

## J3 : Convergence apprentissage et sémantique

Construction de graphes à partir des variables décrivant  
l'environnement et la biodiversité

R. DAVID



## **J3 : Convergence apprentissage et sémantique**

Construction de graphes à partir des variables décrivant  
l'environnement et la biodiversité

R. DAVID

Sources de ces diapo :

- Ateliers curation de données (FRB, GBIF, IMBE)
- Ateliers visualisation de données (FRB, GBIF, IMBE)
- Soutenance De la conception d'un système d'observation à large échelle au déploiement et à l'exploitation de son système d'information : application à l'observation des habitats coralligènes et à la colonisation de récifs artificiels (ARMS)
- Rapport des ateliers

Introduction sur les inférences et les liens de causalité

**1/ Construction de systèmes d'observation à large échelle**

**2/ Construction de systèmes d'information pour les systèmes d'observation à large échelle**

**3/ Discussion sur la fouille de données basée sur les graphes  
(ou ses balbutiements dans les domaines de la biodiversité)**

Recommandations et perspectives

## Les objectifs du système d'observation (SO)

**Produire des données réutilisables** et  
Optimiser leur coût / répétabilité  
Permettre le déploiement à large échelle  
de cette production des données

### Principes :

Des protocoles réalistes, simples et explicites

Des observateurs, des opérateurs formés et inter-calibrés,

Une gestion de la qualité de la donnée pour améliorer itérativement les protocoles

Demande de gérer des données de contextes !

## Les objectifs du système d'observation (SO)

**Produire des données réutilisables** et  
Optimiser leur coût / répétabilité  
Permettre le déploiement à large échelle  
de cette production des données

### Principes :

Des protocoles réalistes, simples et explicites

Des observateurs, des opérateurs formés et inter-calibrés,

Une gestion de la qualité de la donnée pour améliorer itérativement les protocoles

Demande de gérer des données de contextes !

**Sources et méthodes d'enrichissement ???**

Les objectifs du système d'information (SI)

**Rendre des données réutilisables** et  
Optimiser leur accessibilité  
Permettre le développement de  
nouvelles méthodes d'utilisation efficace des données

### Principes :

Des bases de données ouvertes, interopérables, stables (et donc recensées) et des données traçables et bien qualifiées,

en adaptant les méthodes issues d'autres disciplines

**Astronomie, Physique, Mathématiques, Economie,  
Informatique, Fouille de données...**

→ INTEROPERABLES et MULTIDISCIPLINAIRES

# A large échelle : quels enjeux sur la biodiversité?



La dernière “une” de l’AFB

à titre d’exemple,  
en une génération  
30 % d’oiseaux en moins  
80 % d’insectes en moins



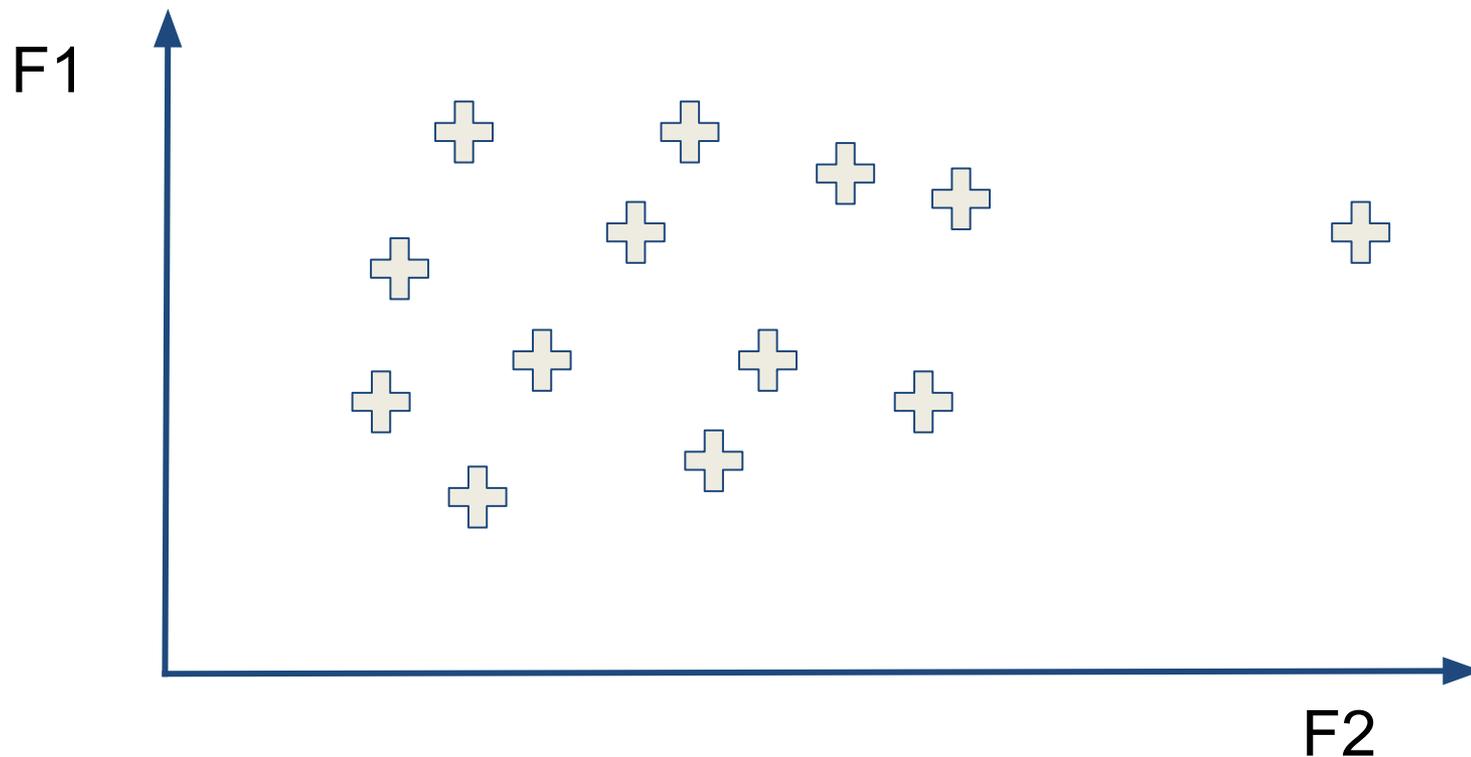
Traiter et représenter plus efficacement les données

**Systemes** non centralisé, open source, open data

Par quoi commencer : les statistiques ou la visualisation ?

# A large échelle : que cherche-t-on ?

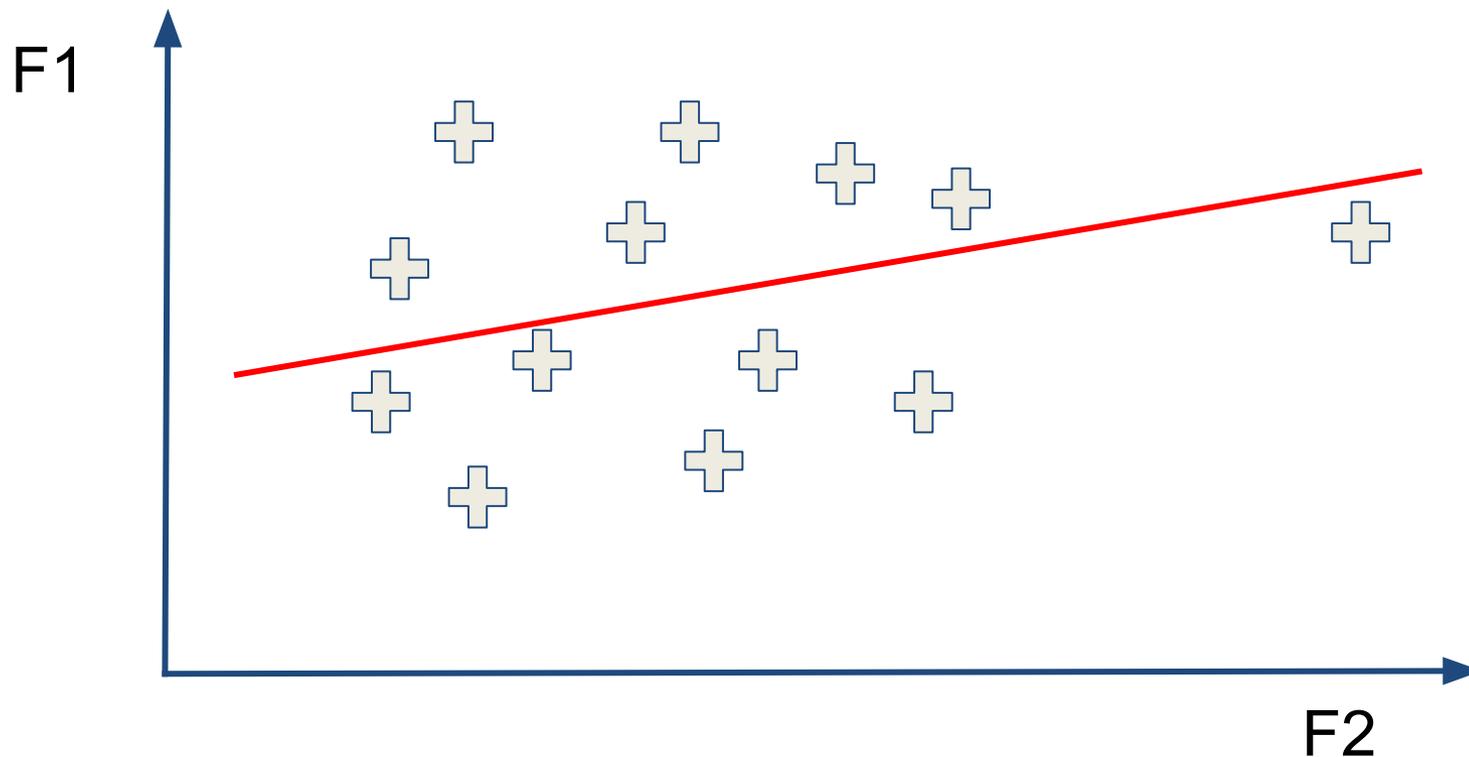
De manière classique, en science : recherche des “inférences”



## A large échelle : que cherche-t-on ?

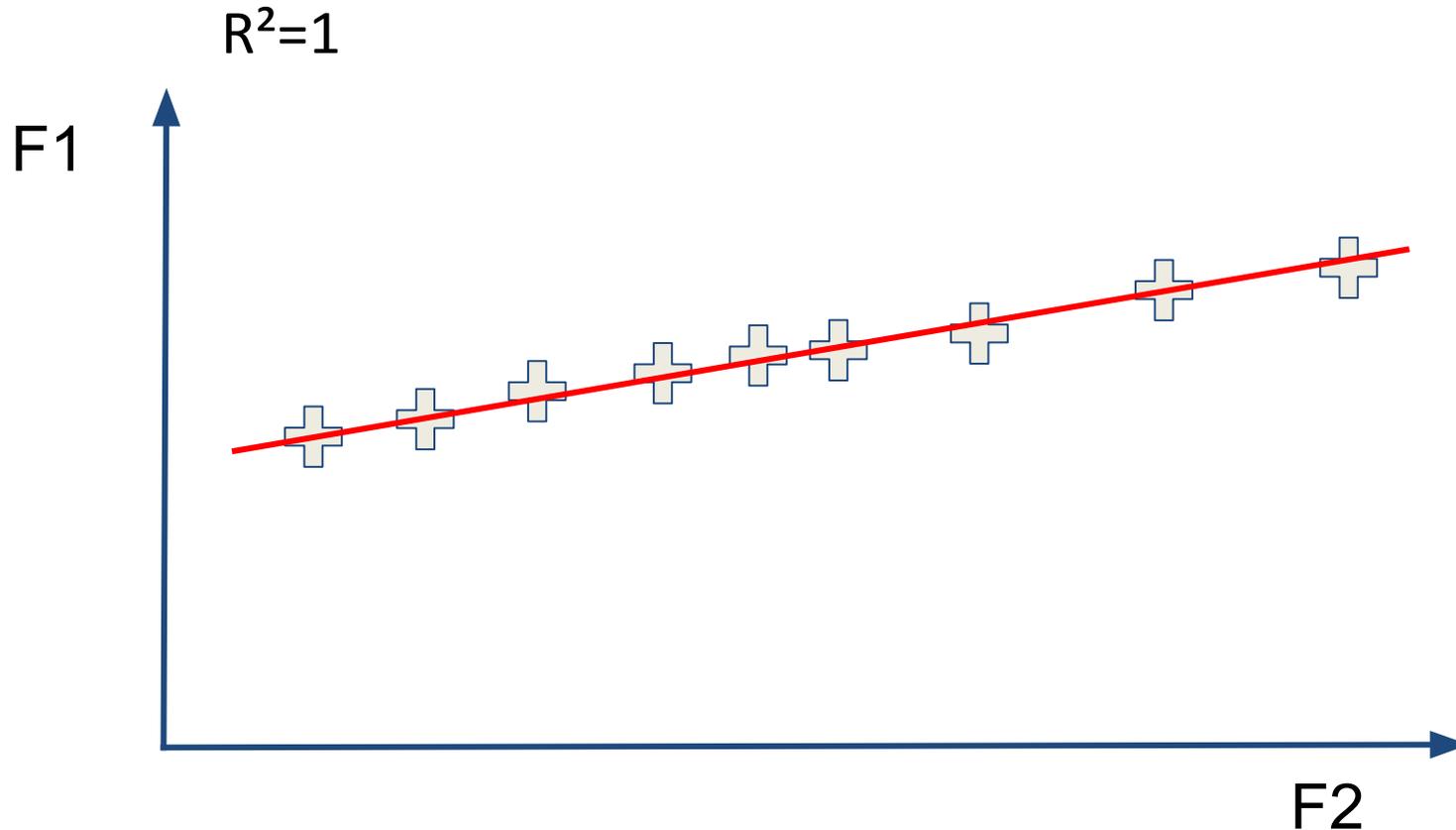
De manière classique, en science : recherche des “inférences”

$R^2$  = variabilité expliquée par la régression / variabilité totale



# A large échelle : que cherche-t-on ?

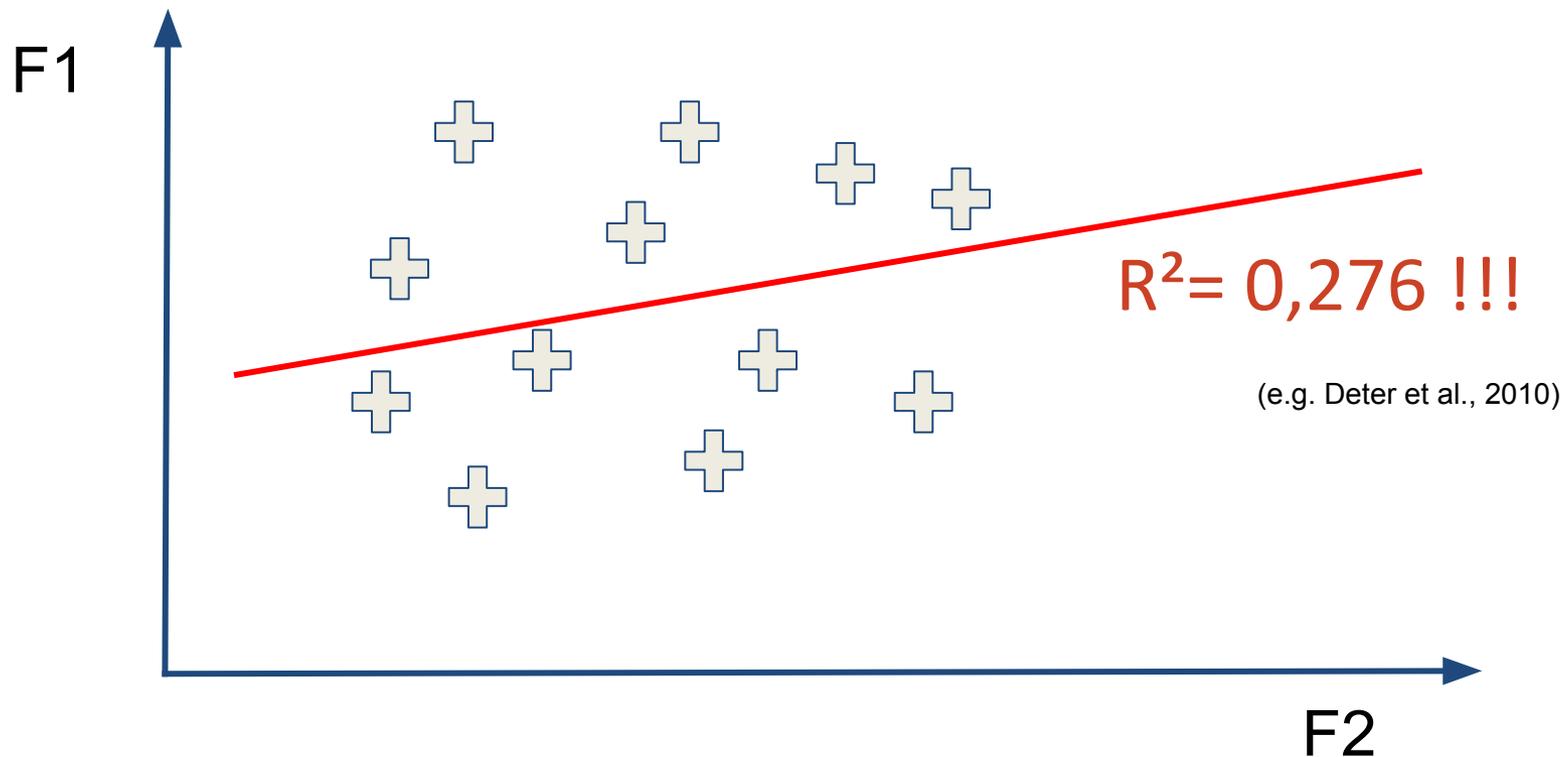
De manière classique, en science : recherche des “inférences”



## A large échelle : que cherche-t-on ?

De manière classique, en science : recherche des “inférences”

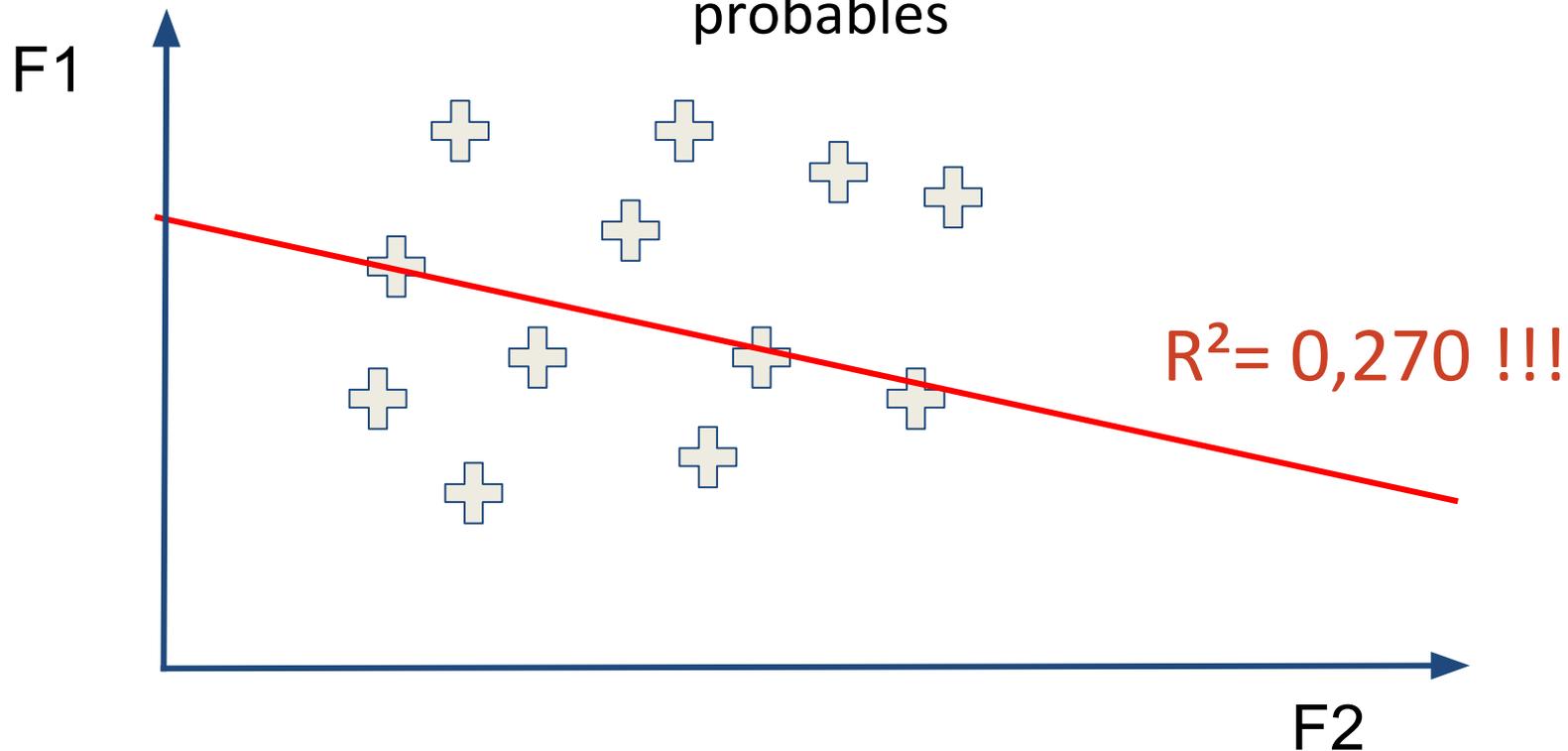
$R^2$  = parfois pas si proche de 1 pour les études en écologie



