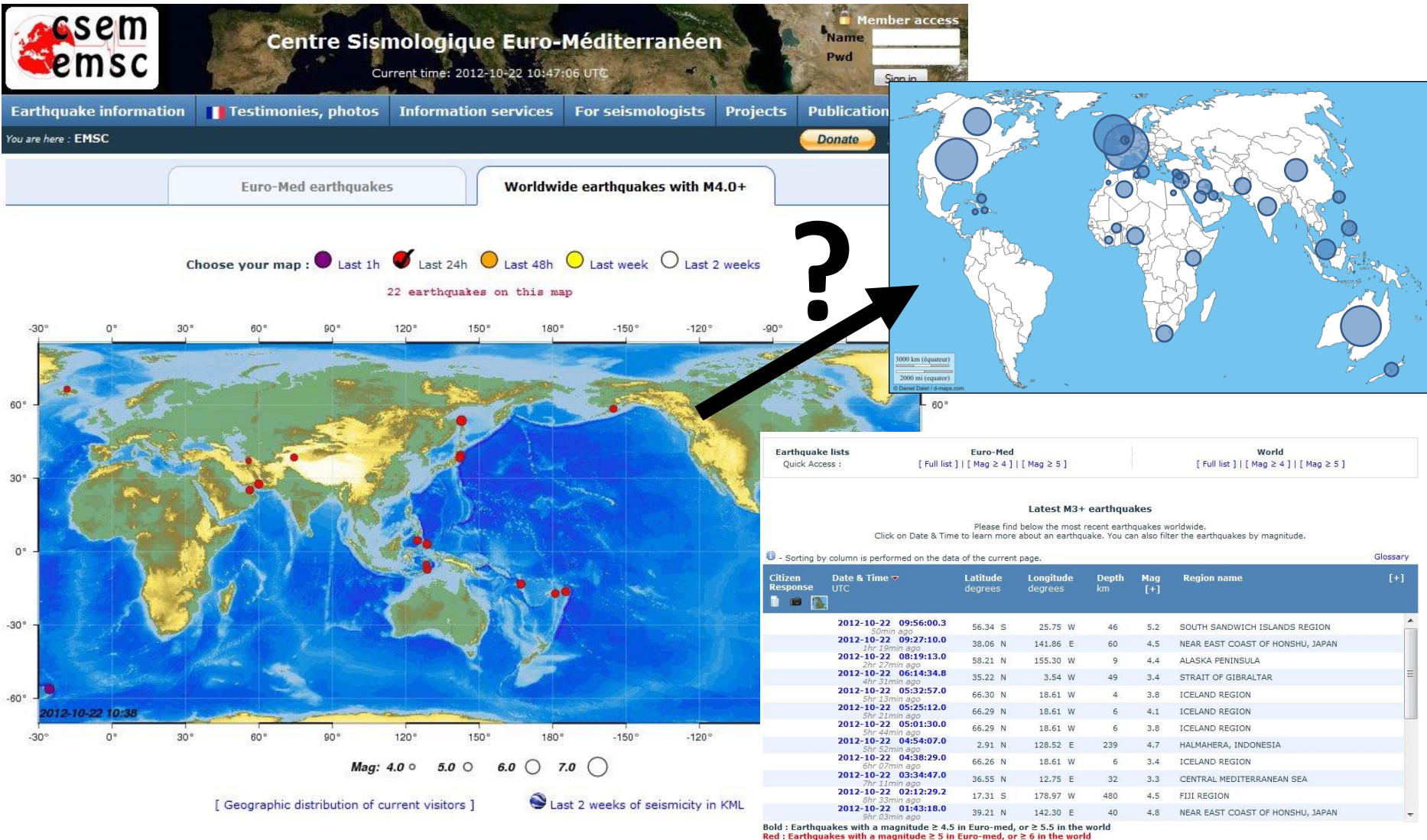


# Observation d'un évènement



# Variable exogène

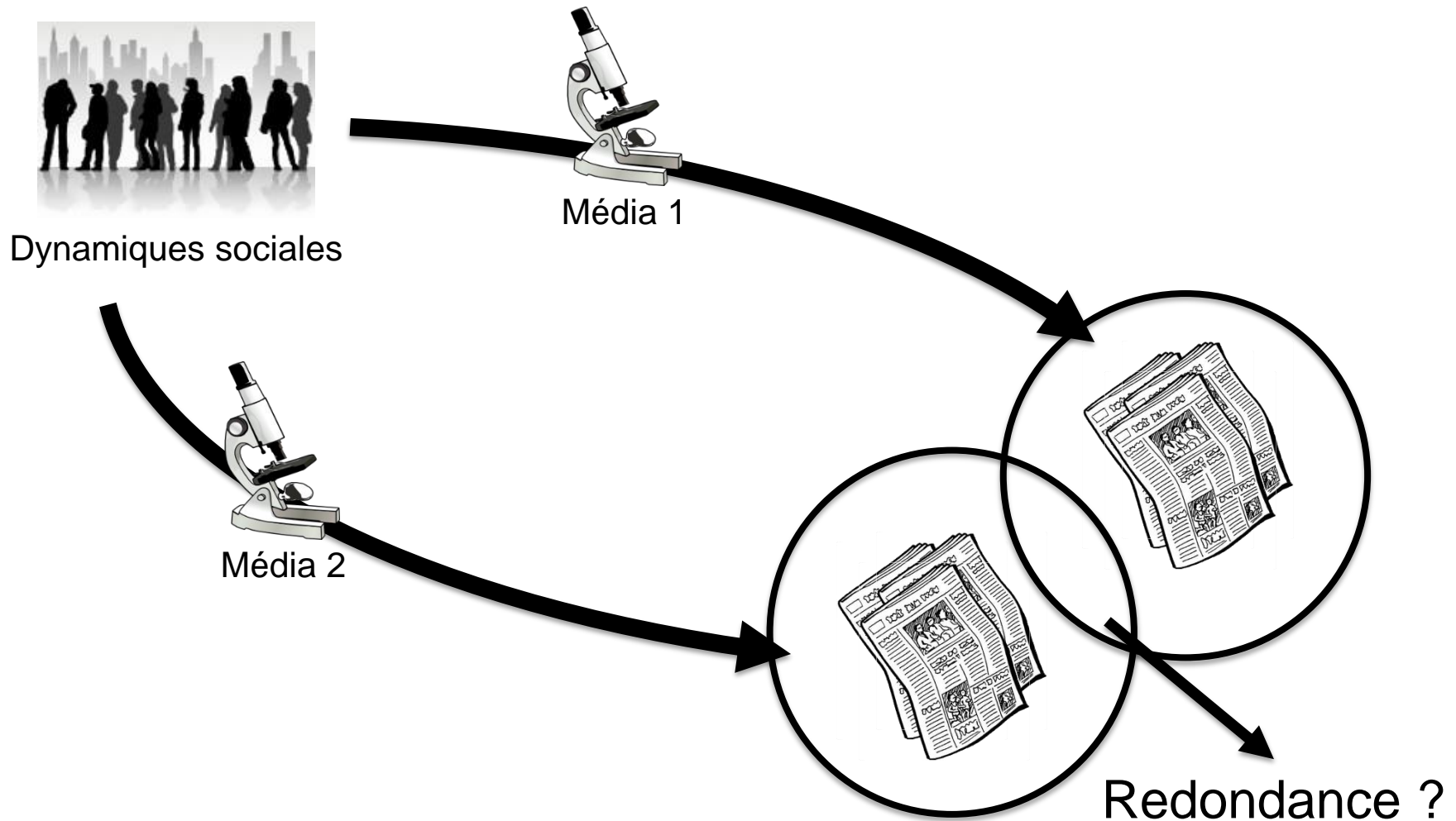
<http://www.emsc-csem.org/#2w>



# Couverture du réseau de capteurs ?

- Couverture thématique
- Couverture linguistique
- Couverture politique
- Couverture temporelle
- *Etc.*