## A large échelle : que cherche-t-on ?

De manière classique, en science : recherche des “inférences”

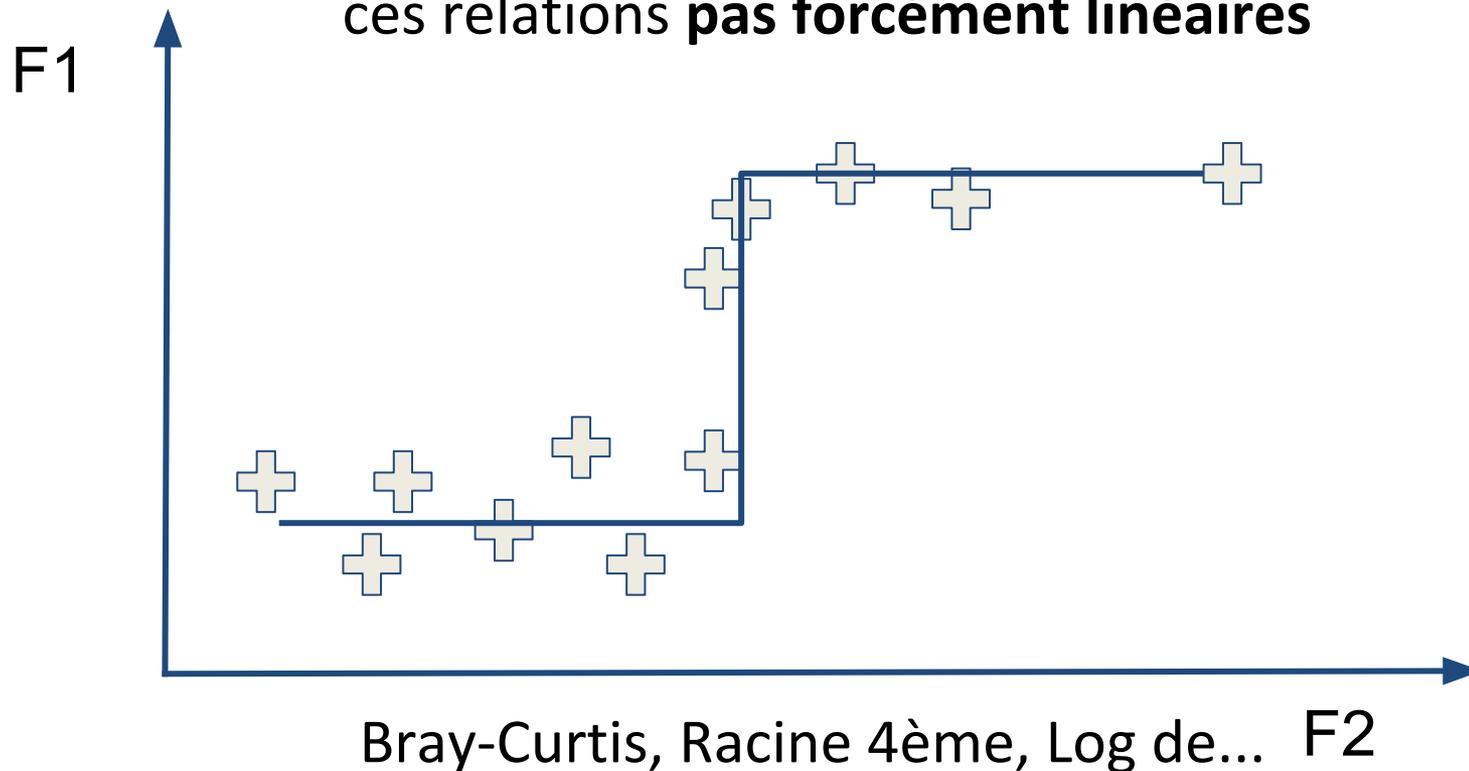
Avec parfois des explications contraires pratiquement aussi probables



## A large échelle : que cherche-t-on ?

De manière classique, en science : la recherche des “inférences”

Les distances sont souvent transformées pour mettre en évidence ces relations **pas forcément linéaires**



# Visualisation et statistiques

Quartet d'Anscombe

I		II		III		IV	
x	y	x	y	x	y	x	y
10,0	8,04	10,0	9,14	10,0	7,46	8,0	6,58
8,0	6,95	8,0	8,14	8,0	6,77	8,0	5,76
13,0	7,58	13,0	8,74	13,0	12,74	8,0	7,71
9,0	8,81	9,0	8,77	9,0	7,11	8,0	8,84
11,0	8,33	11,0	9,26	11,0	7,81	8,0	8,47
14,0	9,96	14,0	8,10	14,0	8,84	8,0	7,04
6,0	7,24	6,0	6,13	6,0	6,08	8,0	5,25
4,0	4,26	4,0	3,10	4,0	5,39	19,0	12,50
12,0	10,84	12,0	9,13	12,0	8,15	8,0	5,56
7,0	4,82	7,0	7,26	7,0	6,42	8,0	7,91
5,0	5,68	5,0	4,74	5,0	5,73	8,0	6,89

Ces 4 jeux de données ont été construits en 1973 par le statisticien [Francis Anscombe](#)

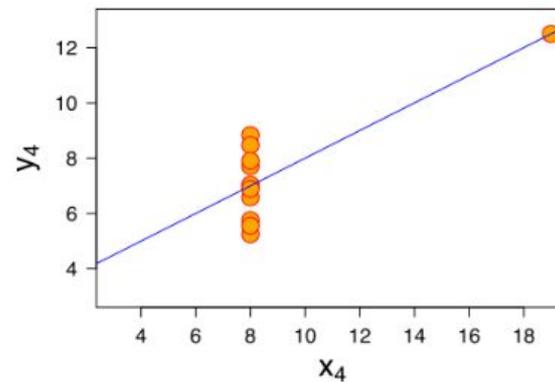
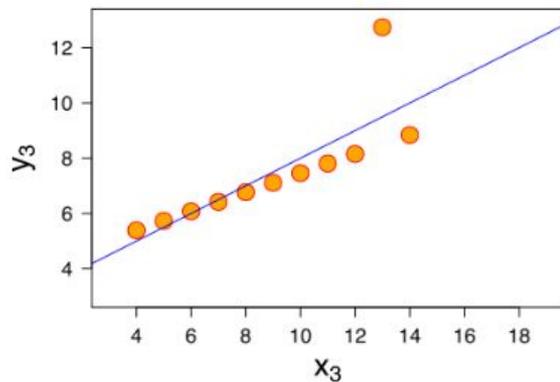
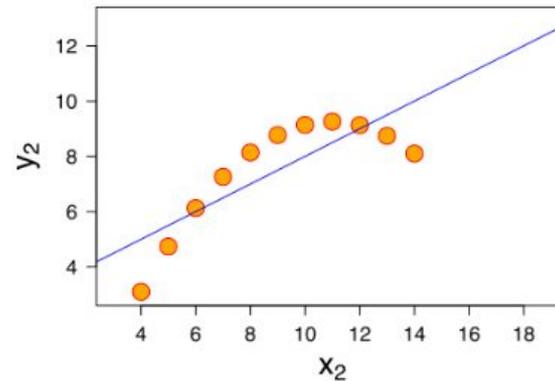
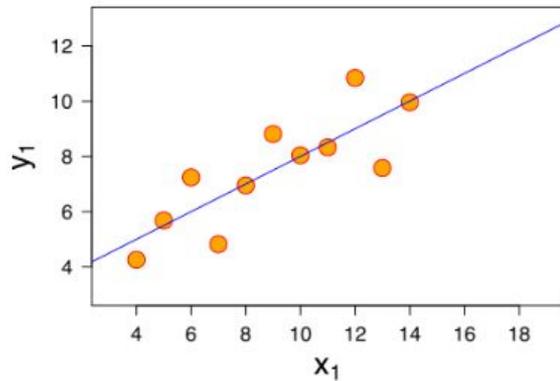
## Visualisation et inférences statistiques

Propriété	Valeur
Moyenne des $x$	9,0
Variance des $x$	10,0
Moyenne des $y$	7,5
Variance des $y$	3,75
Coefficient de <b>corrélation</b> entre les $x$ et les $y$	0,816
Équation de la droite de <b>régression linéaire</b>	$y=3 + 0,5 x$
Somme des carrés des erreurs relativement à la moyenne	110,0

[https://fr.wikipedia.org/wiki/Quartet\\_d%27Anscombe](https://fr.wikipedia.org/wiki/Quartet_d%27Anscombe)

# Visualisation et statistiques

Ce que ces statistiques ne montrent pas



**$R^2 = 0,816$  dans les 4 cas !**

Le but du mathématicien est de démontrer l'importance de tracer des graphiques avant d'analyser des données, notamment pour apprécier l'incidence des données aberrantes.

## A large échelle : que cherche-t-on ?

De manière classique, en science : des “inférences”

Ces inférences sont elles suffisantes pour comprendre des liens de causalité ? Pas souvent (en tout cas dans le domaine de la biodiversité) :

F1 a un effet sur F2    **ou**    F2 a un effet sur F1?

## A large échelle : que cherche-t-on ?

De manière classique, en science : des “inférences”

Ces inférences sont elles suffisantes pour comprendre des liens de causalité ? Pas souvent :

F1 a un effet sur F2    **ou**    F2 a un effet sur F1?

**ou**    F3 a un effet sur F1 et F2

## A large échelle : que cherche-t-on ?

De manière classique, en science : des “inférences”

Ces inférences sont elles suffisantes pour comprendre des liens de causalité ? Pas souvent :

F1 a un effet sur F2    **ou**    F2 a un effet sur F1?

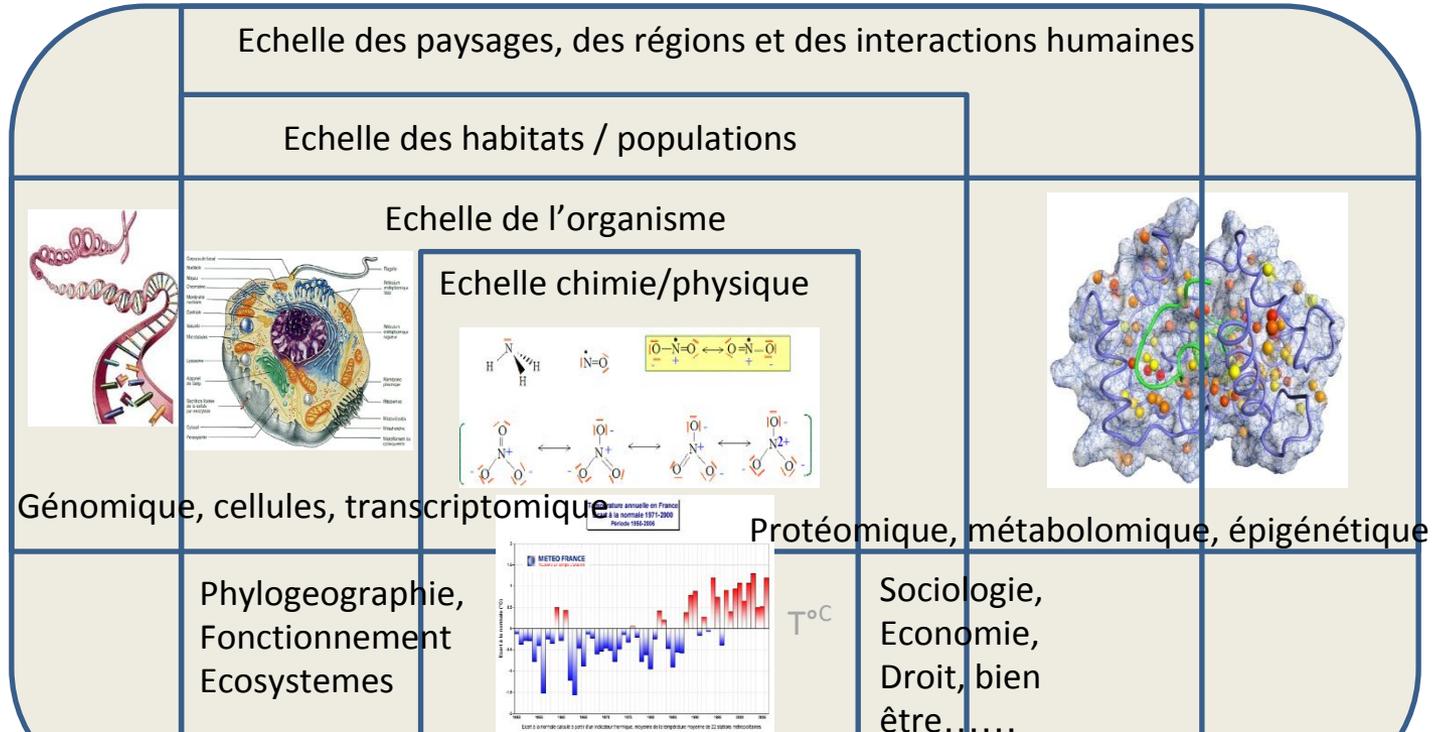
**ou**    F3 a un effet sur F1 et F2

**ou**    F3 et F4 à partir d'un dosage de chacun,  
agissant sur F1 et F2

**ou**    souvent bien plus complexe...  
nécessitant des approches “systémiques”

# Les données sur la biodiversité et l'environnement ?

Multi-échelles – Multi formats – Multi-sources...

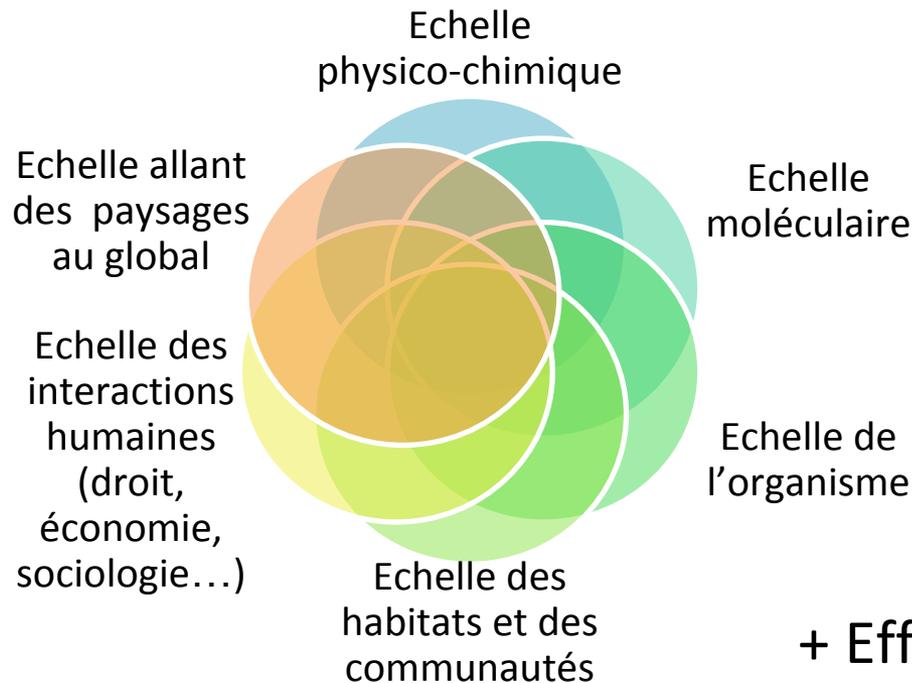


Quelle approche intégrative pour ce patchwork de contextes à chaque niveau d'organisation?

Ces facteurs peuvent être considérés via de nombreux prismes d'observation

# Le contexte : des données très « sectorielles »

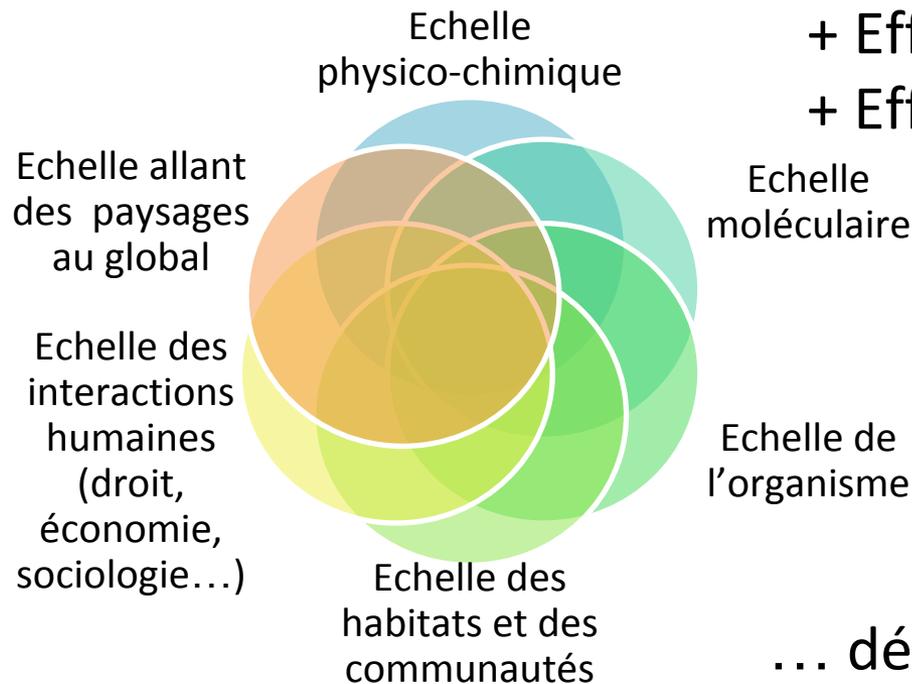
Pourtant, il faut décrire  
un système fortement interconnecté et interdépendant



- + Effets antagonistes
- + Effets « potentialisateurs »
- + Effets de cascades

# Le contexte : des données très « sectorielles »

Pourtant, il faut décrire  
un système fortement interconnecté et interdépendant



- + Effets antagonistes
- + Effets « potentialisateurs »
- + Effets de cascades

... dépendants des  
contextes

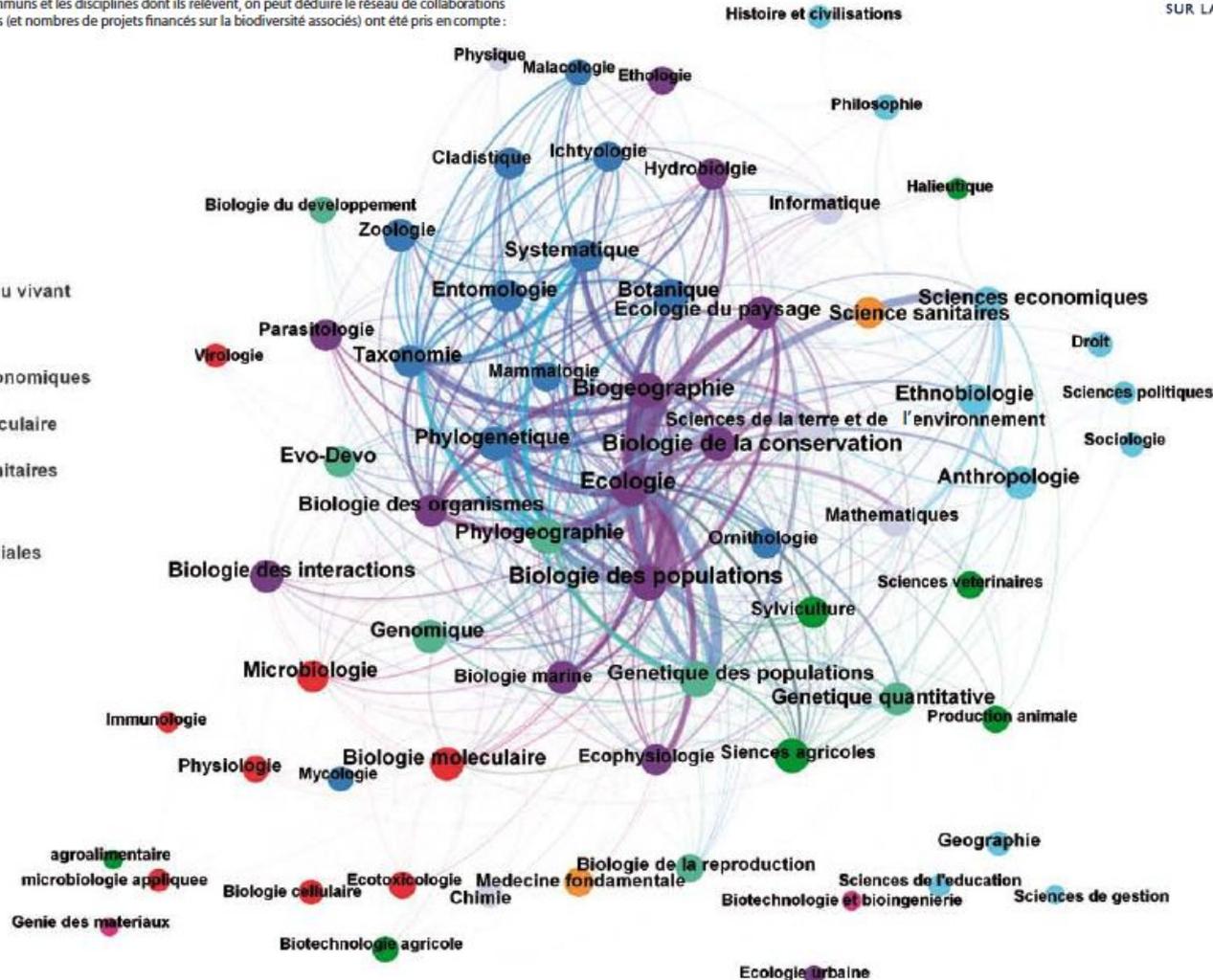
## Interdisciplinarité : des disciplines encore très compartimentées