# Redondance des sondes



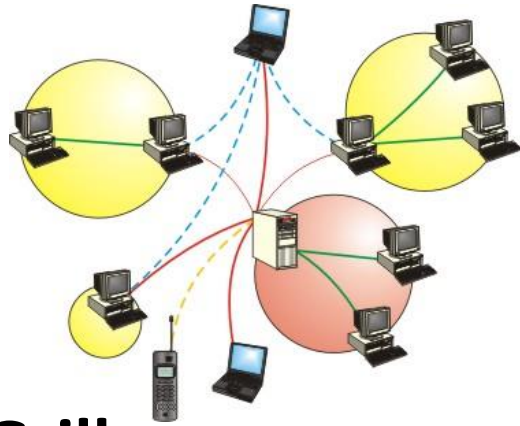
# Méthode générale

- Ce que l'on gagne :
  - En couverture, en précision
  - **Divergence de Kullback-Leibler**
- Ce que l'on perd :
  - En complexité, en redondance
  - **Entropie de Shannon**

# **GRANULARITÉ DE L'INFORMATION (APPROCHE ENDOGÈNE)**

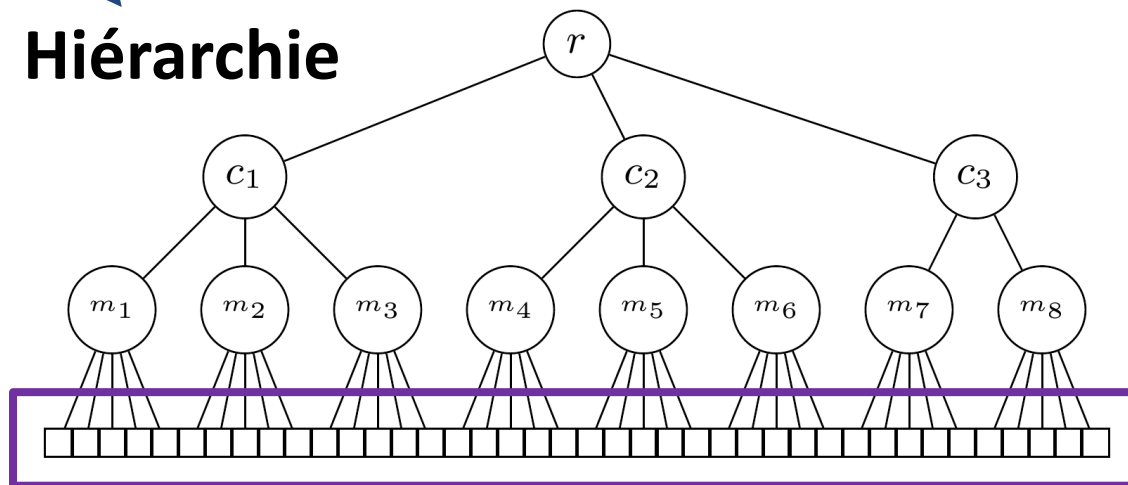
# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués

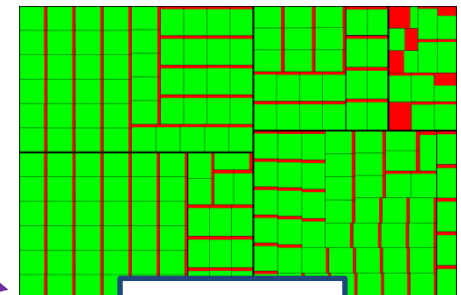


**Grille**

**Hiérarchie**



**Représentation  
treemap**

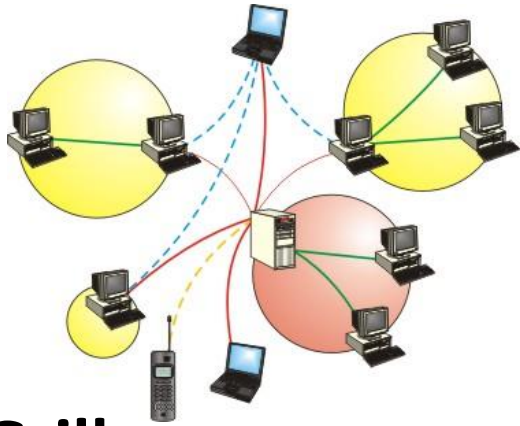


**Processus**



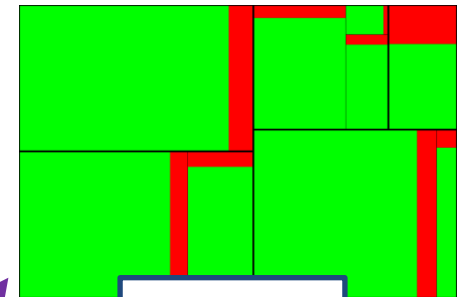
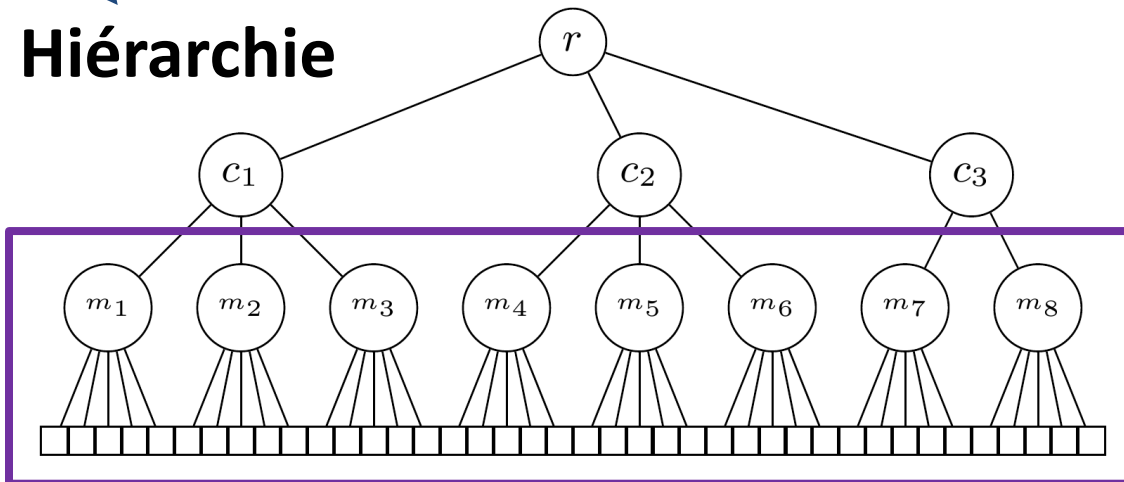
# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués



Grille

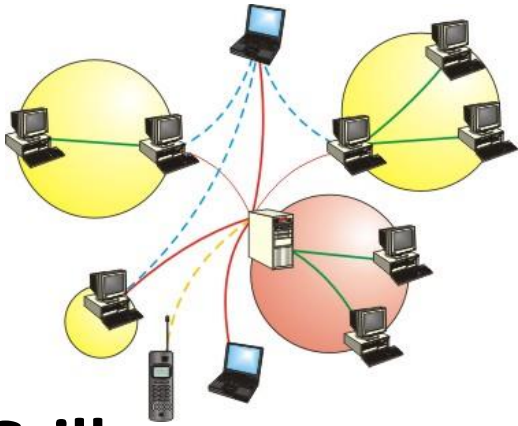
Hiérarchie





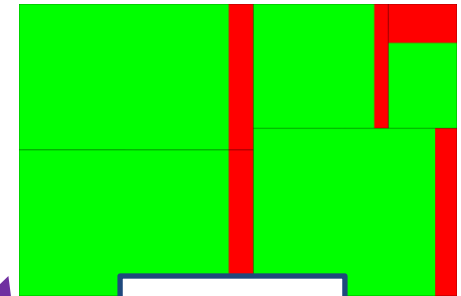
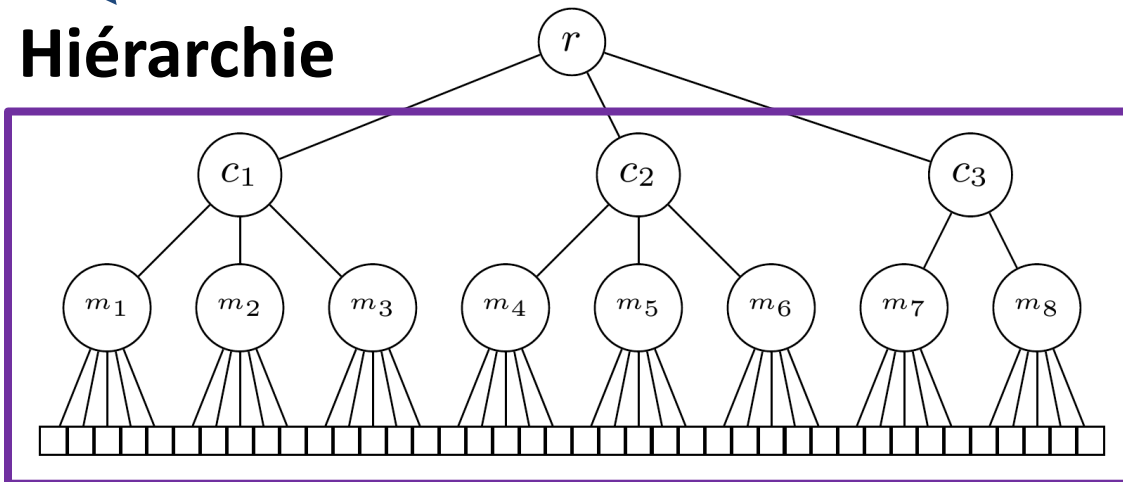
# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués

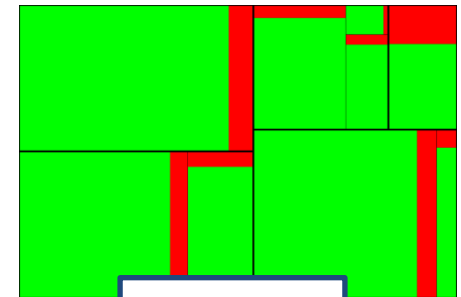


Grille

Hiérarchie



Clusters



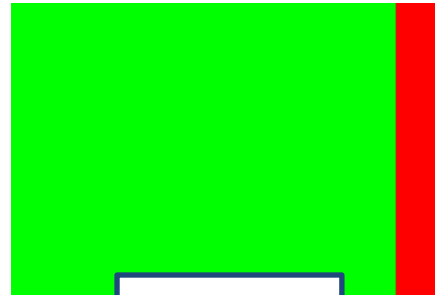
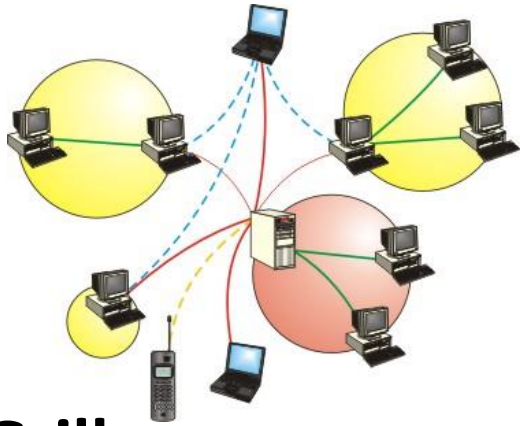
Machines



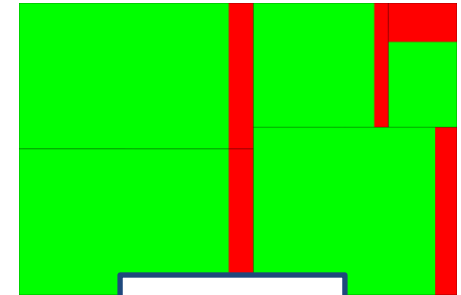
Processus

# Projet TRIVA - ANR SONGS

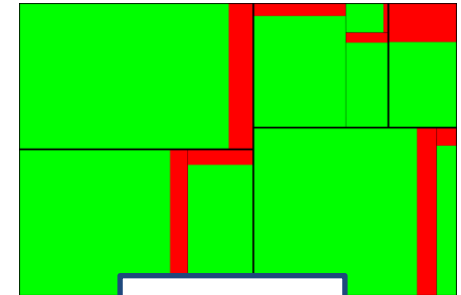
## Visualisation de grands systèmes distribués