### La recherche sur la biodiversité en France Réseau de collaborations entre disciplines

Les acteurs de la base de données FRB sont associés à des disciplines et à des projets de recherche sur la biodiversité. En considérant les chercheurs qui participent à des projets communs et les disciplines dont ils relèvent, on peut déduire le réseau de collaborations entre les disciplines. Les appels d'offres suivants (et nombres de projets financés sur la biodiversité associés) ont été pris en compte :

ANR-6extinction-2009 (12)  
ANR-biodiv-2005 (15)  
ANR-biodiv-2006 (15)  
ANR-biodiv-2007 (16)  
BRG-RG-2005 (30)  
BRG-RG-2007 (19)  
FRB-CESAB-2010 (4)  
FRB-innov-2009 (53)  
FRB-scenario-2010 (4)  
IFB-OcéanIndien-2007 (17)

- Ecologie
- Classes et classification du vivant
- Génétique et évolution
- Sciences agricoles et agronomiques
- Biologie cellulaire et moléculaire
- Sciences médicales et sanitaires
- Sciences exactes
- Sciences humaines et sociales
- Sciences de l'ingénieur et technologiques



## Problématiques des systèmes d'observation à large échelle

- Importance de la donnée historique et de sa nécessaire prise en compte
- Multiplicité des systèmes d'observation préexistants ayant tous leurs particularités
- Nécessaire multiplication des observateurs et opérateurs ayant des compétences / expériences différentes
- Besoin de rationalisation des recueils des données (et donc de les regrouper, les rendant ainsi plus complexes)
- Objets observés parfois foncièrement différents

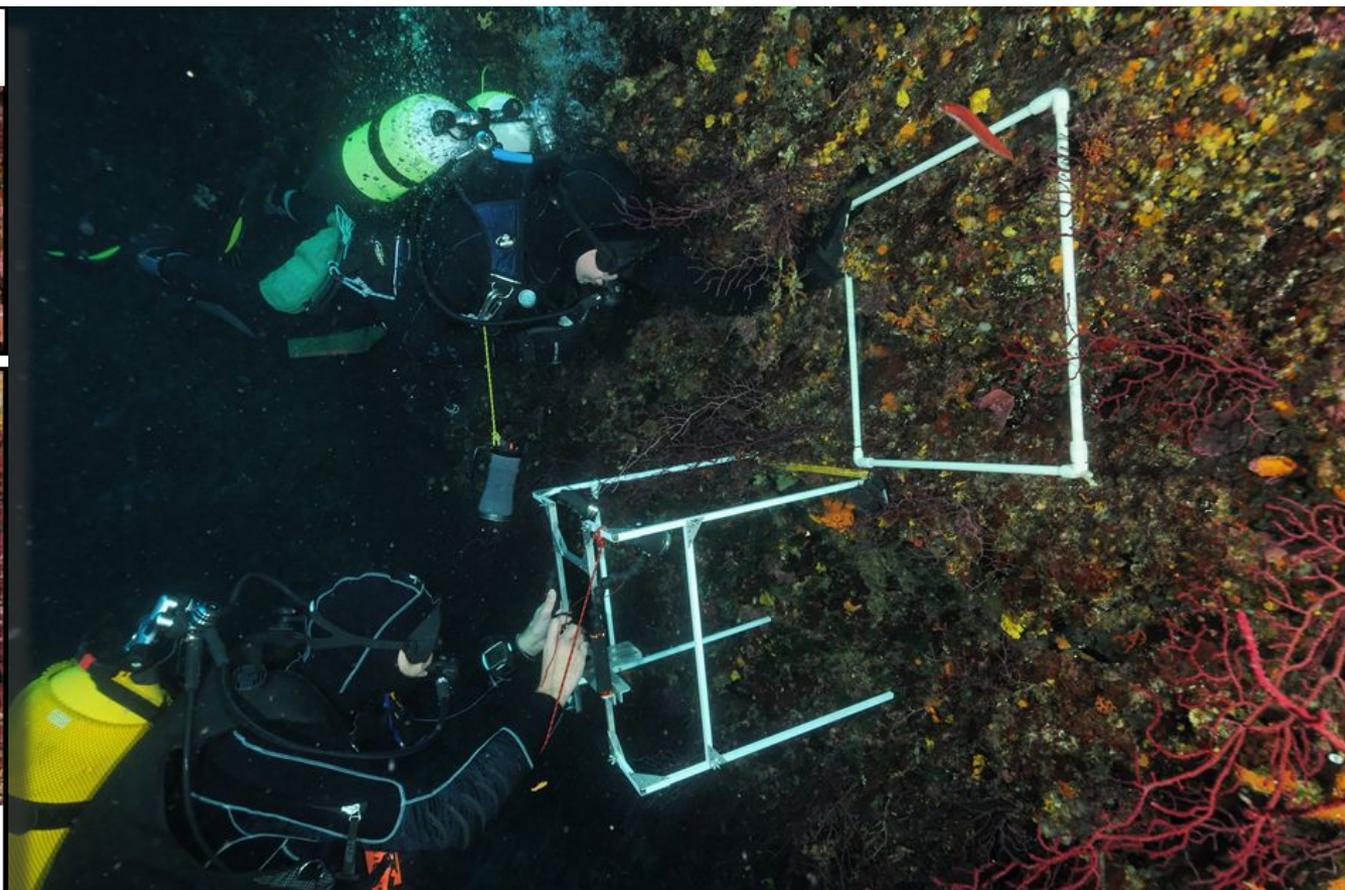
### 3) Sélection de résultats sur les systèmes d'observation

Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience?

Rhodophytes calcaires

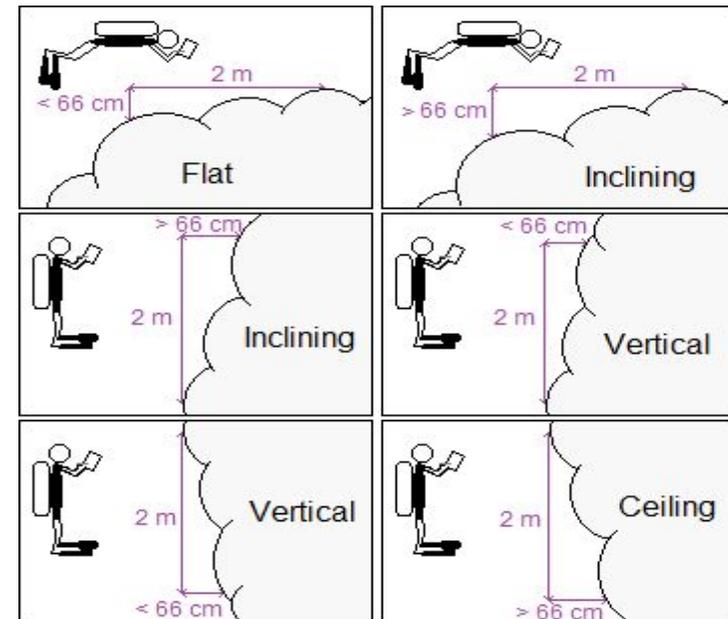
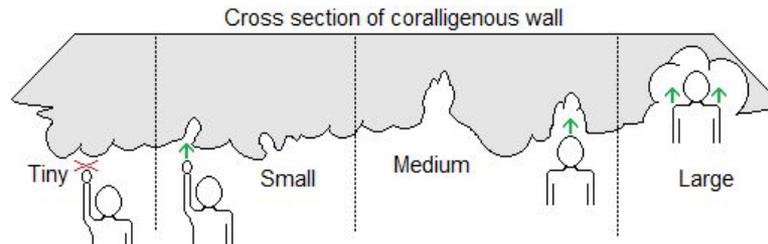
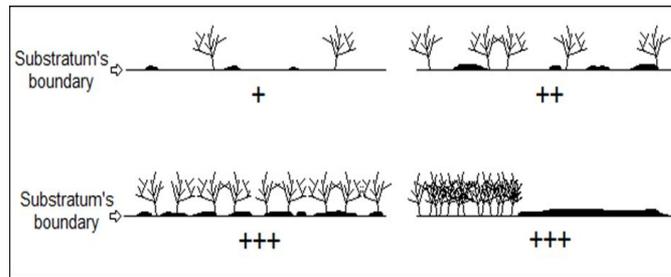


Coralligène à *Paramuricea clavata*, sur paroi verticale



### 3) Sélection de résultats sur les systèmes d'observation

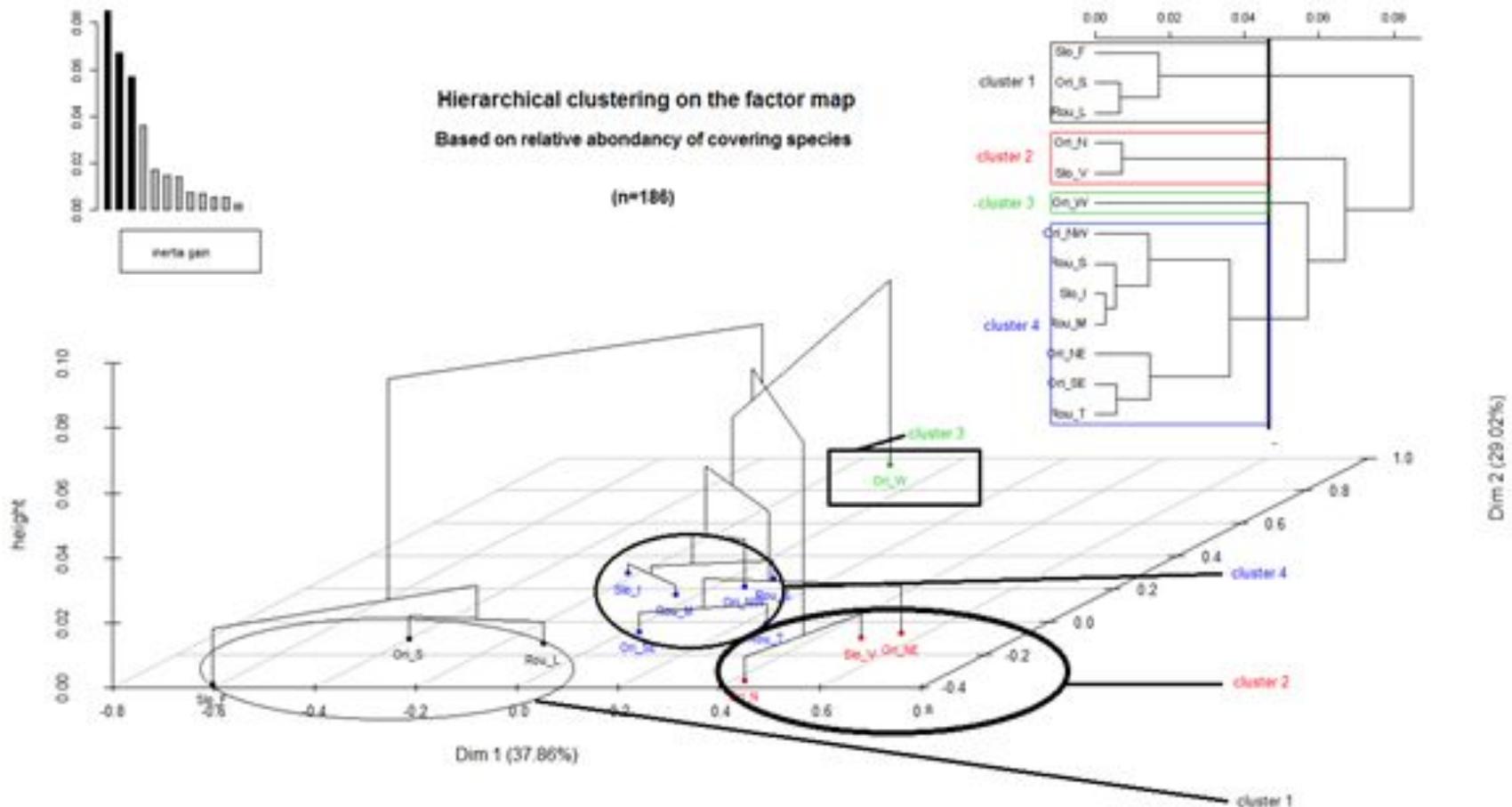
Les assemblages dominants sont localement expliqués par quelques variables très simples (pente, rugosité, orientation, profondeur)



Variables très simples mais pas si simples à mesurer !

### 3) Sélection de résultats sur les systèmes d'observation

Les assemblages dominants sont localement expliqués par quelques variables très simples (pente, rugosité, orientation, profondeur)



### 3) Sélection de résultats sur les systèmes d'observation

Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience ?

Sur la « librairie » commune

-> cas de CIGESMED (concerne le coralligène)

-> cas de DEVOTES (concerne les dispositifs ARMS)

#### Relevés de recouvrement CIGESMED

Site: ..... Date: .../.../..... Depth: ..... m Observer: .....

N°	ORI	INC	RUG	ES	EC	PC	CR	ERA	FRA	EGA	FGA	Sponge	Codium	Bryozoa	Ascidac	Others	Remark ind	Remark pop
1	Orientation																	
2		inclinaison																
3																		
4																		



### 3) Sélection de résultats sur les systèmes d'observation

Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience ?

Sur la « librairie » commune

-> cas de CIGESMED (concerne le coralligène)

-> cas de DEVOTES (concerne les dispositifs ARMS)

**La donnée brute  
(sur les fiches des plongeurs) n'est  
pas explicite (ORI, INC, RUG...) ->**

**Contraintes liées au travail de terrain**

#### Relevés de recouvrement CIGESMED

Site: ..... Date: .../.../..... Depth: ..... m Observer: .....

N°	ORI	INC	RUG	ES	EC	PC	CR	ERA	FRA	EGA	FGA	Sponge	Codium	Bryozoa	Ascidac	Others	Remark ind	Remark pop
1									+	++	+++							
2									1	2	3							
3																		
4																		

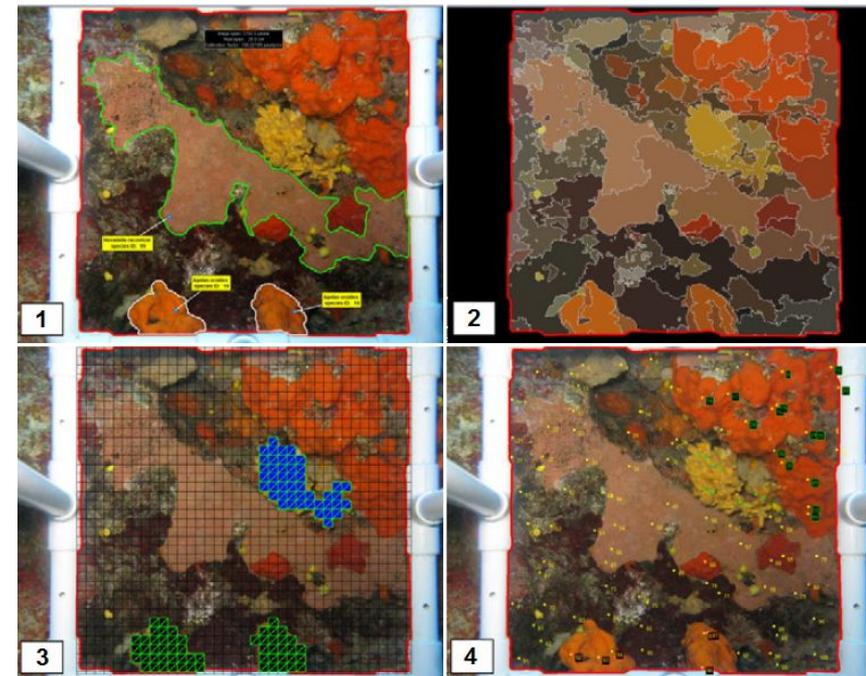
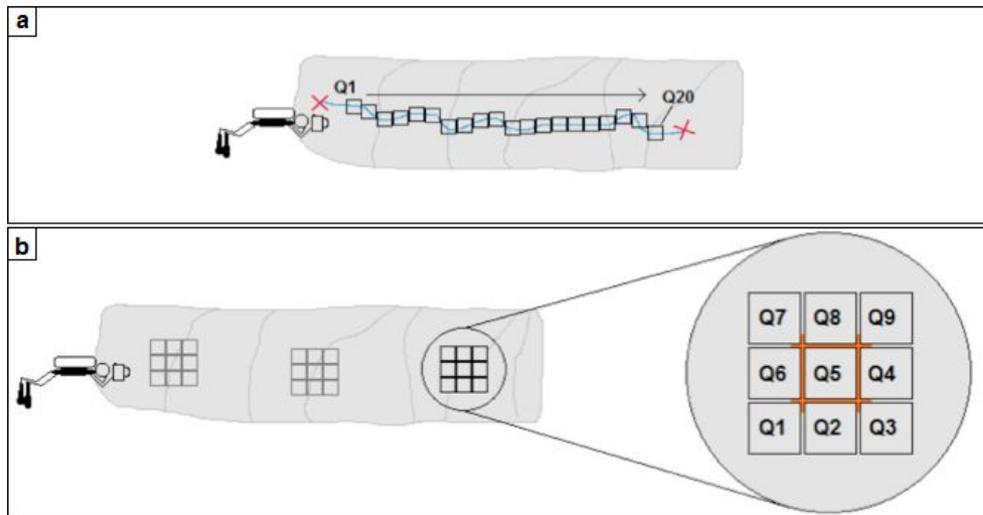
Inclinaison

Orientation

### 3) Sélection de résultats sur les systèmes d'observation

Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience?