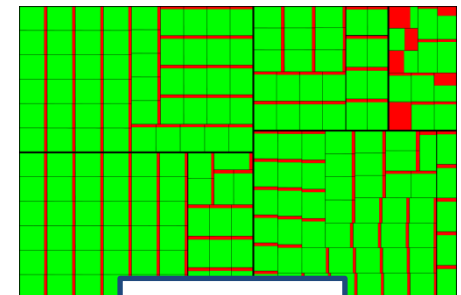
Système



Clusters



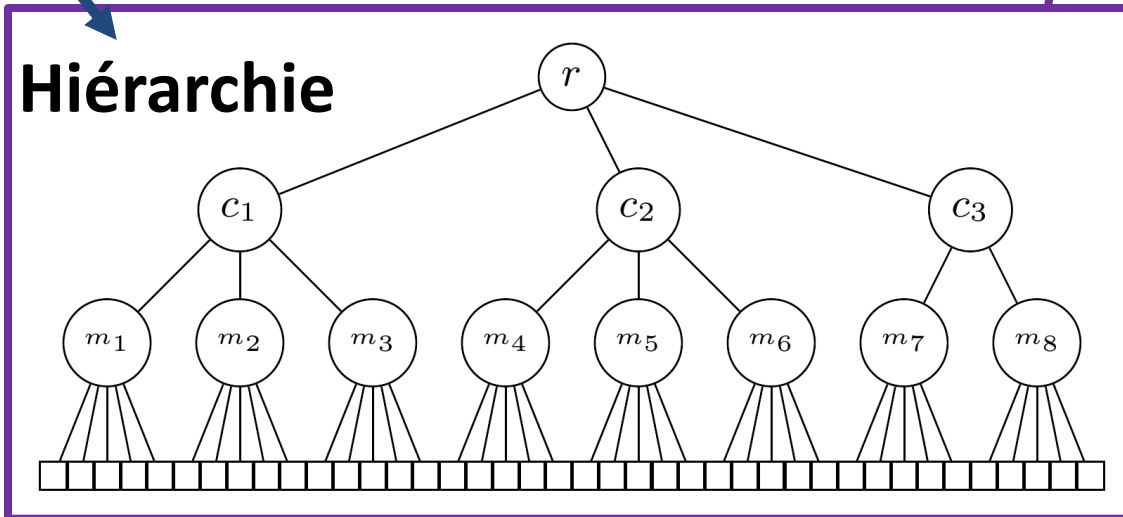
Machines



Processus

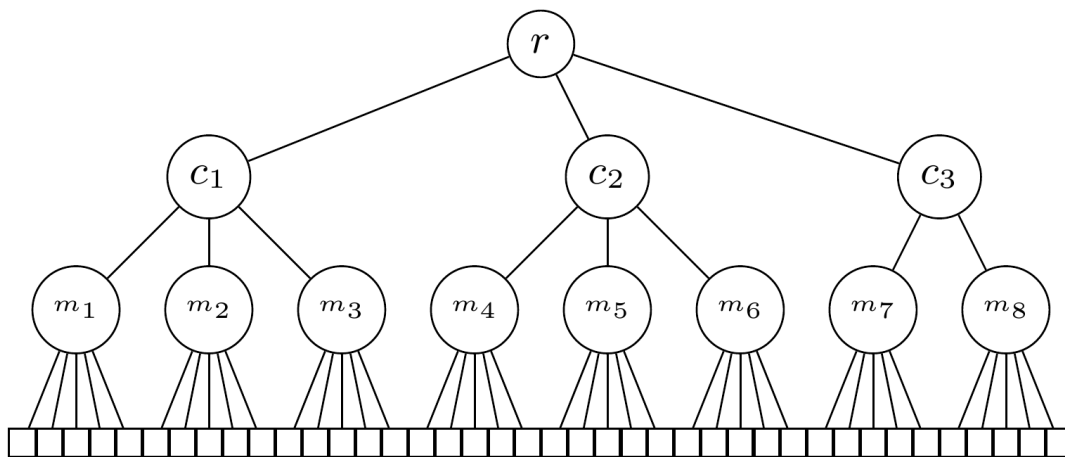
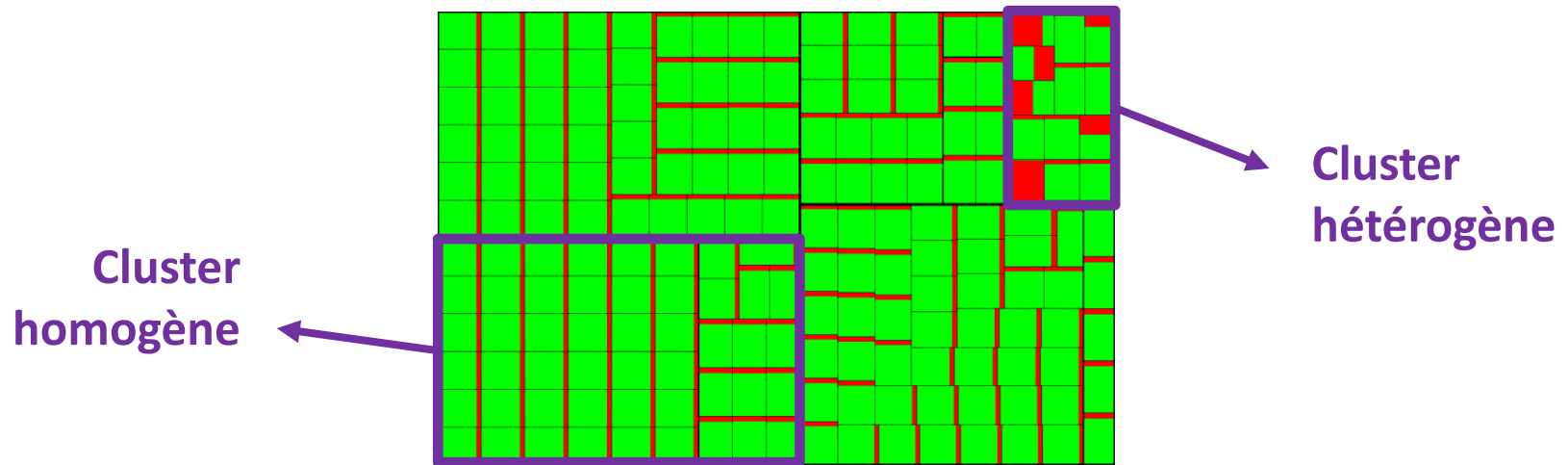
Grille

Hiérarchie



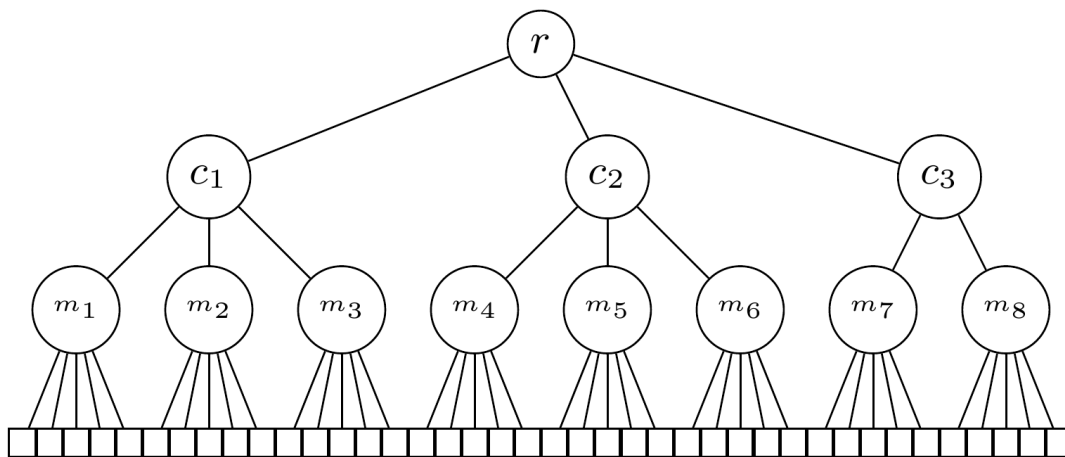
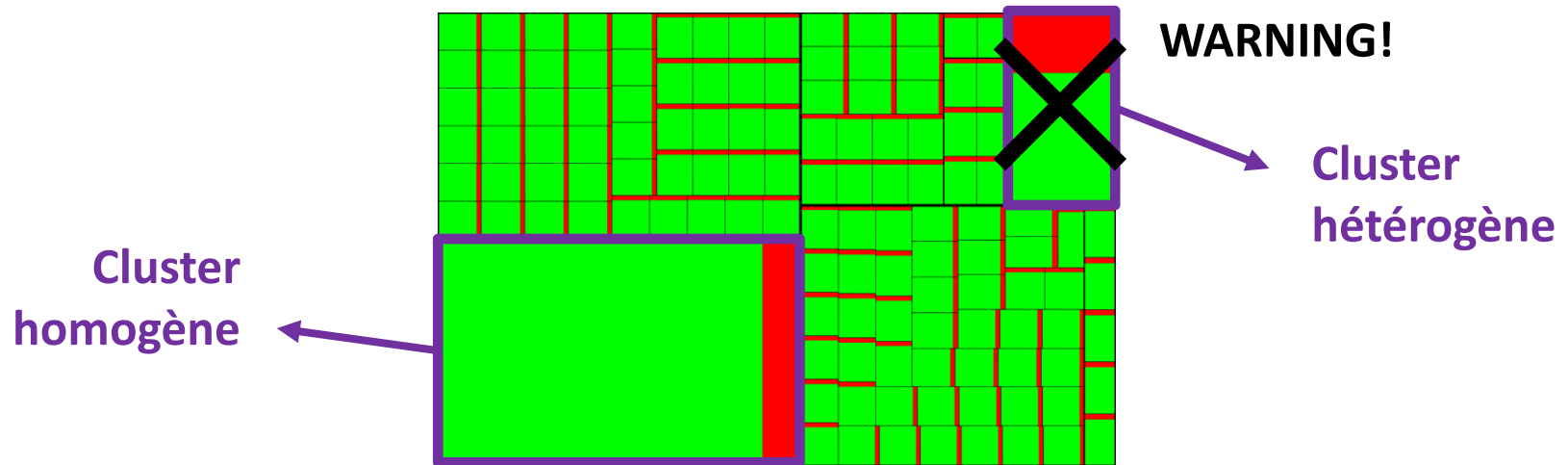
# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués



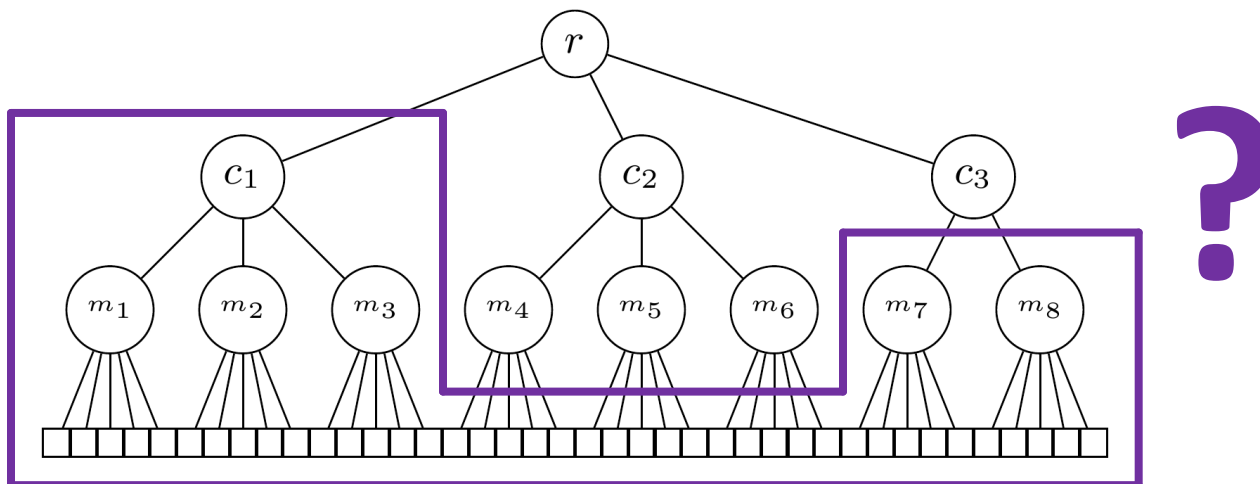
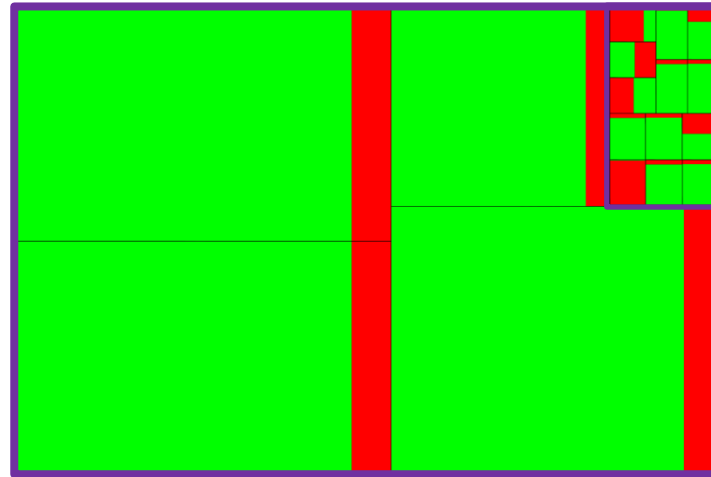
# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués