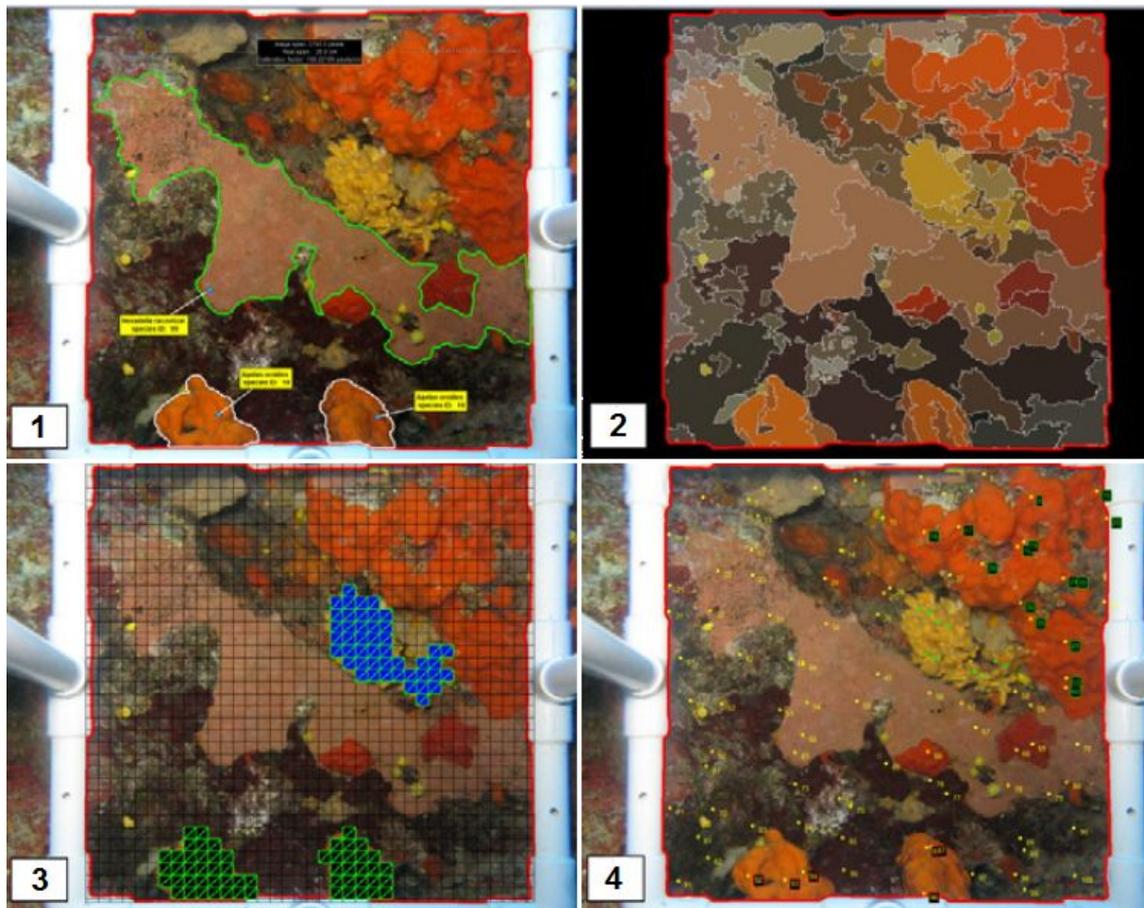
Sur la méthode d'échantillonnage



Exemple de traitement de photographies avec le logiciel PhotoQuad (Trygonis & Sini, 2012)

### 3) Sélection de résultats sur les systèmes d'observation

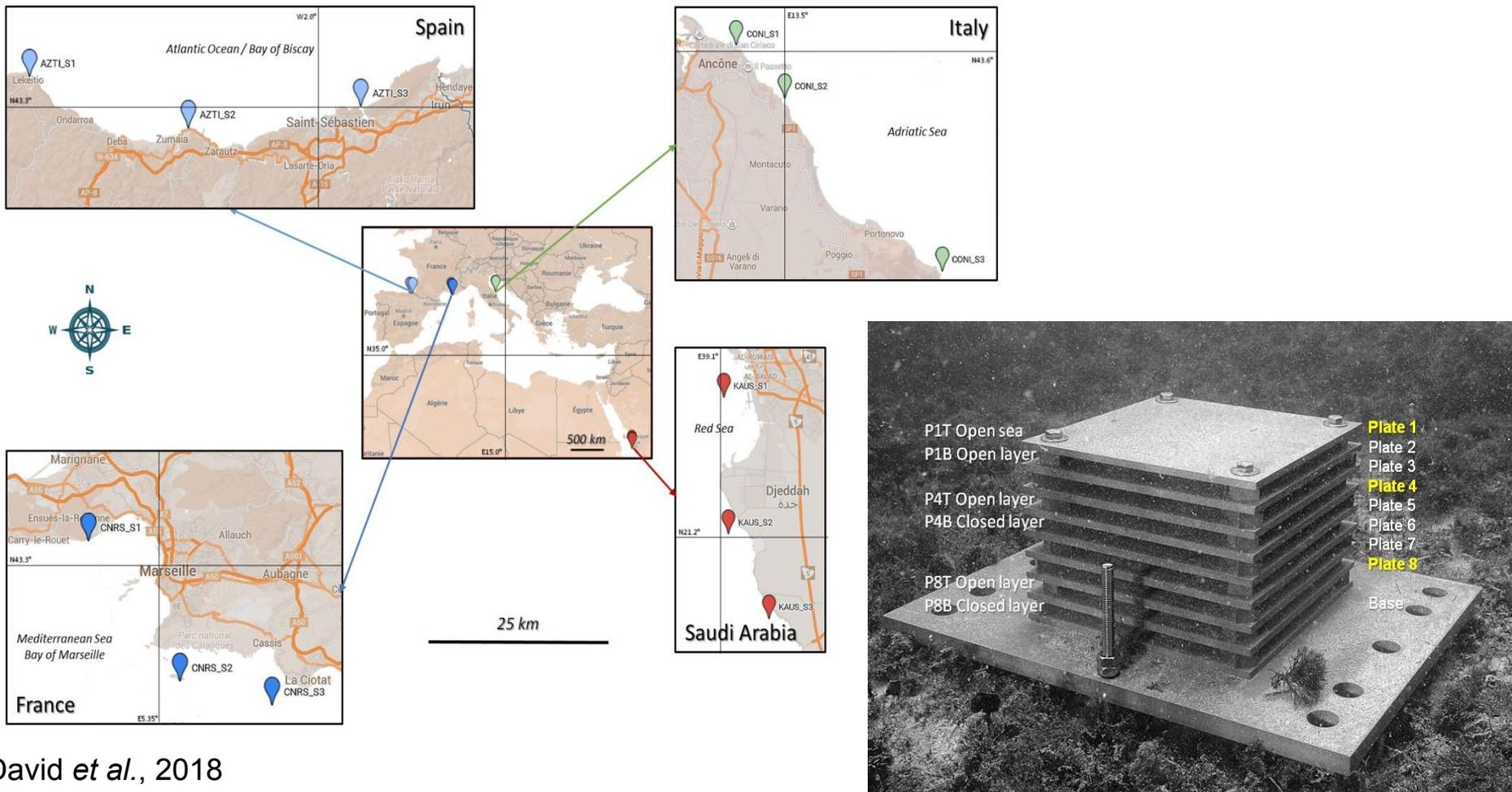
Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience?



Une base pour  
l'apprentissage ?

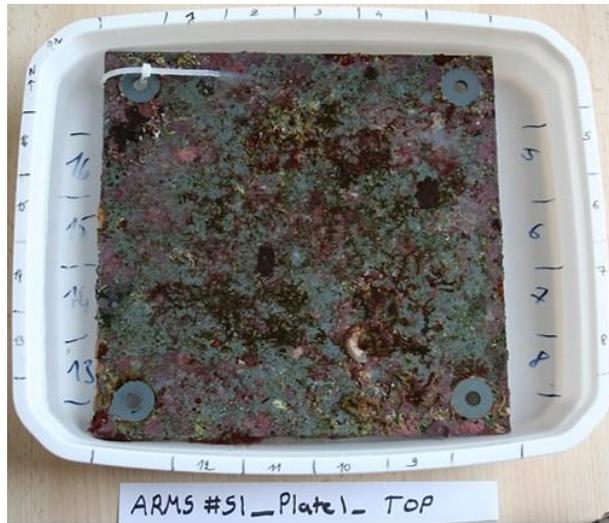
### 3) Sélection de résultats sur les systèmes d'observation

Comment cerner la variabilité à large échelle liée à des différences fondamentales / créer des typologies communes ?



### 3) Sélection de résultats sur les systèmes d'observation

Comment cerner la variabilité à large échelle liée à des différences fondamentales / créer des typologies communes ?



## Problématiques liées aux Systèmes d'Information des systèmes d'observation à large échelle

- Architecture et accessibilité
- Indexation et traçabilité
- Qualification des données – processus de curation
- Equivalence et interopération avec d'autres SI
- Cycle de vie des données et Data Management Plan

**Appliqués aux habitats coralligènes et aux récifs artificiels**

## Problématiques liées aux Systèmes d'Information des systèmes d'observation à large échelle

- Architecture et accessibilité
- Indexation et traçabilité
- Qualification des données – processus de curation
- Equivalence et interopération avec d'autres SI
- Cycle de vie des données et Data Management Plan

**DANS L'OBJECTIF D'UTILISER ou RÉUTILISER les DONNÉES**

**Appliqués aux habitats coralligènes et aux récifs artificiels**

## 2) Nos questionnements scientifiques liés au Système d'Information

Quelle architecture proposée pour rendre les données accessibles pour plusieurs usages ?

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?

**Questionnements appliqués aux données « habitats coralligènes » et ARMS**

## 2) Nos questionnements scientifiques liés au Système d'Information

Quelle architecture proposée pour rendre les données accessibles pour plusieurs usages ?

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité (Web sémantique, vais je arriver à t'utiliser?) ?

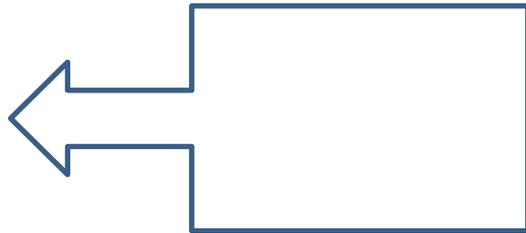
**Quelle qualification des données** a le meilleur impact sur le niveau d'usage (et donc sur la valeur des données) ?

**FOCUS : Comment représenter ces données hétérogènes et forcément imparfaites ?**

**Questionnements appliqués aux données « habitats coralligènes » et ARMS**

### 3) Sélection de résultats sur les systèmes d'information

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?



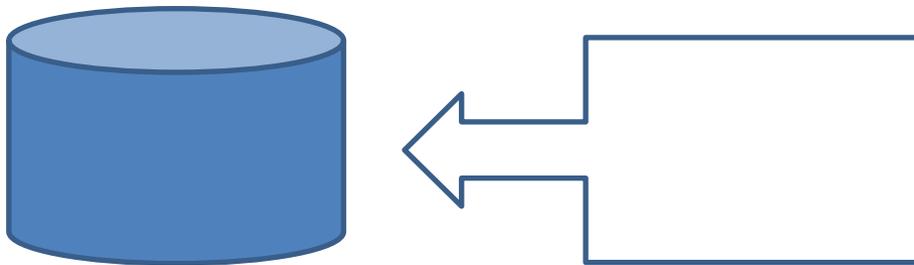
= un descripteur de données

### 3) Sélection de résultats sur les systèmes d'information

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?

## Des descripteurs d'acteurs comme lien

Niveau BDD

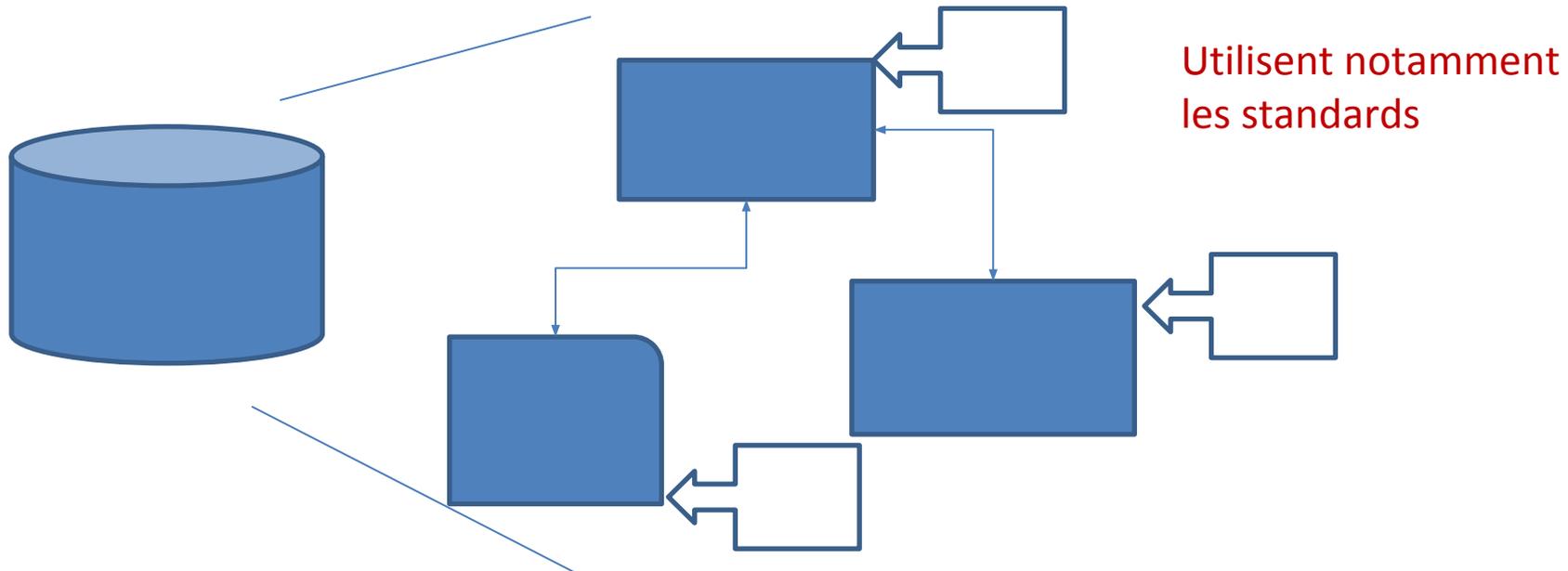


Notamment normes (ISO, INSPIRE...)

### 3) Sélection de résultats sur les systèmes d'information

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?

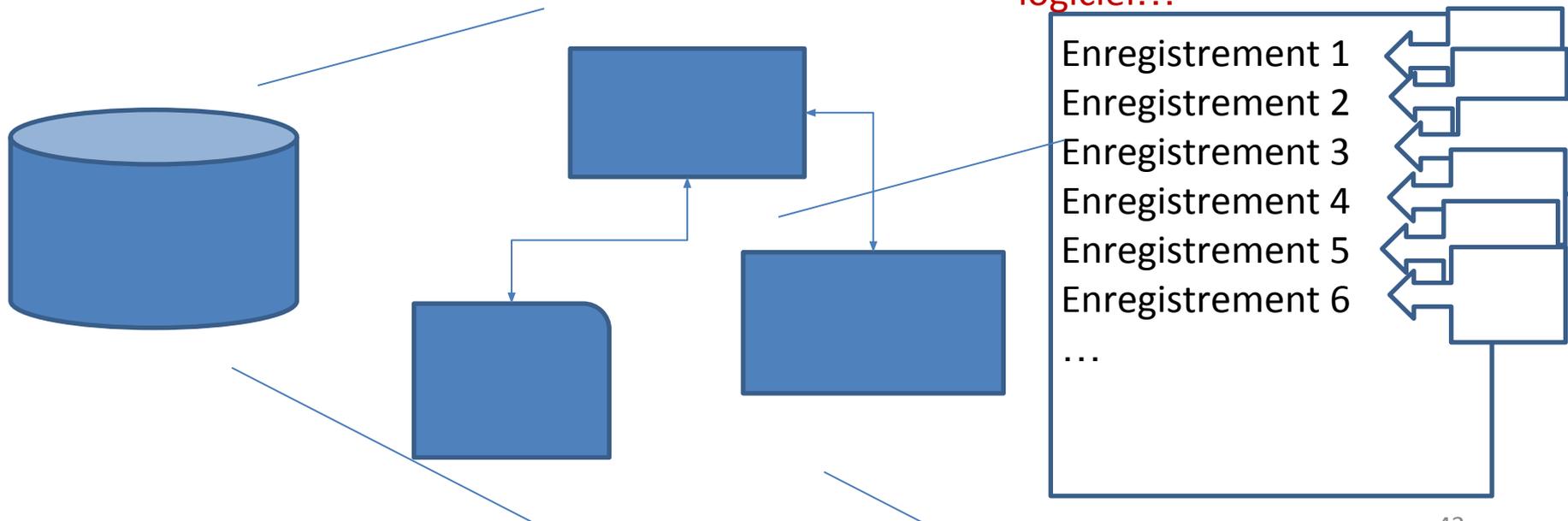
Niveau Tables: dictionnaire de données, mais pas uniquement (équivalences via thesaurus, règles d'intégrité)



### 3) Sélection de résultats sur les systèmes d'information

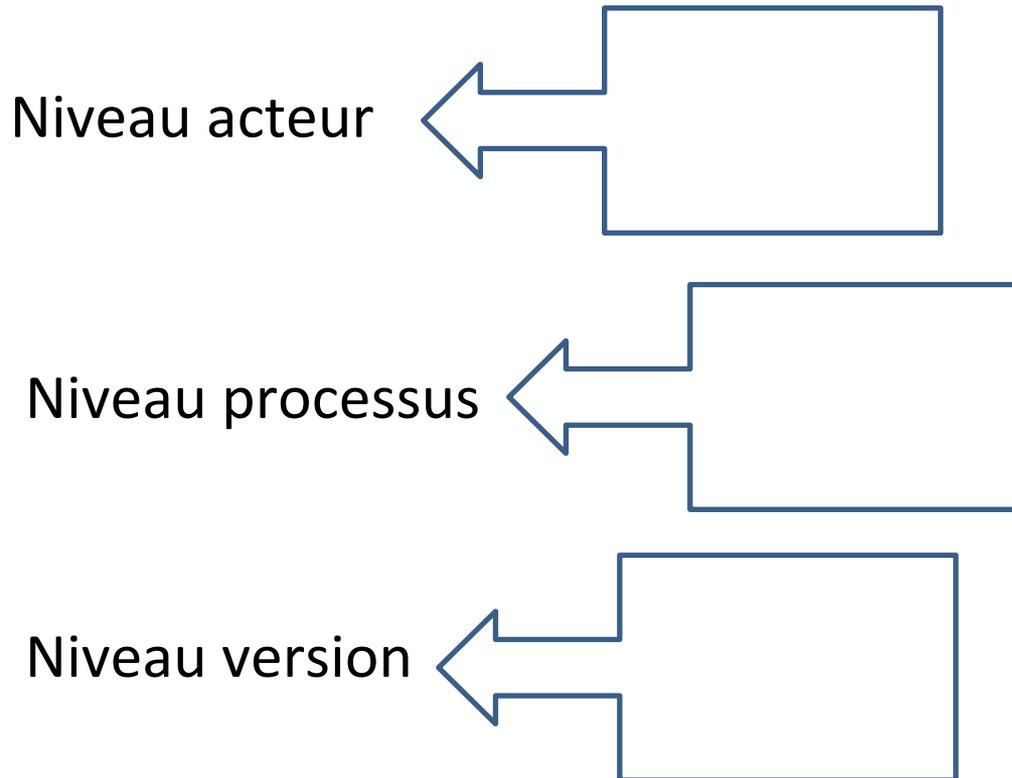
Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?

#### Niveau Enregistrement



### 3) Sélection de résultats sur les systèmes d'information

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?