# Projet TRIVA - ANR SONGS

## Visualisation de grands systèmes distribués



# Mesures en théorie de l'information

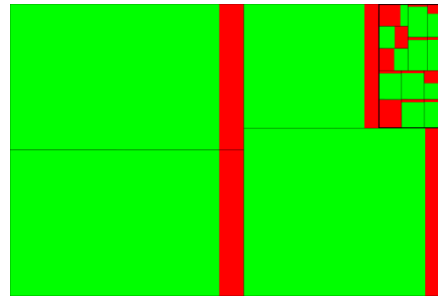
Données  
microscopiques

*Agrégation*

Données  
macroscopiques



$p(x)$



$p(y)$

**Que gagne-t-on ?  
Que perd-on ?**

## Théorie de l'information

**Entropie de Shannon**

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

**Divergence de Kullback-Leibler**

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

**Critère informationnel**

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

# Méthode générale

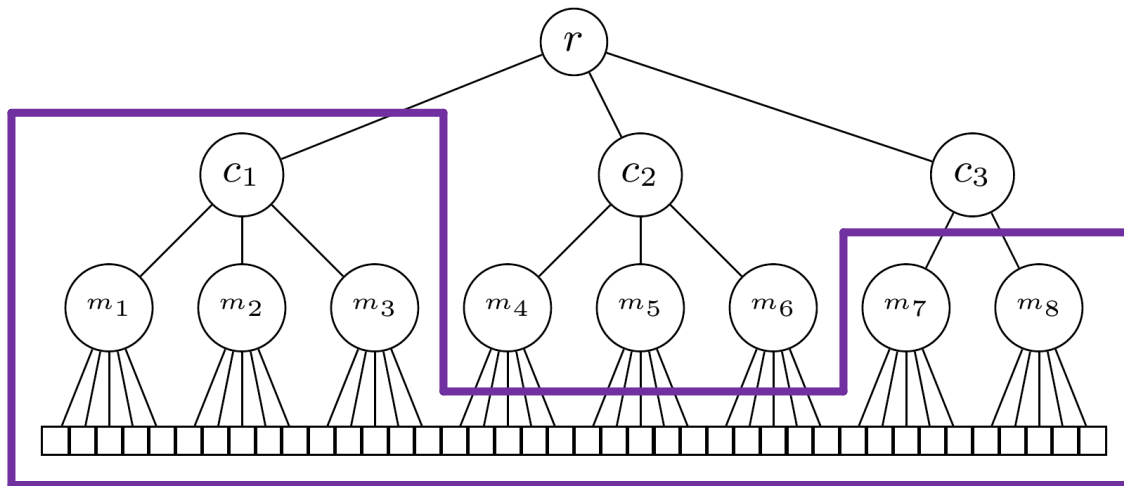
- Ce que l'on gagne :
  - En structure, en généricité
  - En simplicité, en facilité de lecture
  - **Entropie de Shannon**
- Ce que l'on perd :
  - En information
  - En précision, en détails
  - **Divergence de Kullback-Leibler**

# Agrégation pRIC-maximale

Maximiser le pRIC

Gain d'entropie vs. Divergence

$$pRIC = p \times G - (1 - p) \times D$$



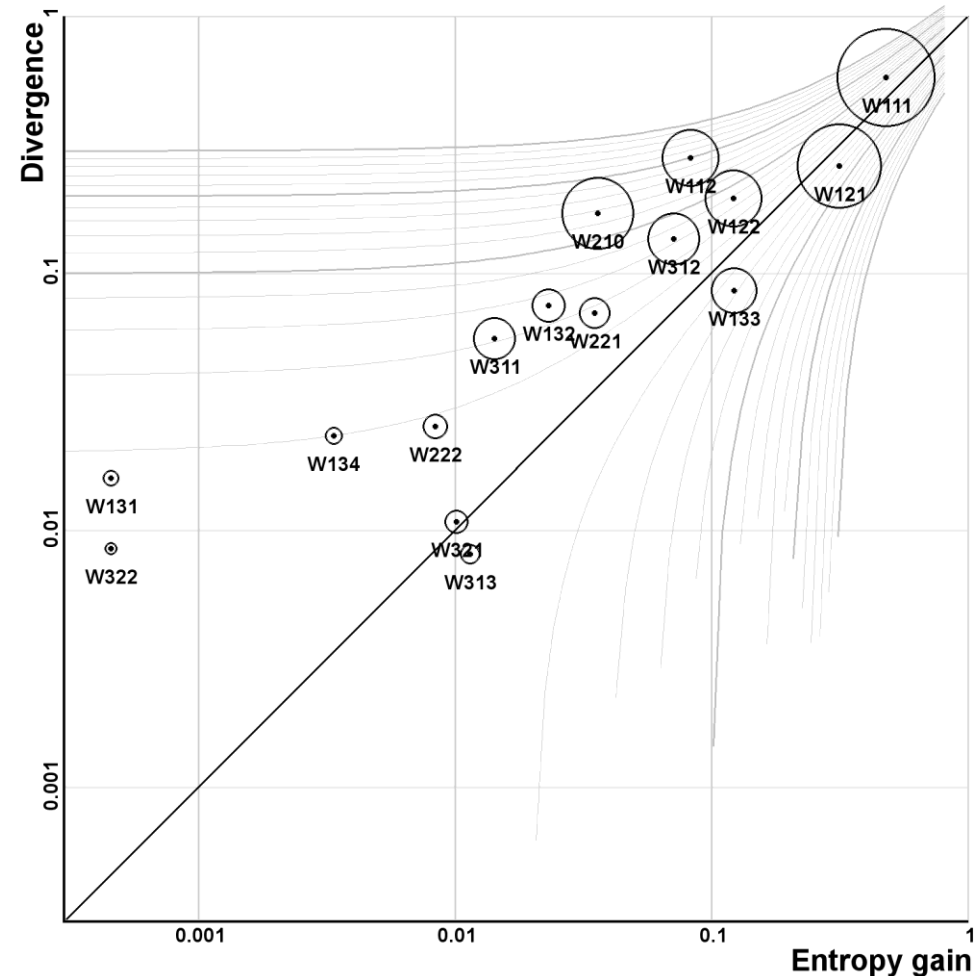
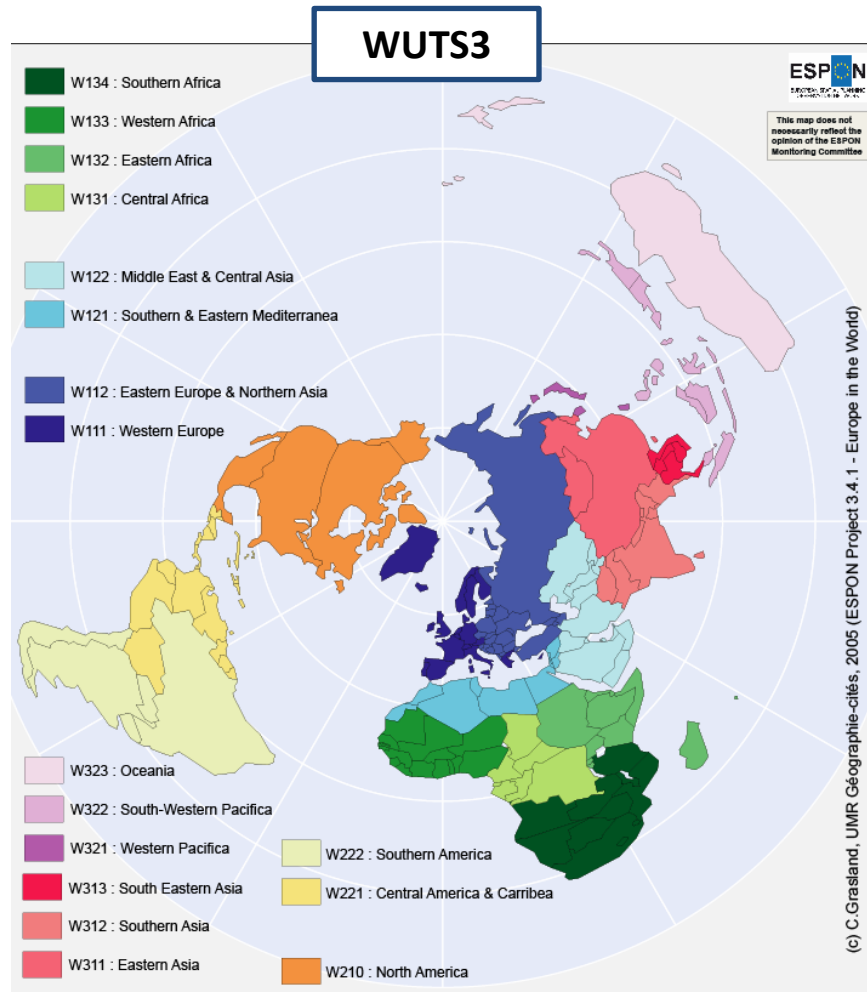
$p = 1$