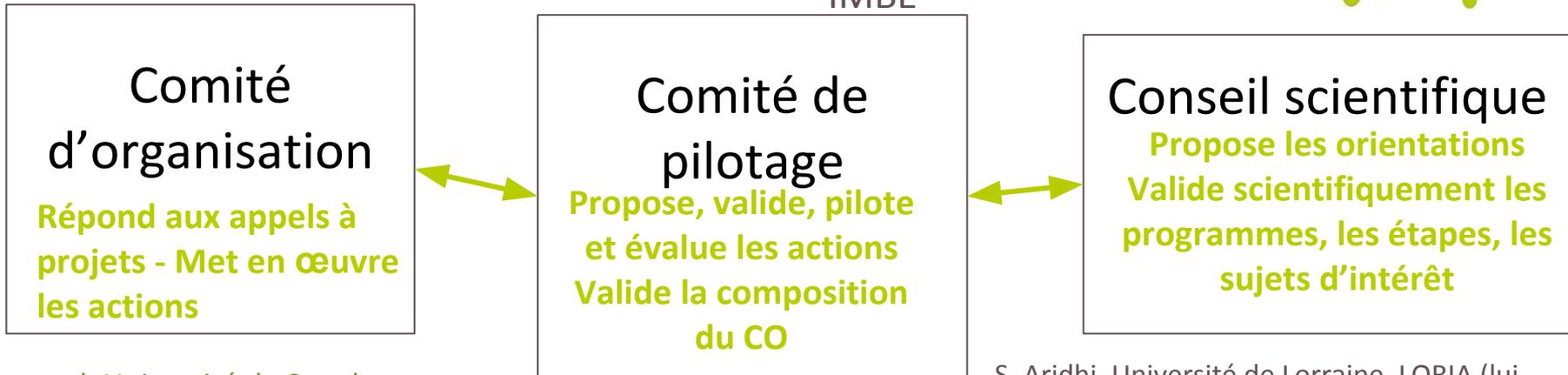
**RDF ???**



Traiter et représenter plus efficacement les données

**Rappel : non centralisé, open source, open data**

Proposition : utiliser les graphes pour explorer les données distantes et hétérogènes



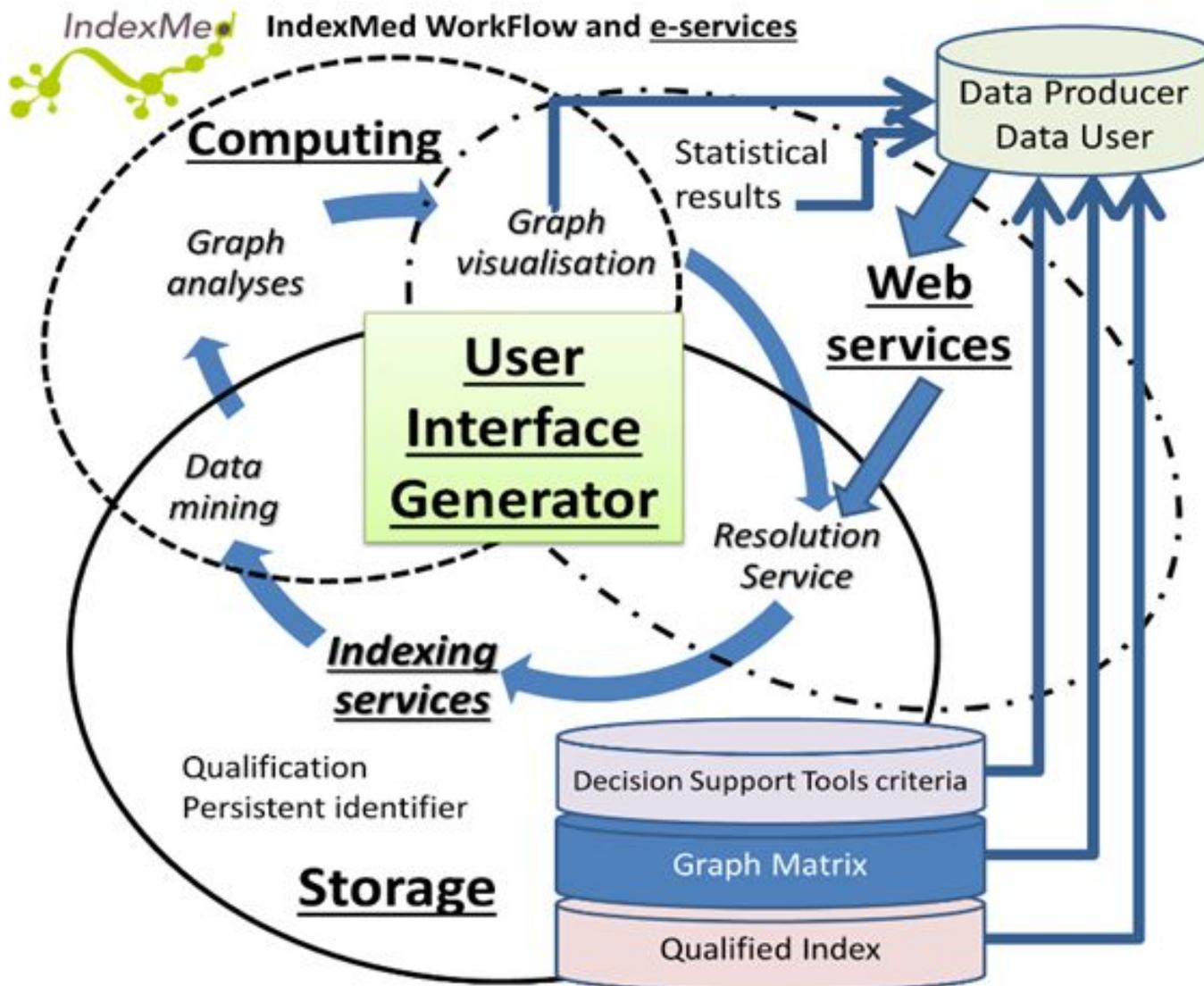
L. Bernard, Université de Strasbourg, ArkéoGis  
A. Cohen Nabeiro, FRB  
R. David, Université Aix-Marseille, IMBE  
L. Hogie, Inria  
L. Berti-Equille, Université Aix-Marseille, LIF  
R. Bourqui Université Bordeaux I, LaBRI (lui proposer)  
S Aridhi University of Lorraine , Nancy · LORIA (lui proposer)  
D. Leroux-Michelet, CNRS, OHM Vallée du Rhône (lui proposer)  
J-B. Mihoub, UPMC (lui proposer)  
S. Pamerlon, GBIF-Fr.  
J-C. Raynal, CNRS, ECOREV, OHM Bassin minier de Provence (lui proposer)  
**+ de STIC!!**

**Représentant du CS :** C. Surace, Aix-Marseille Université, LAM (lui proposer)  
**Représentant du CO :** A. Cohen Nabeiro, FRB  
C. Blanpain, OSU Pythéas  
E. Chenin, IRD, GBIF-Fr. (lui proposer)  
A. Delavaud, FRB  
S. Gachet, IMBE  
G. Romier, IDGCC  
T. Tatoni, IMBE  
Y. Le BRAS, MNHN – ex INRIA

**Composition non validée**  
- août 2017

S. Aridhi, Université de Lorraine, LORIA (lui proposer)  
D. Auber, Université Bordeaux I, LaBRI (lui proposer)  
N. Bailly, HCMR, OBIS, Lifewatch  
D. Couvet, MNHN (lui proposer)  
B. Cremilleux, ENSI Caen, GREYC (lui proposer)  
J-P. Féral, Université Aix-Marseille, IMBE  
M. Habib, Université de Paris-Diderot, IRIF (lui proposer)  
D. Ienco, IRSTEA  
R. Julliard, MNHN  
Y. Legre, EGI  
I. Mougnot, Université Montpellier  
C. Pichot, Inra, AnaEE-France (lui proposer)  
A. Specht, CESAB-FRB  
C. Surace, Aix-Marseille Université, LAM  
R. Viggne-Lebbes, MNHN

Une architecture issue de plusieurs jours d'échange en atelier à EGI

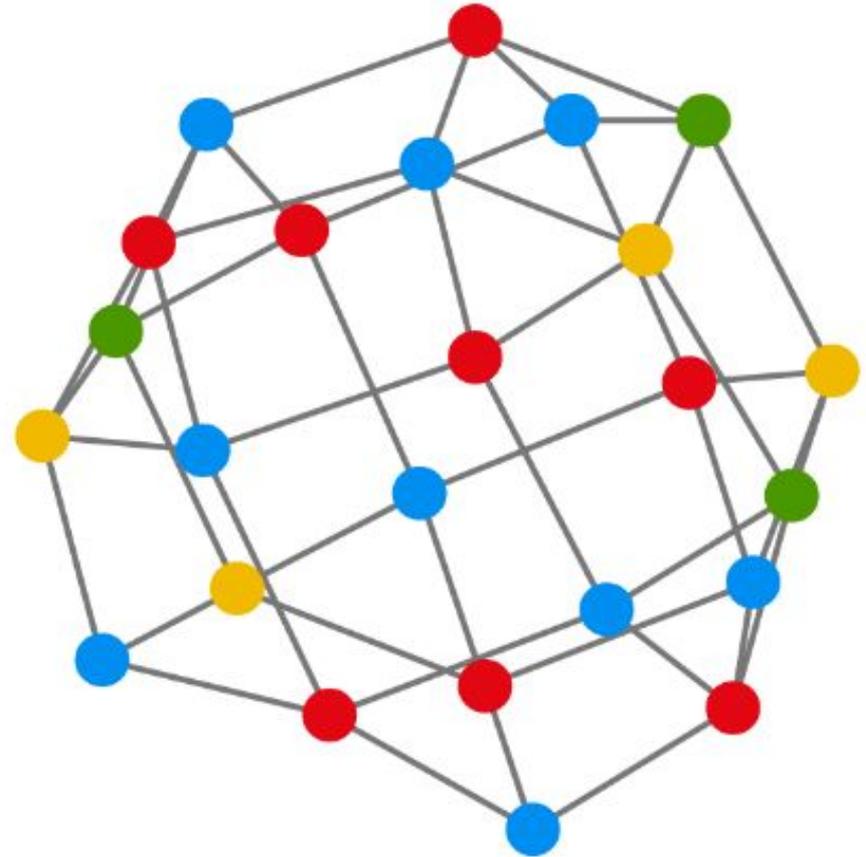


### 3) Sélection de résultats sur les systèmes d'information

Une proposition innovante : l'utilisation des graphes

#### Qu'est ce qu'un graphe ?

- **Des noeuds**  
représentant des objets
- **Des liens**  
représentant des valeurs  
d'attributs communs  
à deux objets  
(= descripteurs)



### 3) Sélection de résultats sur les systèmes d'information

Une proposition innovante : l'utilisation des graphes

#### Qu'est ce qu'un graphe ?

##### - Des noeuds

Objets géographiques (sites)

Transects

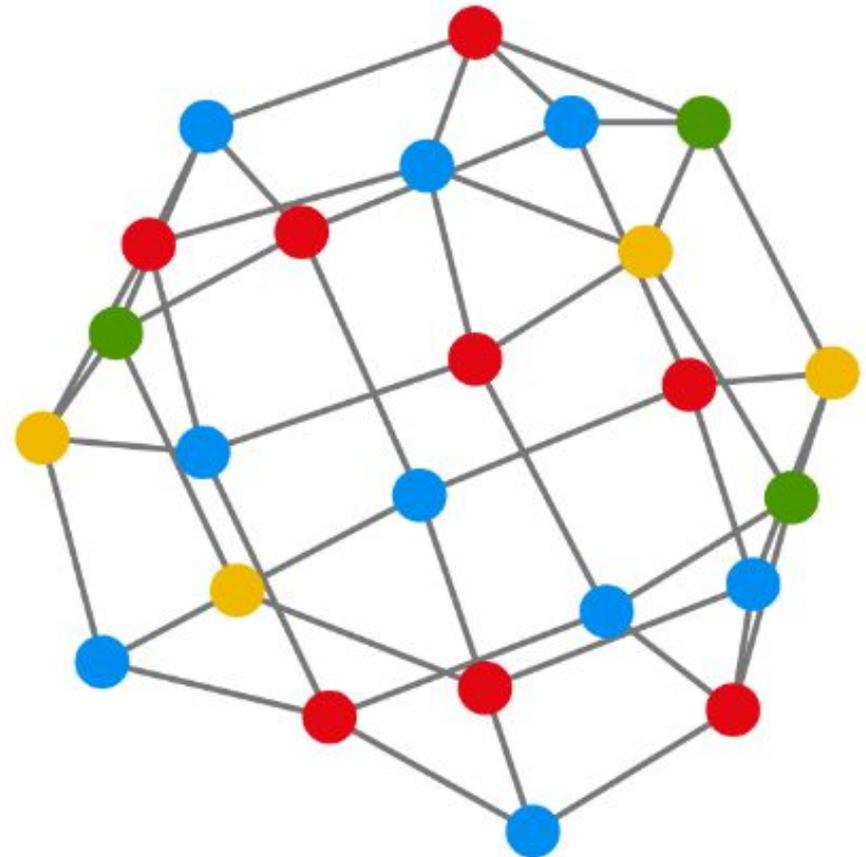
Quadrats photo

Observateurs

Espèces / Taxons

Traits

...



### 3) Sélection de résultats sur les systèmes d'information

Une proposition innovante : l'utilisation des graphes

#### Qu'est ce qu'un graphe ?

- **Des noeuds**
- **Des liens**

Objets géographiques (sites)

Transects

Quadrats photo

Observateurs

Espèces / Taxons

Traits

Définitions

Services

Concepts / idées

Combinaison de traits

Objets combinatoires

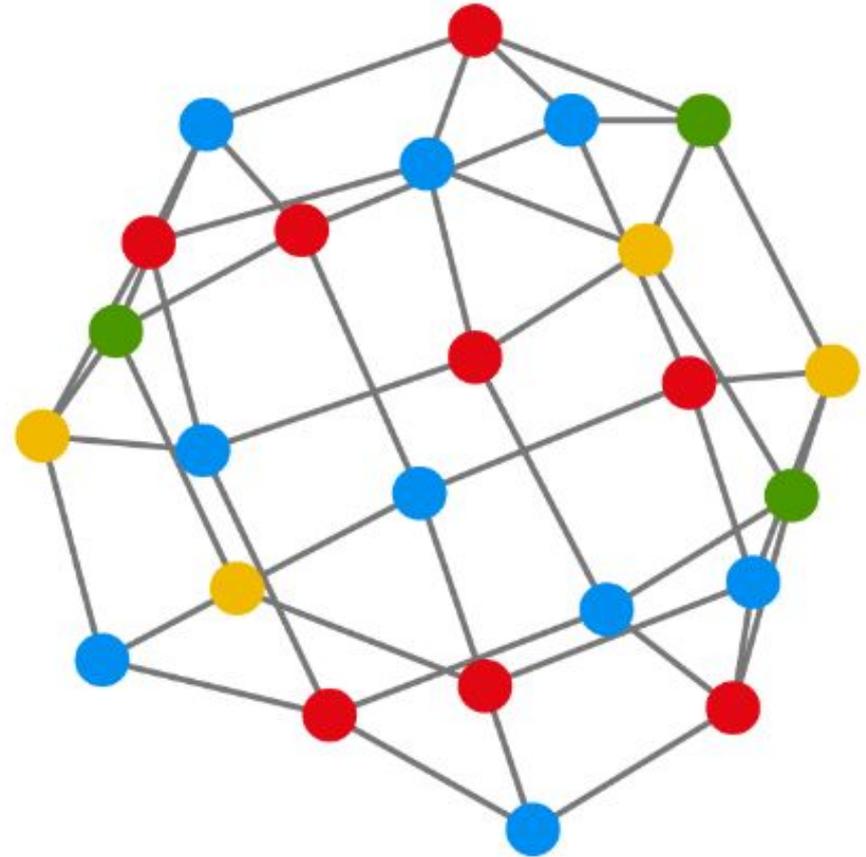
Echantillons

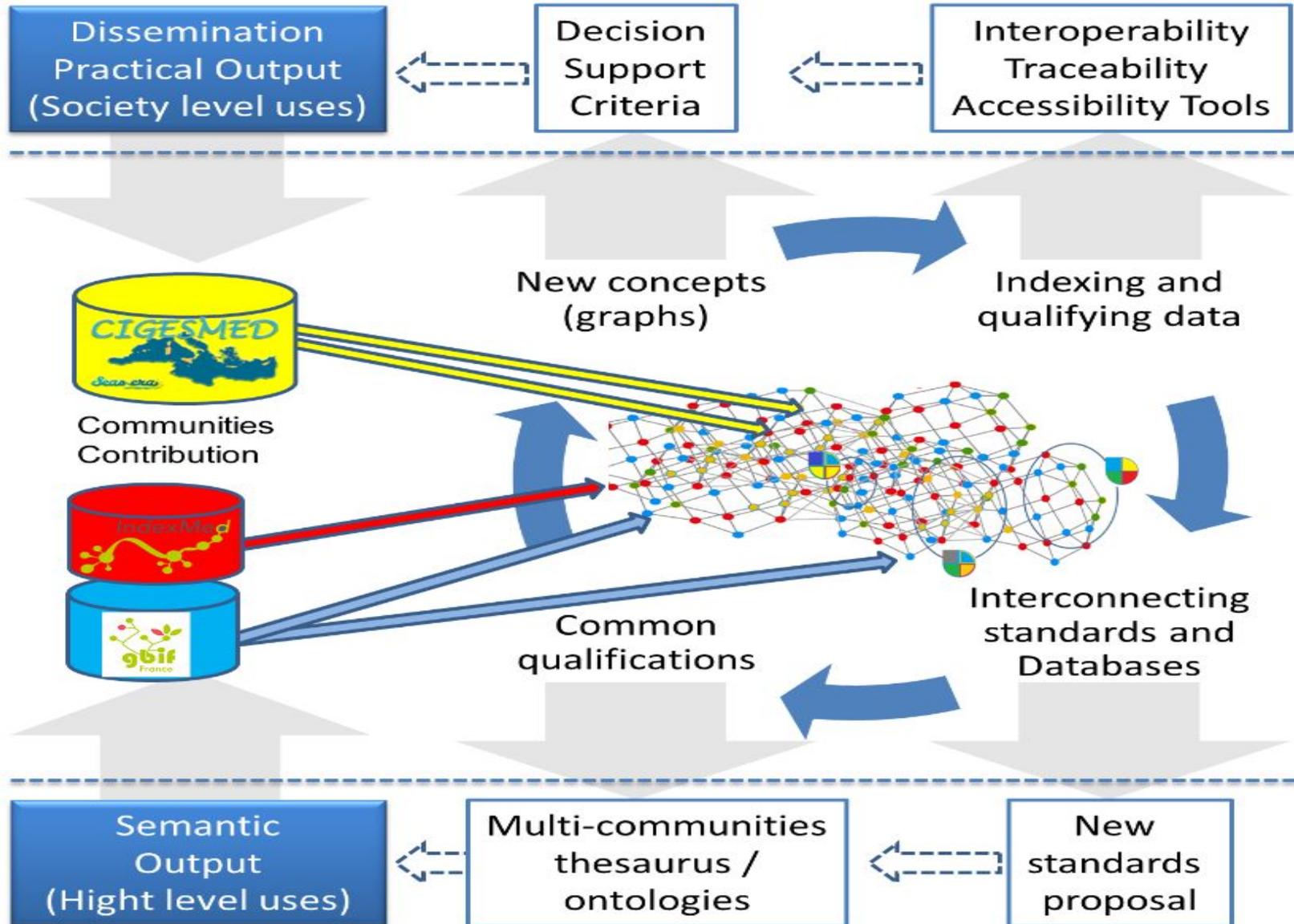
Objets "moléculaires"

Photos

Groupes citoyens

...





### 3) Sélection de résultats sur les systèmes d'information

Le prototype

Online IndexMed interface (test phase) : input and output fonctionnalités

Using Neo4J

Cypher commands

Formular

XML – JSON

Generated

```
Graph json
{
  "nodes": [
    {
      "title": "INDEXCOR_FTF_20130307_D1_SS01_40x6",
      "label": "TraitementPhotos",
      "id": "514",
      "index": 0,
      "weight": 14,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    },
    {
      "title": "Parazoanthus axinellae",
      "label": "EspècesAssignees",
      "id": "418",
      "index": 1,
      "weight": 84,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    },
    {
      "title": "INDEXCOR_FTF_20130307_D1_SS01_40x6",
      "label": "TraitementPhotos",
      "id": "515",
      "index": 2,
      "weight": 14,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    }
  ]
}
```

- 1) choose external databases
- 2) choose nodes
- 3) choose links
- 4) run the graph construction
- 5) move the nodes
- 6) remove links/nodes
- 7) change links/nodes properties
- 8) change links/nodes color/form
- 9) generate the graph flux
- 10) download the result
- 11) generate persistent web URL

Results :

- a) a graphic representation
- b) a web service accessible for data computing

The screenshot shows the IndexMed interface with a graph visualization on the right and a list of nodes on the left. The graph features a central node 'Sp 1' (blue) and another 'Sp 2' (blue), surrounded by numerous 'SS17' nodes (pink). A 'Marseille' node (green) is highlighted with a red arrow. The interface includes a toolbar with icons for search, save, and share, and a text input field for queries. Below the graph, there are filters for 'Lieu', 'EspècesAssignees', and 'TraitementPhotos', along with a table of node properties.

Label : TraitementPhotos obs : SS17 program : CIGESMED photo : INDEXCOR\_FTF\_20130307\_D1\_SS01\_40x60 T01\_TC00\_Q01 site : FTF camera : Nex9 frame : 40x76 date : 20130323 transectype : Linear density : 116 operator : ValidLDRDDG transectnum : T01\_TC10 depth : D17

Graph json

```
{
  "nodes": [
    {
      "title": "INDEXCOR_FTF_20130307_D1_SS01_40x6",
      "label": "TraitementPhotos",
      "id": "514",
      "index": 0,
      "weight": 14,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    },
    {
      "title": "Parazoanthus axinellae",
      "label": "EspècesAssignees",
      "id": "418",
      "index": 1,
      "weight": 84,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    },
    {
      "title": "INDEXCOR_FTF_20130307_D1_SS01_40x6",
      "label": "TraitementPhotos",
      "id": "515",
      "index": 2,
      "weight": 14,
      "x": 655.6954305619001,
      "y": 667.9262476041913,
      "px": 655.6954305619001,
      "py": 667.9262476041913
    }
  ]
}
```

Results :

- a) a graphic representation
- b) a web service accessible for data computing

### 3) Sélection de résultats sur les systèmes d'information

Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?



The screenshot shows the IndexMed web interface. At the top, there is a navigation bar with the 'IndexMed' logo and a menu. Below the logo, there is a toolbar with icons for 'Import JSON', a list, a link, a cloud, and a play button. A blue arrow points to a search filter menu that is open, showing three dropdown menus and a search icon. Below the search filter menu, there is a yellow box with the text: *GENERIQUE!!  
Paramétrage  
des OBJETS  
Par type / par attributs ou  
par valeurs d'attributs*. To the right of the search filter menu, there is a large grey box with the text: *Zone d'affichage  
et de manipulation  
du graphe*.