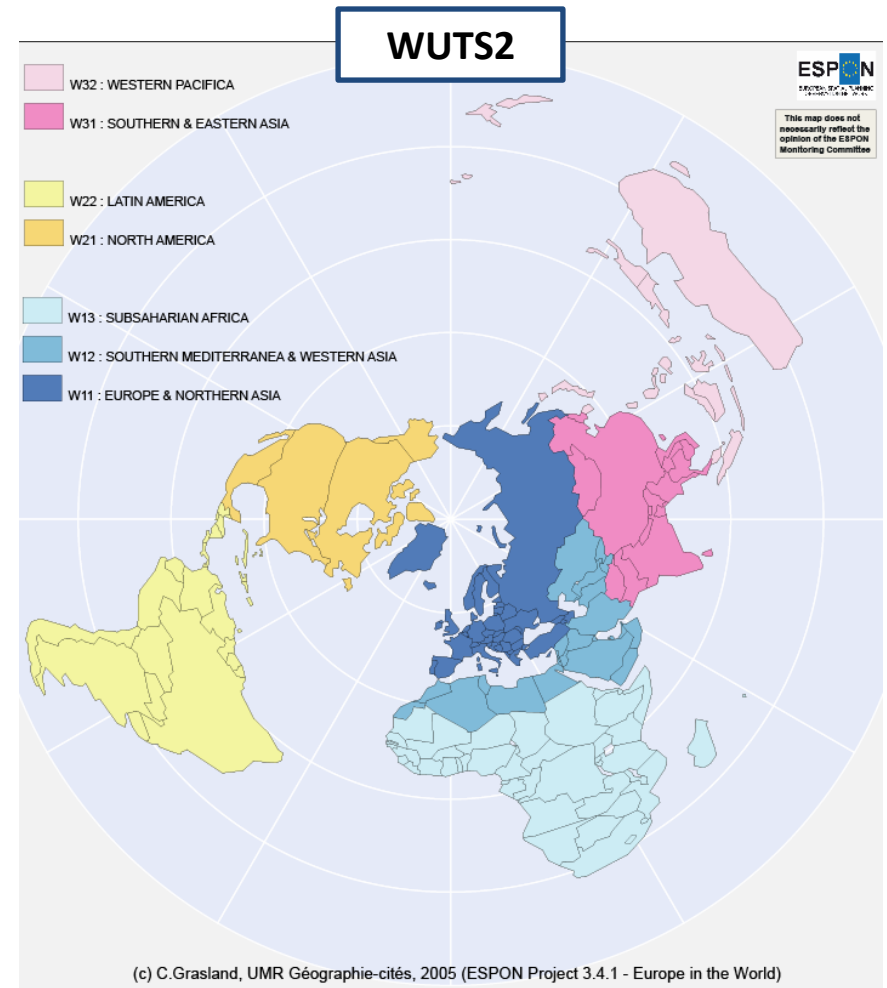
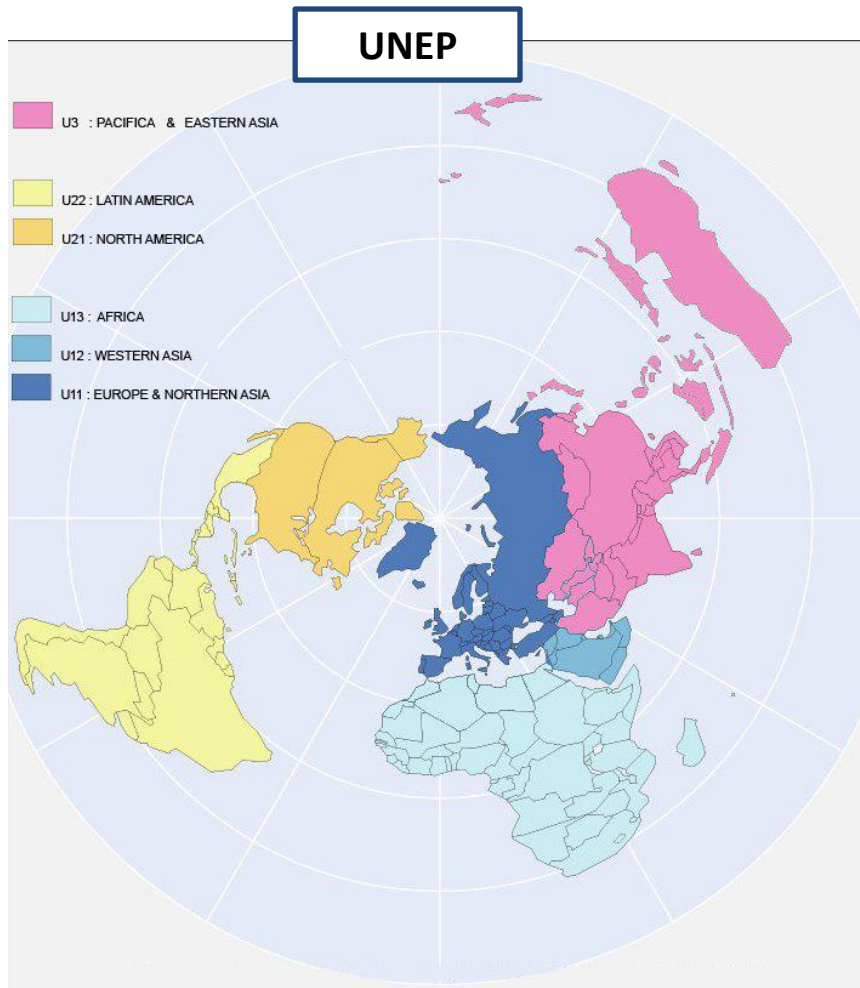
$p = 0$