### 3) Sélection de résultats sur les systèmes d'information

Une proposition innovante (pour notre communauté) : l'utilisation des graphes

The screenshot displays a software interface for graph-based data analysis. On the left, a 'Journal de requêtes' (Query Log) lists several Cypher queries with their respective limits and sorting options. A blue arrow points from the text 'Journal de requêtes' to this log. The main area shows a network graph with nodes of various colors (green, red, blue, yellow) and sizes, connected by edges. Below the graph, there are 'Nodes properties' and 'Attributes properties' sections. The 'Attributes properties' section contains a grid of input fields for various parameters like Frame, Transect, Operator, Obs, Density, Program, Orientation, Source, Camera, Slope, Date, Quadrat, Site, Transect\_Num, Lights, and Depth. A legend at the bottom indicates that node colors correspond to different categories and node sizes correspond to another set of categories.

**Journal de requêtes**

```

MATCH ()-[r:FREQ_ESPECE]->() WHERE toInt(r.target1)>50
RETURN r LIMIT 10000

MATCH ()-[r:FREQ_ESPECE]->() WHERE
toInt(r.target1)>50 RETURN r LIMIT 25

MATCH ()-[r:FREQ_ESPECE]->() RETURN r
LIMIT 25

MATCH ()-[r:FREQ_ESPECE]->() WHERE
toInt(r.target1)>50 RETURN r ORDER BY
toInt(r.target1) LIMIT 25

MATCH ()-[r:FREQ_ESPECE]->() WHERE
toInt(r.target1)>50 RETURN r ORDER BY
toInt(r.target1) LIMIT 50

MATCH ()-[r:FREQ_ESPECE]->() WHERE
toInt(r.target1)>50 RETURN r ORDER BY
toInt(r.target1) LIMIT 300

MATCH (n0:EspeciesAssignees)-
[r0:SE_TROUVE_DANS_CETTE_PHOTO]->
(n1:TraitementPhotos) WHERE
(n0.speciesname='Adeonella calveti')

MATCH (n0:EspeciesAssignees)-
[r0:SE_TROUVE_DANS_CETTE_PHOTO]->
(n1:TraitementPhotos) WHERE
(n0.speciesname='Adeonella calveti')

MATCH (p:Personne)-[r:POSSEDE]->
(o:objet) OPTIONAL MATCH (l:Labo)-
[r1:EMPLOI]->(p1:Personne) OPTIONAL
  
```

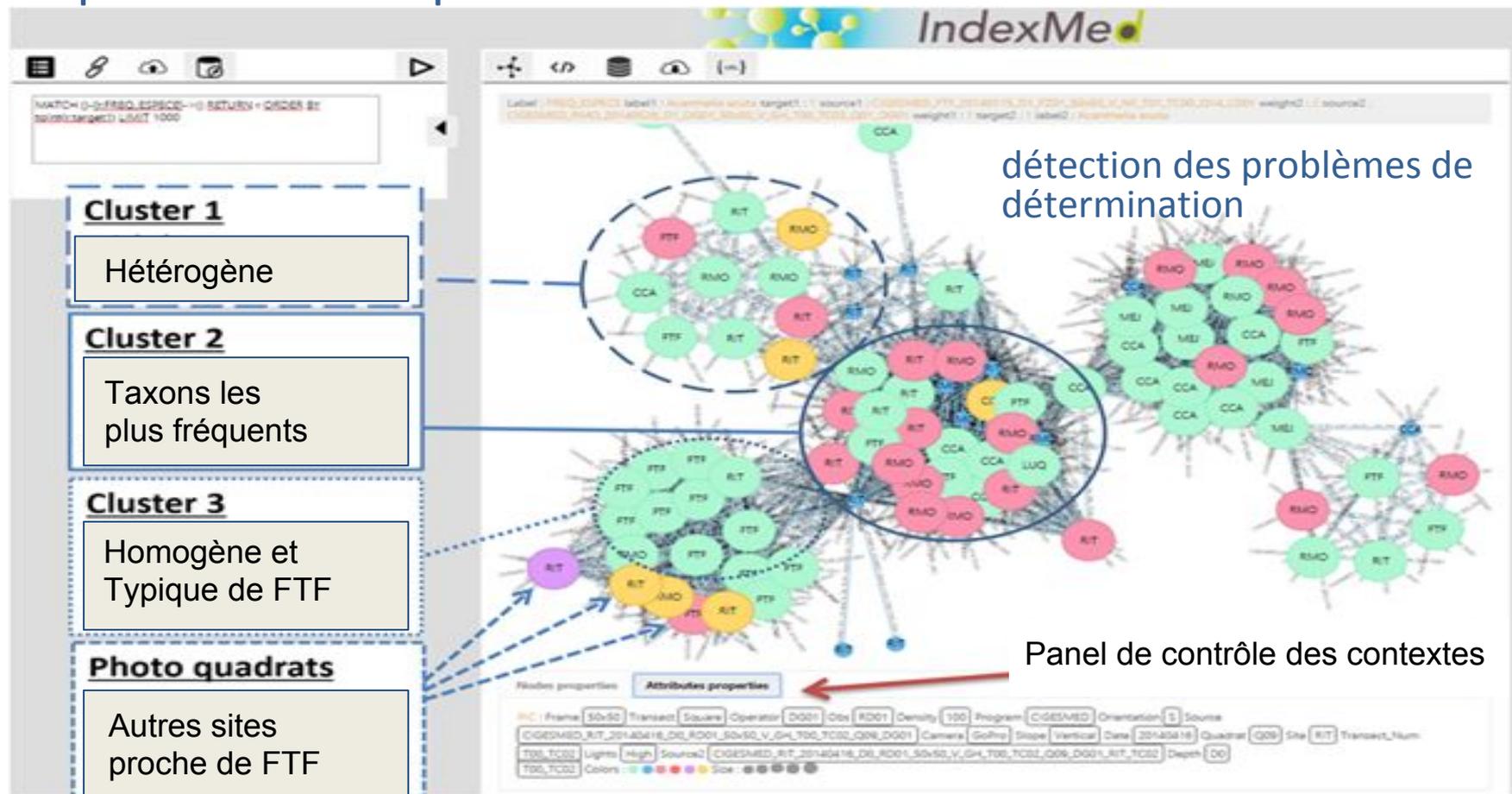
**Nodes properties**    **Attributes properties**

PIC : Frame 50x50 Transect Square Operator DG01 Obs DG01 Density 100 Program CIGESMED Orientation S Source CIGESMED\_RMO\_20140526\_D1\_DG01\_50x50\_V\_GH\_T00\_TC02\_Q02\_DG01 Camera GoPro Slope Vertical Date 20140526 Quadrat Q02 Site RMO Transect\_Num T00\_TC02 Lights High Source2 CIGESMED\_RMO\_20140526\_D1\_DG01\_50x50\_V\_GH\_T00\_TC02\_Q02\_DG01\_RMO\_TC02 Depth D1

T00\_TC12 Colors: ●●●●●● Size: ●●●●●●

### 3) Discussion : la fouille de données

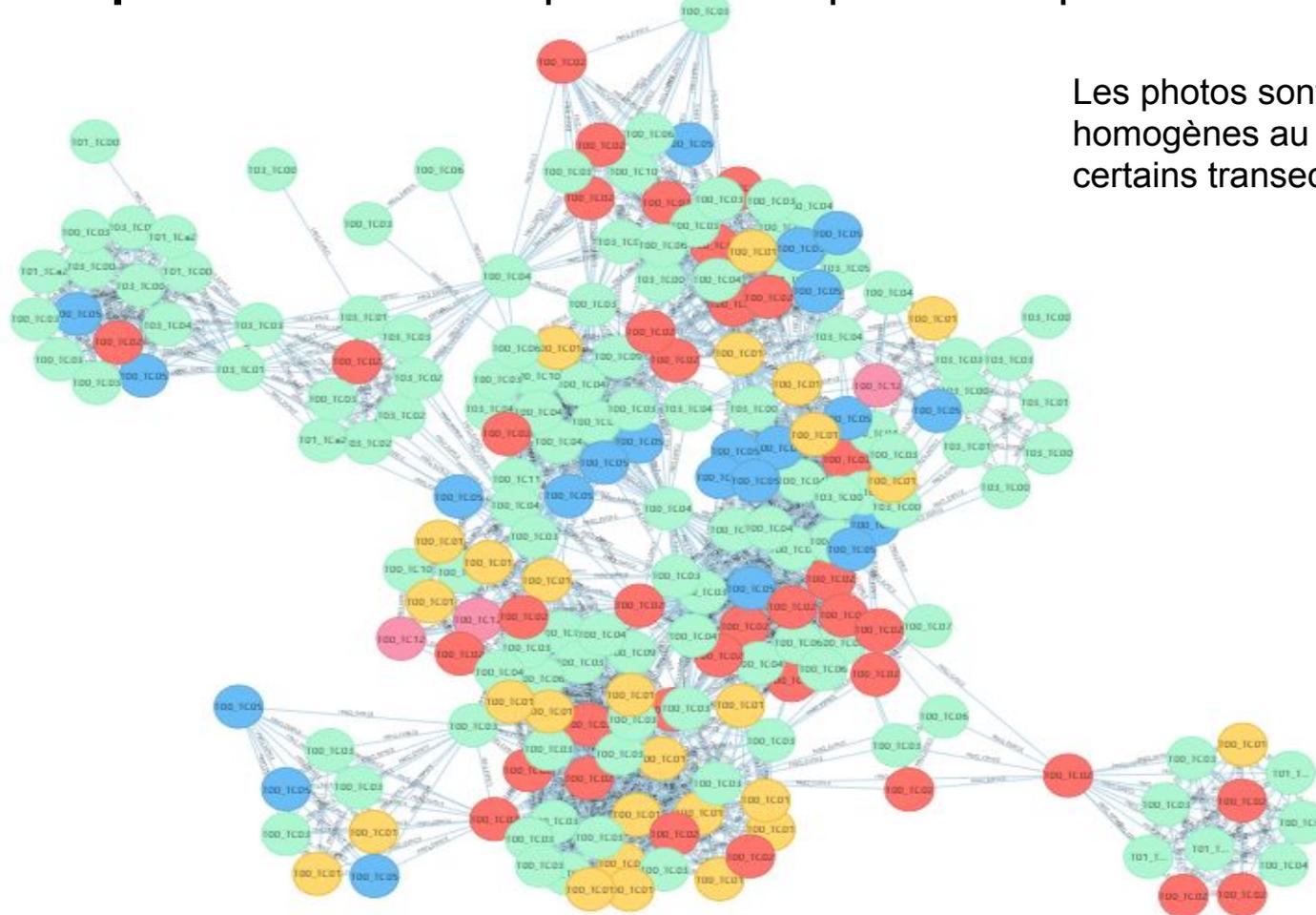
La fouille de données basée sur les graphes, utilisables pour étudier les  
Compétences des opérateurs :



### 3) Discussion : la fouille de données : les graphes valués (CIGESMED)

**les noeuds** : des photos de différents transects (une sélection)

**les liens pondérés** : les fréquences d'espèces les plus abondantes

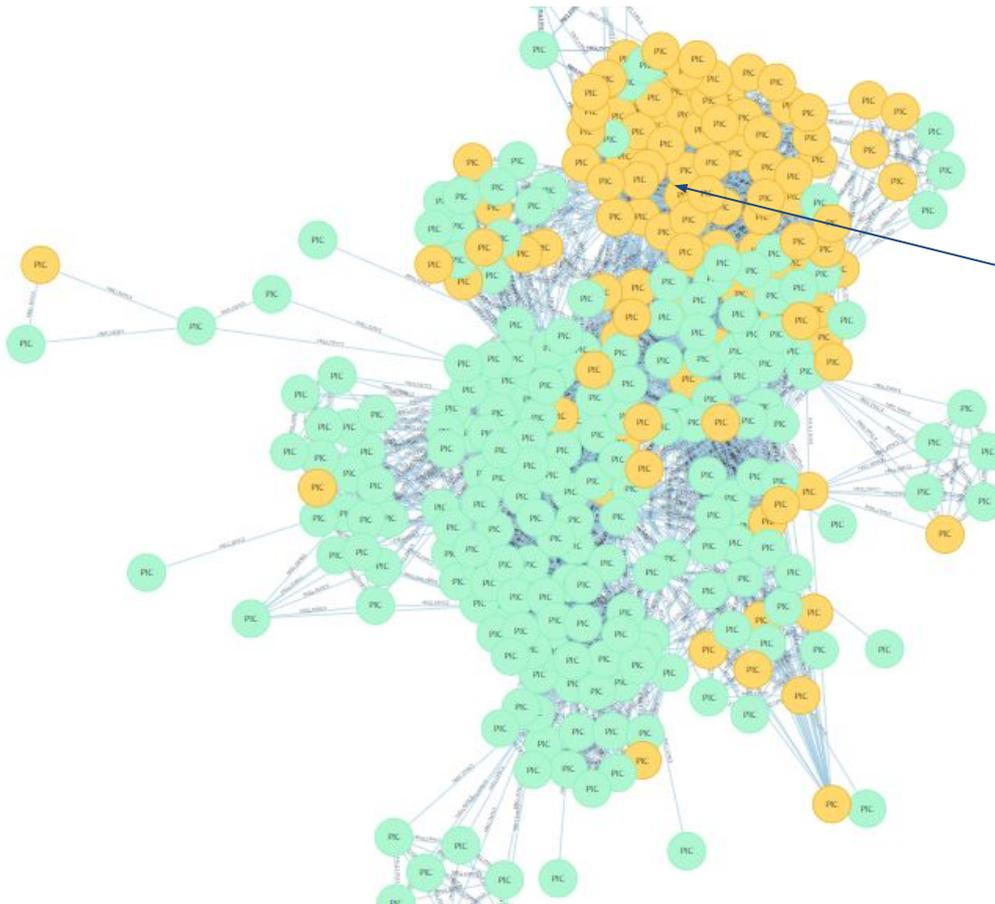


Les photos sont elles  
homogènes au sein de  
certains transects ?

### 3) Discussion : la fouille de données

**les noeuds** : des photos

**les liens pondérés** : les fréquences d'espèces les plus abondantes



L'inclinaison est elle un facteur explicatif de la structure des assemblages d'espèces ?

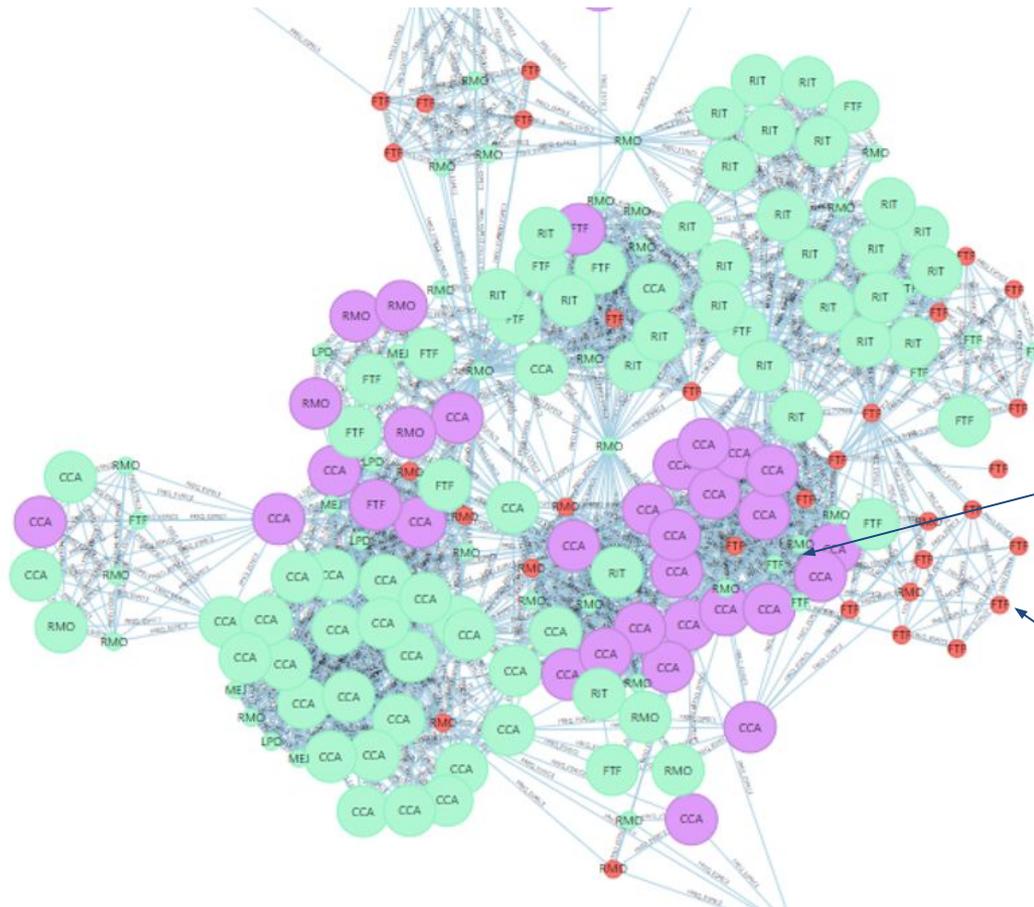
Pente Intermédiaire



### 3) Discussion : la fouille de données

**les noeuds** : des photos

**les liens pondérés** : les fréquences d'espèces les plus abondantes



La combinaison orientation-profondeur est elle un facteur possiblement explicatif de la structure des assemblages d'espèces ?