# Comparaison des agrégats



# Comparaison des agrégats



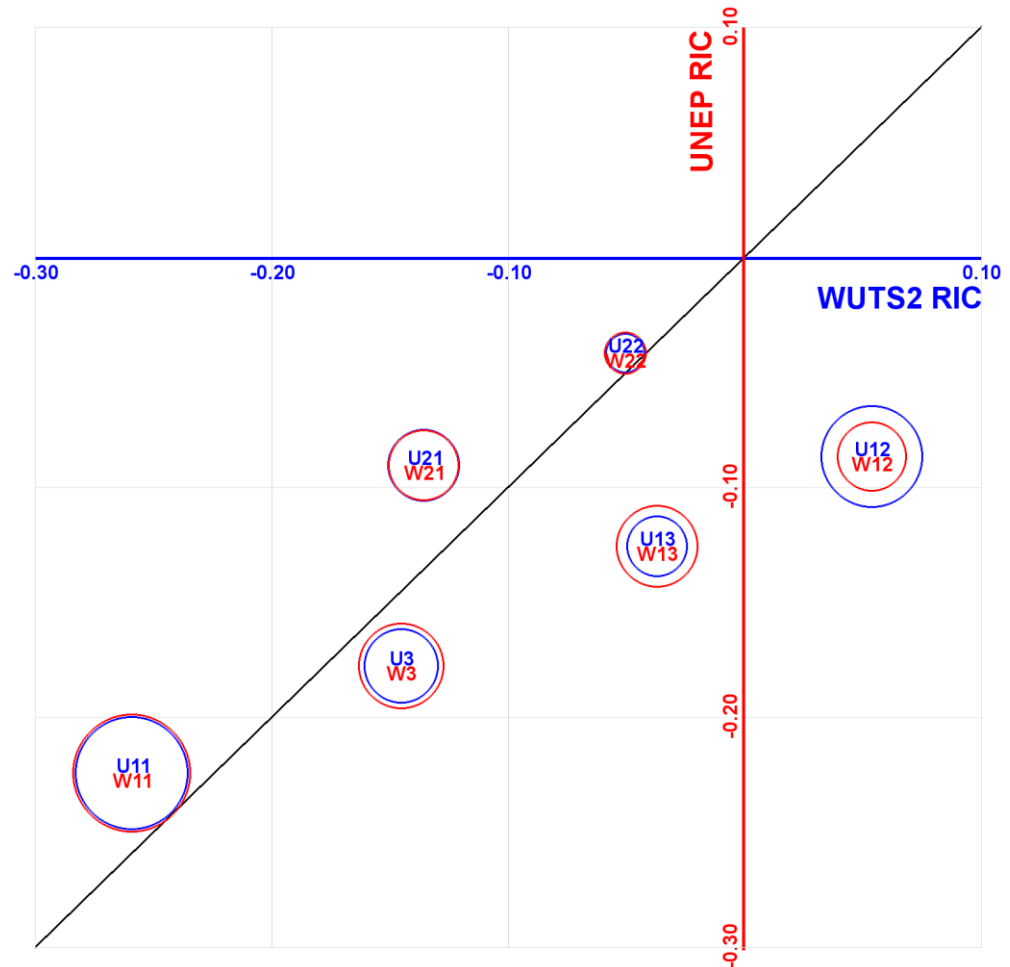
# Comparaison des agrégats



UNEP

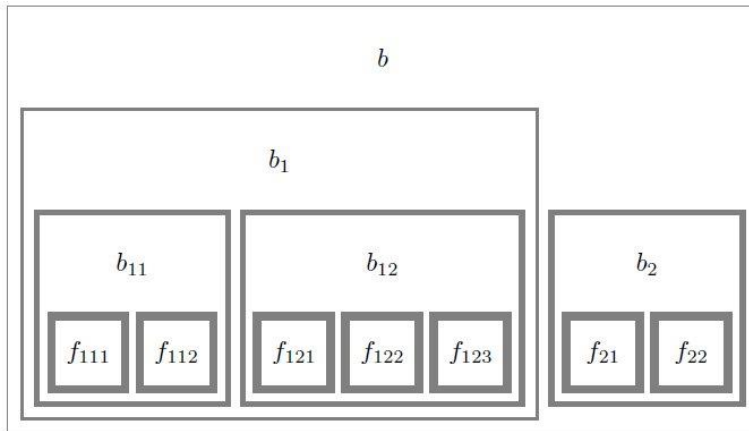


WUTS2

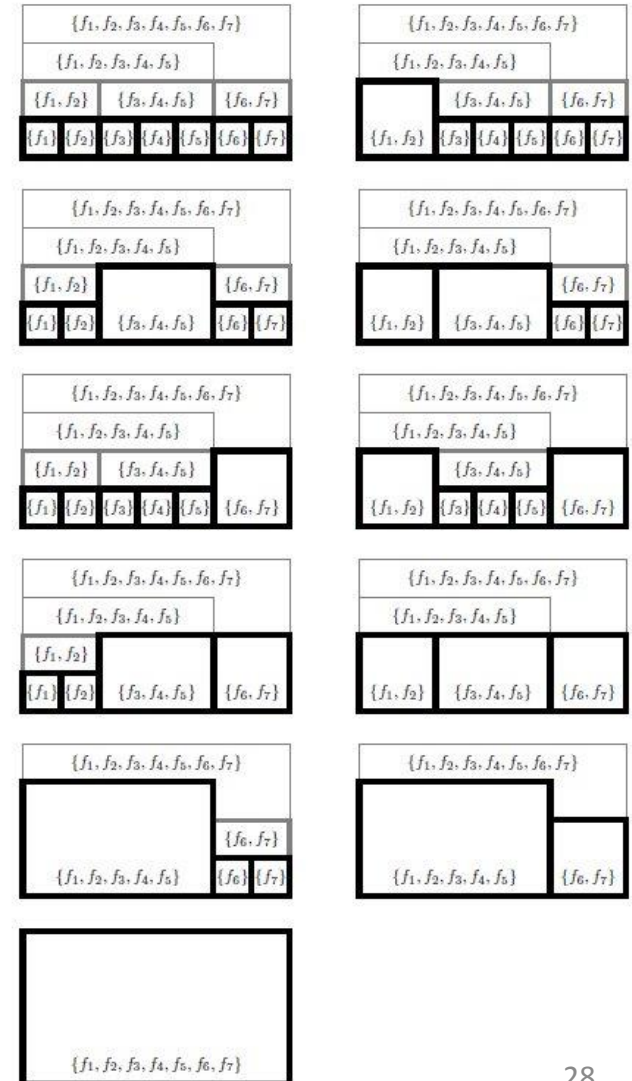


# Topologies et partitionnement

## Hiérarchie



Agrégations  
possibles



D'autres topologies  
intéressantes ?

# Bilan

- Méthode applicable
  - Aux approches exogènes (pour évaluer les capteurs)
  - Aux approches endogènes (pour contrôler la granularité)