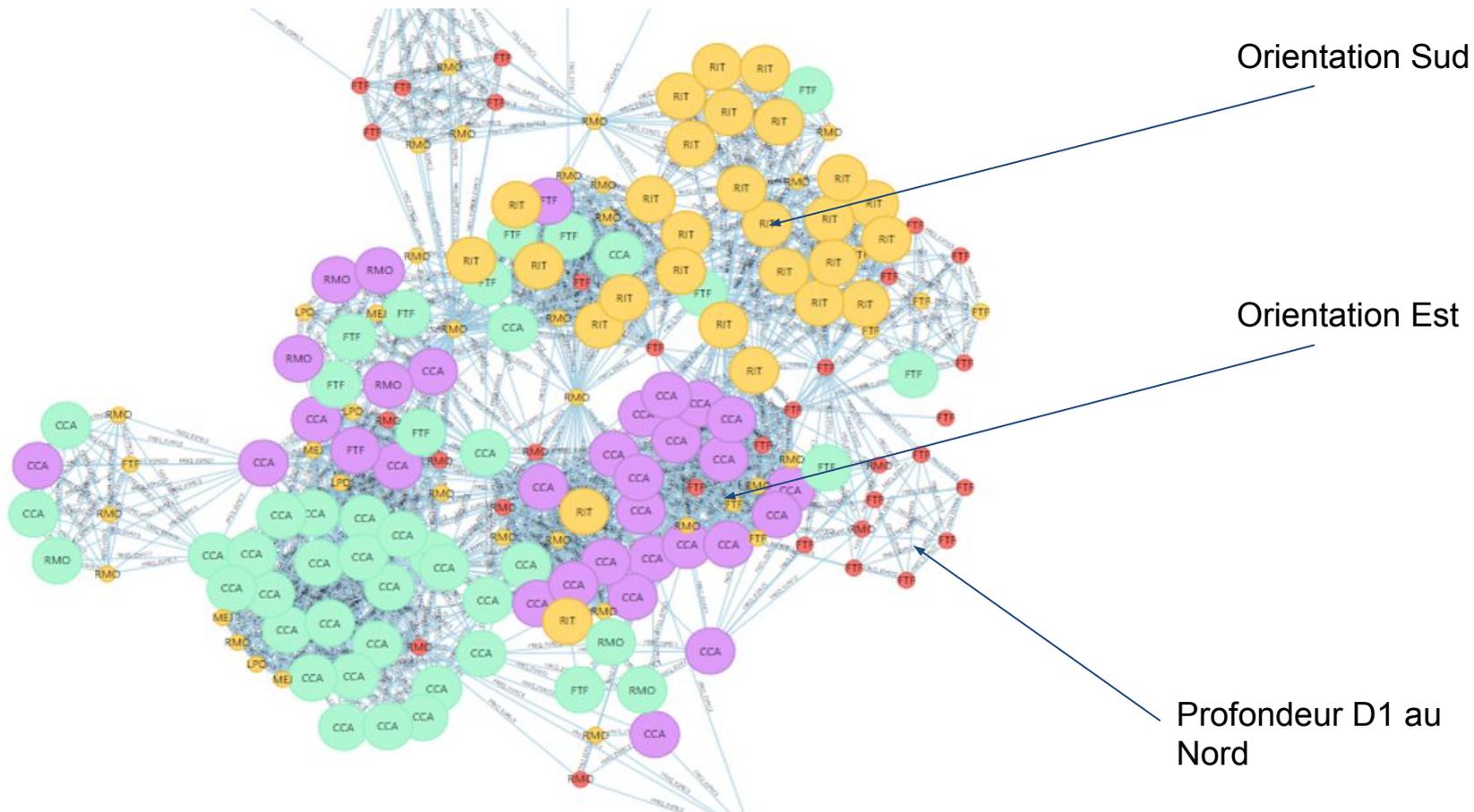
Orientation Est

Profondeur D1 au Nord

### 3) Discussion : la fouille de données

**les noeuds** : des photos

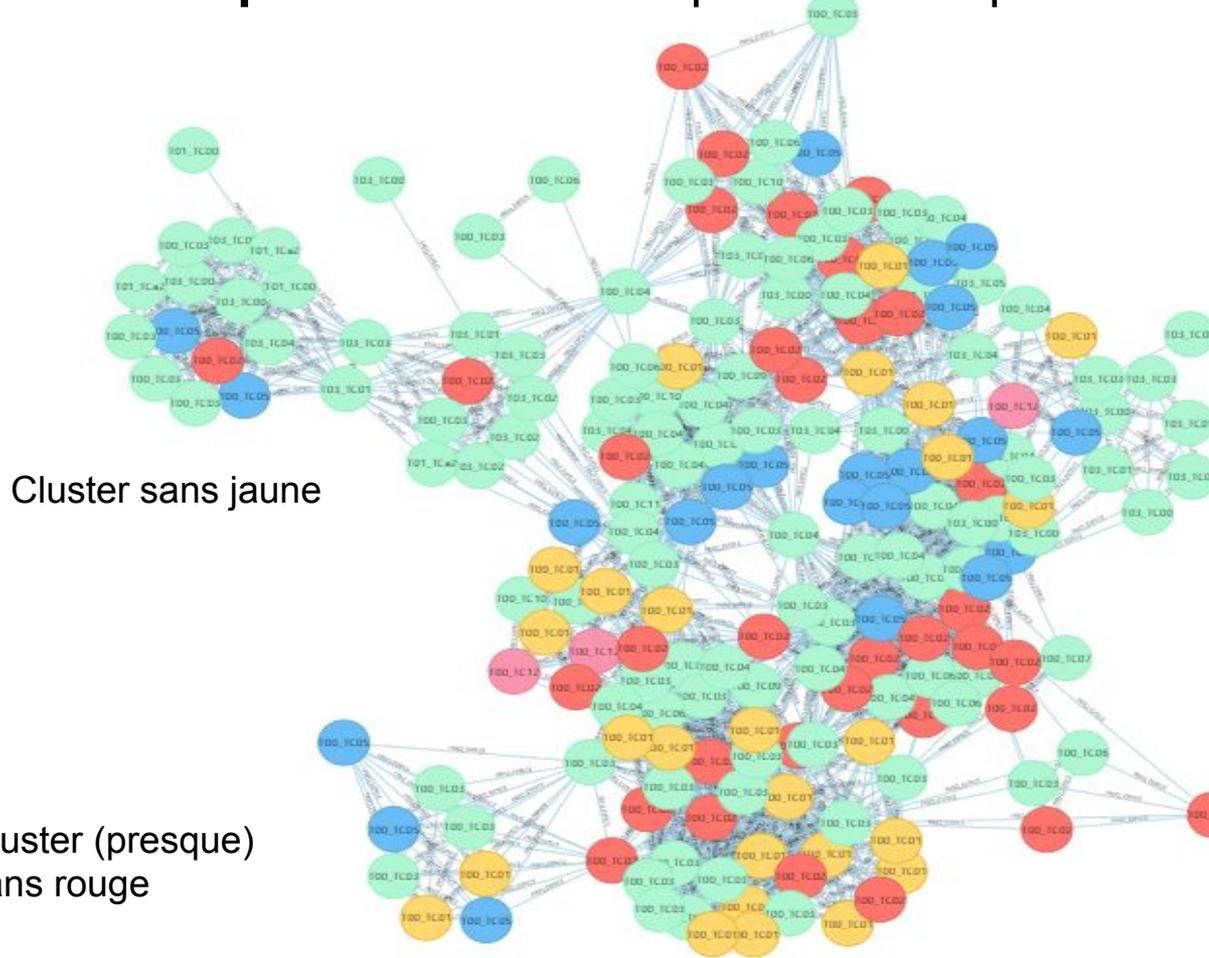
**les liens pondérés** : les fréquences d'espèces les plus abondantes



### 3) Discussion : la fouille de données

**les noeuds** : des photos de différents transects (une sélection)

**les liens pondérés** : les fréquences d'espèces les plus abondantes



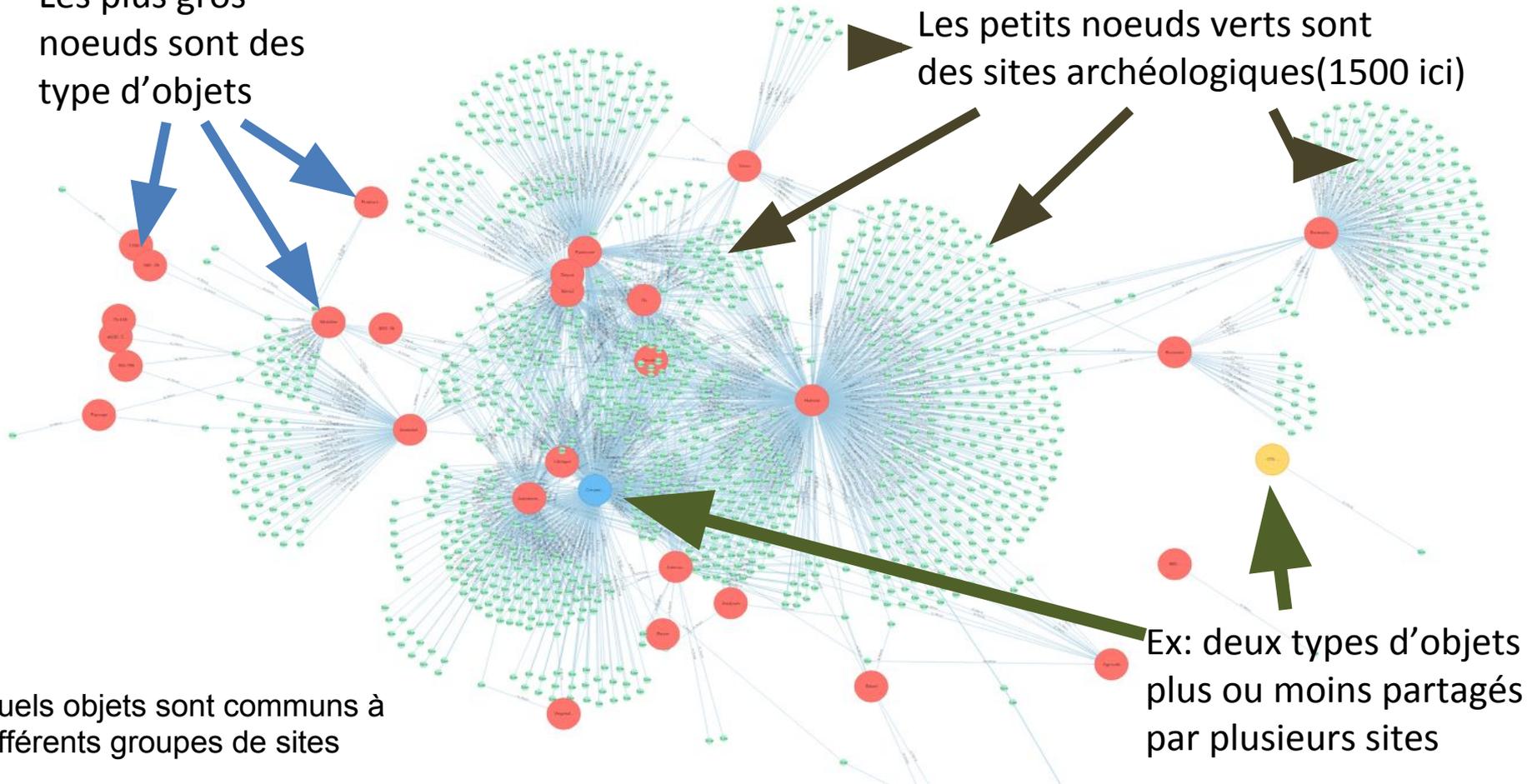
Peut on visualiser l'absence de certains transects dans certains clusters représentant une structure particulière des assemblages d'espèces ?

### 3) Discussion : la fouille de données

La fouille de données basée sur les graphes, utilisables pour étudier les

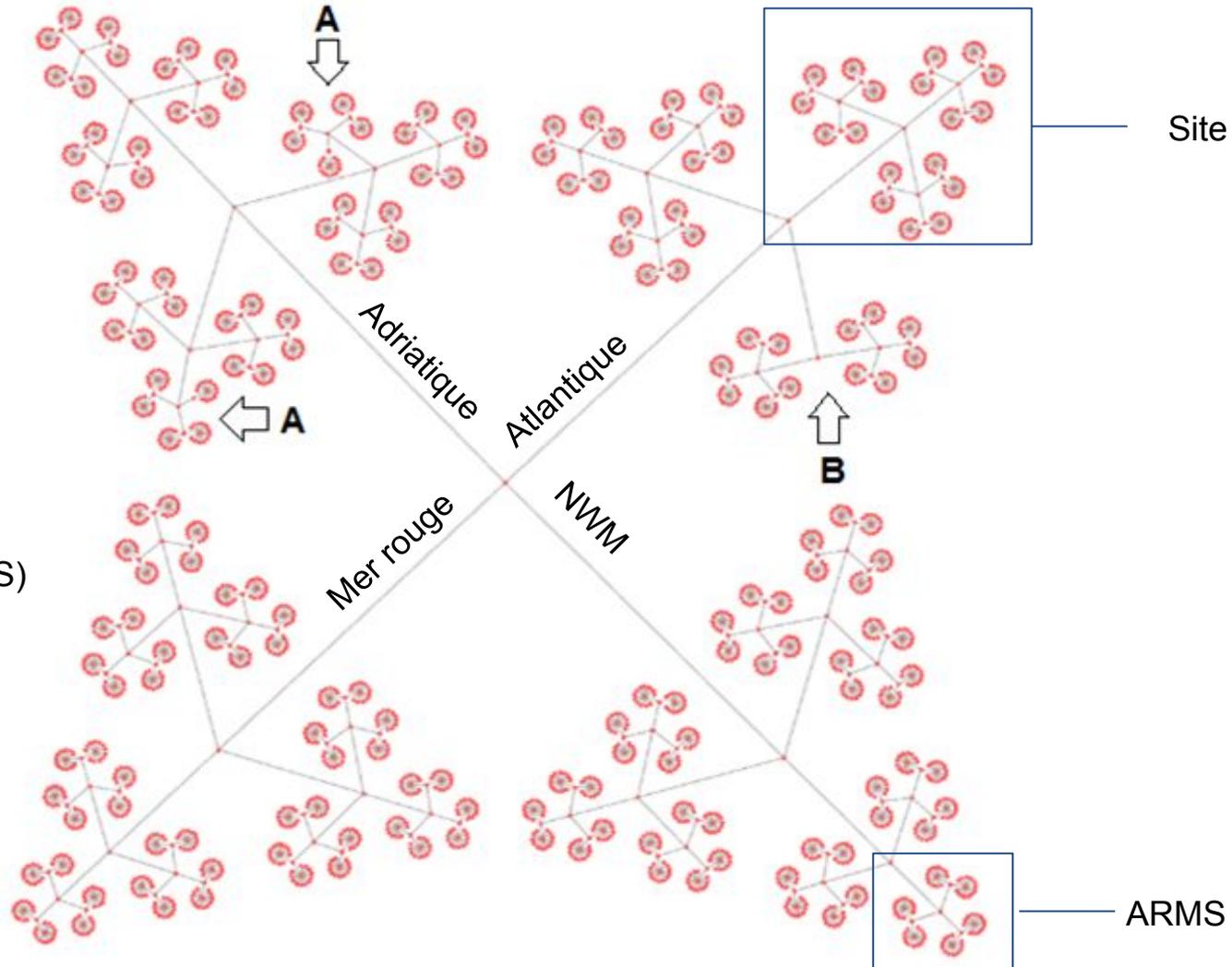
Les plus gros  
noeuds sont des  
type d'objets

Les petits noeuds verts sont  
des sites archéologiques(1500 ici)



systèmes d'observation, les systèmes observés et les **systèmes d'information**

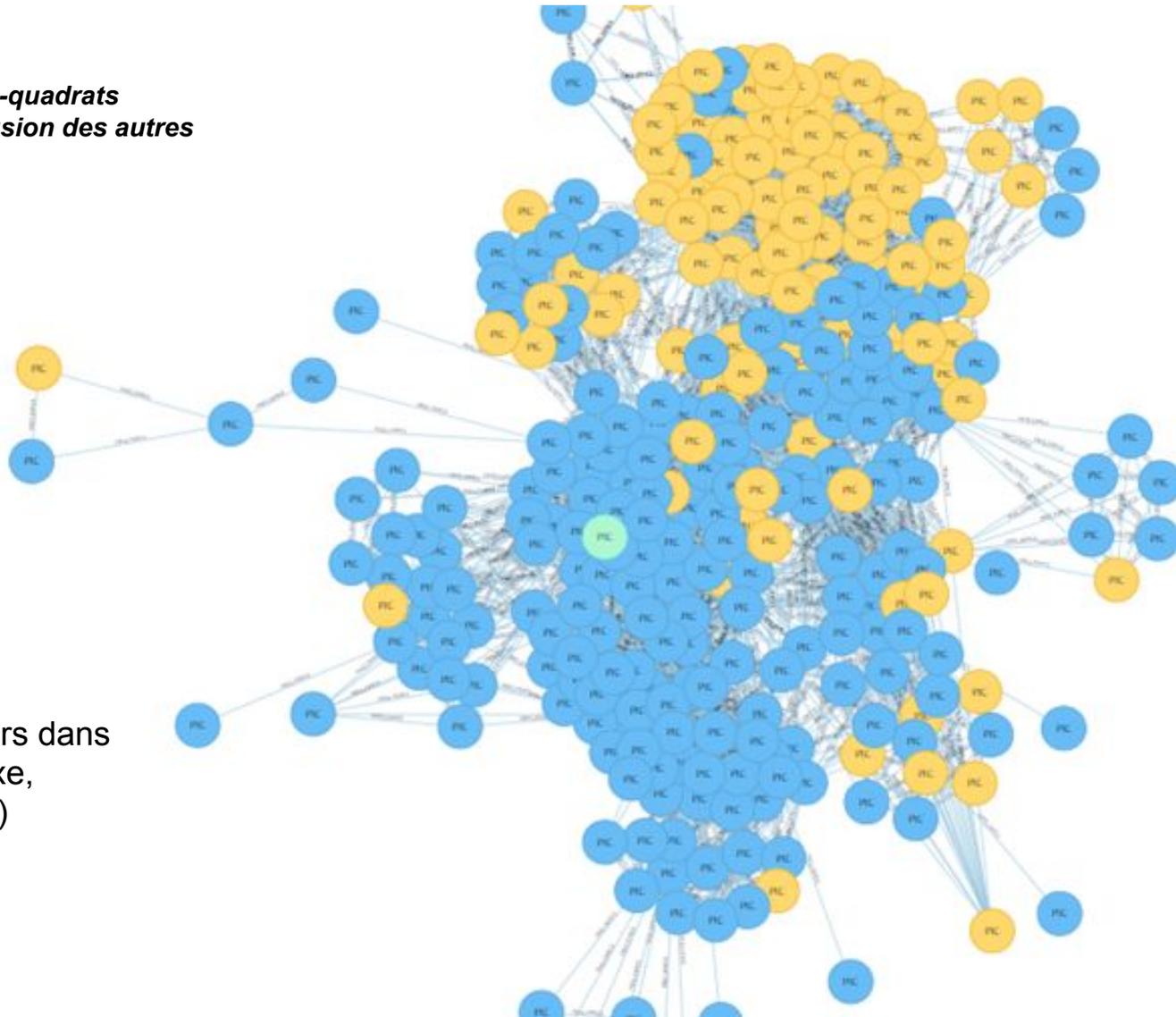
Visualisation du plan  
d'échantillonnage  
des faces de plaques  
d'ARMS  
(programme DEVOTES)



systèmes d'observation, les systèmes observés et les **systèmes d'information**

*sélection des photos-quadrats  
CIGESMED par exclusion des autres  
catégories*

*En orange : incliné  
En bleu : vertical*



Détection des erreurs dans  
les données (syntaxe,  
donnée manquante)

systèmes d'observation, les systèmes observés et les **systèmes d'information**

## 4) Quelques messages à retenir

### Pour rendre opérationnelle l'observation à large échelle :

- **Explorer les standards existants** avant de proposer un protocole
- Prévoir **l'adaptabilité** d'une mesure d'un facteur **pour plusieurs usages**
- **Éviter la surqualité**, mesurer les temps de mise en oeuvre
- **Préférer la répliquabilité** à la précision

## 4) Quelques messages à retenir

### L'exploration de données hétérogènes basée sur les graphes : un effet vertueux pour l'amélioration des standards et des SI

- Les graphes permettent de modéliser et représenter les **systèmes d'observation, les systèmes observés, et les systèmes d'information**
- Les graphes permettent d'explorer à large échelle l'**effet de combinaison de contextes** sur la topologie du graphe
- Les graphes permettent de **représenter un ensemble de données quelque soit sa complexité**, sans projection (et donc sans déformation)
- L'utilisation des graphes est prometteuse pour améliorer les systèmes d'aide à la décision, et permet d'utiliser des **systèmes d'information répartis** (utilisation modulaire et temporalisation)

## 4) Quelles perspectives

### Les futurs développements

#### **Pour rendre opérationnelle l'observation à large échelle :**

Améliorer les standards concernant les facteurs de contextes

Renforcer les outils de formation et d'intercalibration des observateurs

## 4) Quelles perspectives

### Les futurs développements

#### Pour rendre opérationnelle l'observation à large échelle :

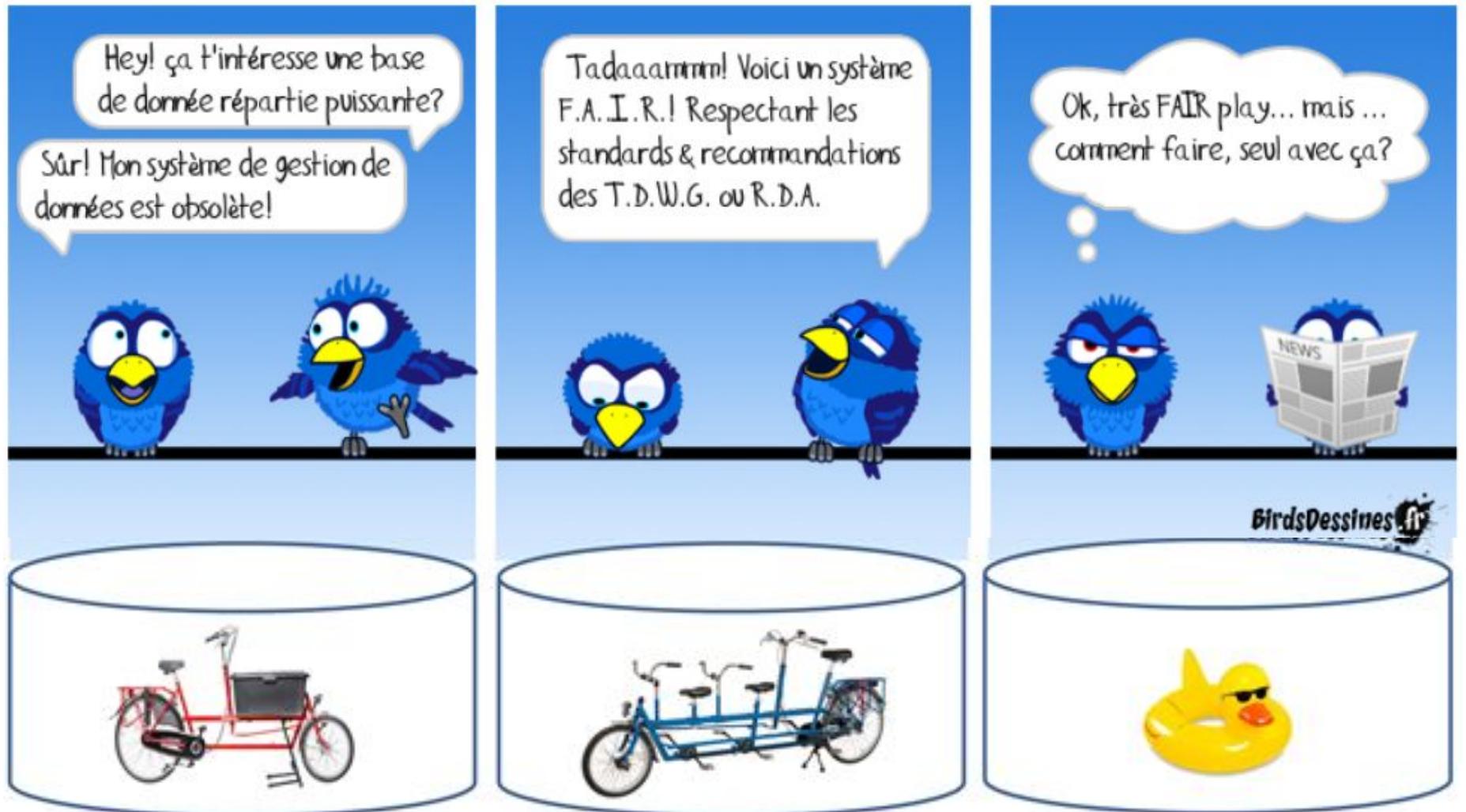
Améliorer les standards concernant les facteurs de contextes

Renforcer les outils de formation et d'intercalibration des observateurs

#### Le futur de l'exploration de données hétérogènes basée sur les graphes :

- Construire des arbres de décision pour choisir “**la bonne statistique**” à appliquer aux graphes (en fonction des données, de leur structure ET de leur qualité)
- Utiliser la **parallélisation des calculs** pour fouiller de très grands graphes ( passage à l'échelle avec des milliards de noeuds et centaines de milliards de liens)
- Créer une **bibliothèque de motifs de (valeur de) contextes** ayant une incidence sur la structure des populations
- Utiliser le **machine learning pour “compléter”** les données grâce aux contextes
- **Alimenter les systèmes d'aide à la décision** avec ces motifs de contextes
- Elaborer une stratégie pour **construire des ontologies à partir de graphes valués** (ANF en novembre 2018)

# RDF or PERISH ?



Cycle pictures from : [http://cargocycling.org/category/riding\\_type/family-cycling](http://cargocycling.org/category/riding_type/family-cycling)

FAIR or PERISH ?

## J3 : Convergence apprentissage et sémantique

Atelier de programmation GRAMINEES (GRAphe data Mining In Natural, Ecological and Environnemental Sciences)

Romain David, IMBE, INEE, Stéphane Pérennes, I3S, INS2i)

# GRAphe data Mining In Natural, Ecological and Environnemental Sciences



## MaDICS

Masses de Données, Informations et Connaissances en Sciences

### Big Data - Data Science

[Accueil](#) [Actions](#) [Manifestations](#) [Réseaux](#) [Actualités](#) [Offres d'emploi](#) [Calendrier](#) [Diffuser](#) [Intranet](#)

**Accueil** [Actions](#) [Actions en cours](#) [GRAMINEES](#)

[Calendrier](#) [Documents](#) [Envoyer un email](#) [Liste des intéressés](#)

## Action GRAMINEES

*GRAphe data Mining In Natural, Ecological and Environnemental Sciences*

### Responsables

- Romain David, IMBE, INEE
- Dino Ienco, TETIS, IRSTEA

### Thématiques

fouille de données  
écologie et biodiversité

### Données concernées

fréquences d'espèces géo spatialisées et contextualisées

### Résumé

Récemment, les 3èmes journées organisées par le consortium IndexMeed ont mis en évidence le potentiel des approches basées sur les graphes, les lacunes en termes de compétences et d'expérience de la communauté des écologues pour adapter et utiliser ces méthodes pour analyser leurs jeux de données. Les 2 objectifs de l'action proposée sont: 1/ La mise en place d'une dynamique d'échange entre des experts en écologie/biodiversité et des experts du domaine des STIC. 2/ La réalisation d'une première carte des compétences de laboratoires qui pourront apporter des outils méthodologiques ou des techniques algorithmiques adéquates pour l'analyse des données sur la biodiversité.

### Objectifs

Les deux objectifs principaux de l'action proposée sont:

Pour participer aux Actions et manifestations MaDICS, vous devez [devenir membre](#).  
*In order to participate in MaDICS Actions and Events, you have to [register](#).*

### Accès Rapide

[Nous contacter](#)

### Bienvenue Romain David !

[Se déconnecter](#)

Search ...

### Suivez-Nous Sur Twitter

Tweets by @GDR\_MaDICS

GdR MaDICS Retweeted

IFB\_Bioinformatique  
@IFB\_Bioinfo

Venez nombreux #ReproHackathon ifb-elixirfr.github.io/ReproHackathon... @BioinfoFr @u\_psud @GDR\_MaDICS



## MOYENS:

L'indexation, l'identification  
pérenne des données  
avec le consortium IndexMeed

Consortium IndexMeed : cadre de l'action  
GRAMINEES

### Objectif principal d'IndexMeed:

développer la culture des bases de données et leur accessibilité

Mais surtout de nouvelles méthodes d'utilisation efficace des données  
dans le milieu de la recherche en écologie et biodiversité.

### Conditions :

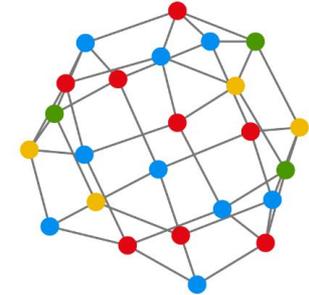
Des bases de données ouvertes, interopérables, stables (et donc recensées) et des données traçables et bien qualifiées,

en adaptant les approches issues d'autres disciplines

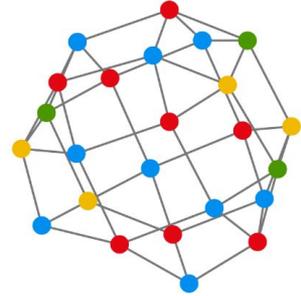
**Astronomie, Physique, Mathématiques, Economie,  
Informatique, Fouille de données...**

-> INTEROPERABLES et MULTIDISCIPLINAIRES

## Objectifs de GRAMINEES



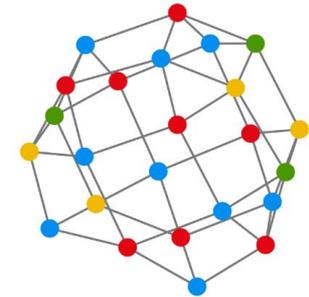
1/ Développer **la dynamique d'échange** entre des experts en écologie/biodiversité et des experts du domaine des STIC concernant la **fouille de graphes** ; (GRAMINEES)



## Objectifs de GRAMINEES

1/ Développer la **dynamique d'échange** entre des experts en écologie/biodiversité et des experts du domaine des STIC concernant la **fouille de graphes** ; (GRAMINEES)

2/ Réaliser une première version de **carte des compétences de laboratoires** qui pourront apporter des outils méthodologiques ou des techniques algorithmiques adéquates pour l'analyse des données sur la biodiversité et l'environnement. (GRAMINEES)



## Moments marquants

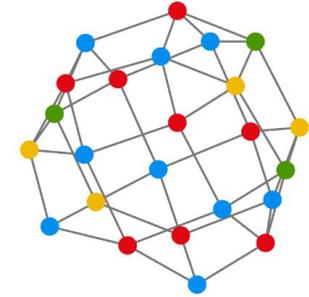
La préparation puis l'animation des événements GRAMINEES a été assurée par des personnes issues des STIC (\*) et des personnes issues des sciences de l'environnement (!).

**15 mai, 29 mai, 26 juin et 15 novembre : Ateliers Curation de données**

Organisation et animation d'ateliers délocalisés autour du nettoyage des jeux de données pour la visualisation des données dans des logiciels de graphes (étape 1 de l'action GRAMINEES). Lors de ces rencontres, il a été décidé de produire un « how to do » réutilisable par d'autres cas d'études.

Ce document a été validé lors des ateliers « visualisation ».

Lors de ces rencontres, la matinée était allouée à des présentations théoriques, et l'après-midi à un travail conjoint de personnes compétentes sur des cas d'études et dans les domaines STICS concernés. (animation Romain DAVID (!), Anna COHEN NABEIRO et Yvan LEBRAS (\*\*))



# Résultats des ateliers sur la curation

Attente d'autres séances de coaching de la part des participants



## La curation de données environnementales étape par étape

*en vue d'une analyse reposant sur la théorie des graphes*



Ce document est issu des échanges ayant eu lieu au cours des ateliers "Curation de données pour la visualisation et l'analyse de données s'appuyant sur les méthodes de graphes", organisés par le consortium [IndexMEED](#) en 2017.

Pour citer ce document : R. DAVID, A. COHEN NABEIRO, A. DELAUD, M. LEYDET-BARBIER, S. PAMERLON, A. QUESNEL-BARBET, L. BERNARD

ici tous les participants aux ateliers "curation"

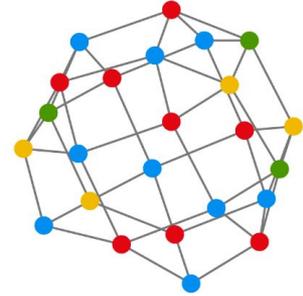
Pour plus d'informations, vous pouvez :

- consulter le site internet <https://indexmeed2017.sciencesconf.org/>
- envoyer un mél à [romain.david@imbe.fr](mailto:romain.david@imbe.fr) et à [ecoscope@fondationbiodiversite.fr](mailto:ecoscope@fondationbiodiversite.fr)

**Objectifs du document :** décrire, étape par étape, les actions préconisées sur un jeu de données pour préparer puis importer des données dans un logiciel de graphes, en mettant en avant les points de difficultés, les erreurs à ne pas commettre, et les bonnes pratiques.

Sommaire

- 1) Enjeux et principes de la curation de données
- 2) Avant : la mise en place du protocole de curation de données
- 3) Pendant : l'application du protocole de curation de données en vue d'une représentation sous forme de graphes
- 4) Après : le contrôle de la qualité de la curation de données



## Moments marquants

La préparation puis l'animation des événements GRAMINEES a été assurée par des personnes issues des STIC (\*) et des personnes issues des sciences de l'environnement (!).

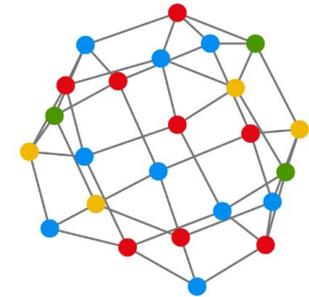
### **22-23 juin : Journées « sciences des données » du GDR MaDICS**

Animation d'une journée de travail sur l'action GRAMINEES et recueil des attentes en compétences et en disponibilité en compétences (animation Dino IENCO (\*), Luc HOGIE (\*), Anna COHEN NABEIRO (!) et Romain DAVID (!))

<http://www.madics.fr/event/journees-madics-2017-23-juin/>

### **04-07 juillet : JDEV Journées nationales du DEVeloppement logiciel, Marseille**

Animation du groupe de travail « Outils, compétences et savoir-faire nécessaires à la fouille de graphes : quelques cas d'utilisation et les défis qui les accompagnent » puis échanges avec les participants sur les définitions des compétences liés à la visualisation et la fouille de graphes en écologie et environnement. (animation Luc HOGIE (\*), Anna COHEN NABEIRO (!) et Romain DAVID (!))



## Moments marquants

La préparation puis l'animation des événements GRAMINEES a été assurée par des personnes issues des STIC (\*) et des personnes issues des sciences de l'environnement (!).

**15-17 novembre : Ateliers Visualisation des données** (Animation Romain BOURQUI (\*), David AUBER (\*), Anna COHEN NABEIRO (!) et Romain DAVID (!).

### **Décembre :**

finalisation des cartes de compétences et graphes sur cas d'étude

production et publications des « how to do »

Ces documents reprenant étape par étape s'appuient tous les éléments des différents ateliers pour les deux périmètres de compétences STIC « Curation de données pour la construction de graphes » et « Visualisation de données sous forme de graphes »

## J3 : Convergence apprentissage et sémantique

Atelier de programmation GRAMINEES (GRAphe data Mining In Natural, Ecological and Environnemental Sciences)

Romain David, IMBE, INEE, Stéphane Pérennes, I3S, INS2i)

**objectif** : préparer les étapes et la répartition des rôles nécessaires à l'élaboration d'un programme de recherche

**focus** : la fouille de données environnementale hétérogènes et réparties basée sur la visualisation et la fouille de graphe.

## **J3 : Convergence apprentissage et sémantique**

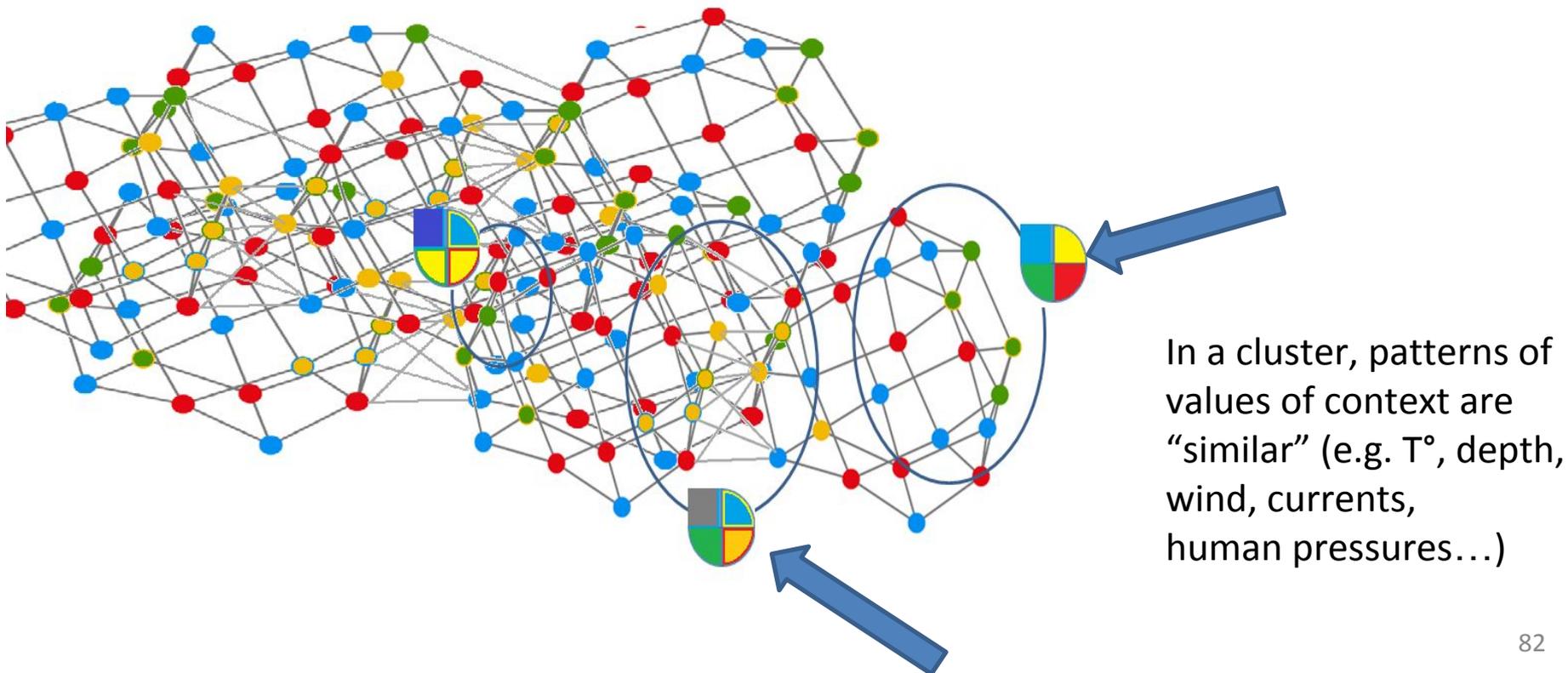
Atelier de programmation GRAMINEES (GRAphe data Mining In Natural, Ecological and Environnemental Sciences)

Romain David, IMBE, INEE, Stéphane Pérennes, I3S, INS2i)

## la fouille de données

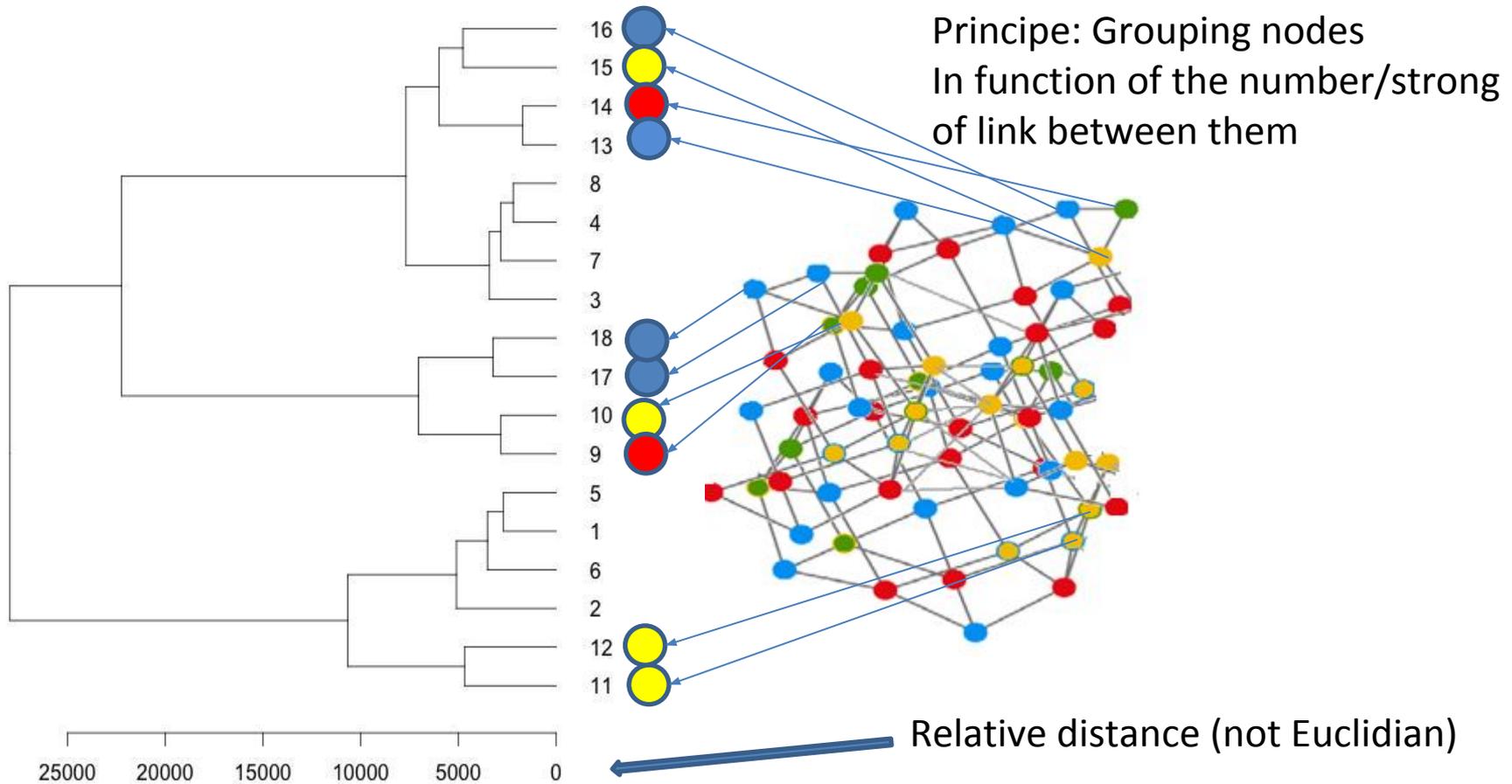
La fouille de données basée sur les graphes, utilisables pour étudier les systèmes d'observation, les systèmes observés et les systèmes d'information

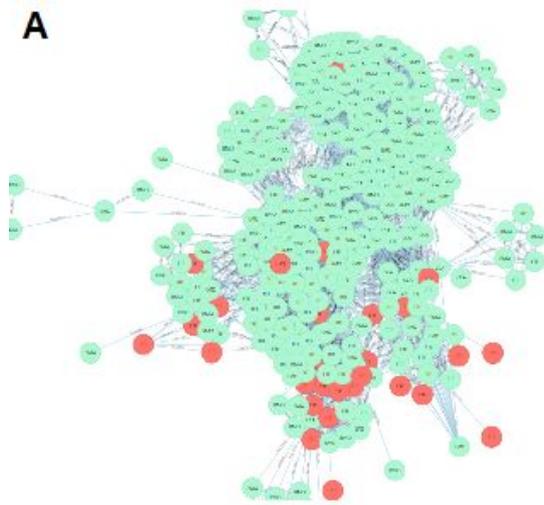
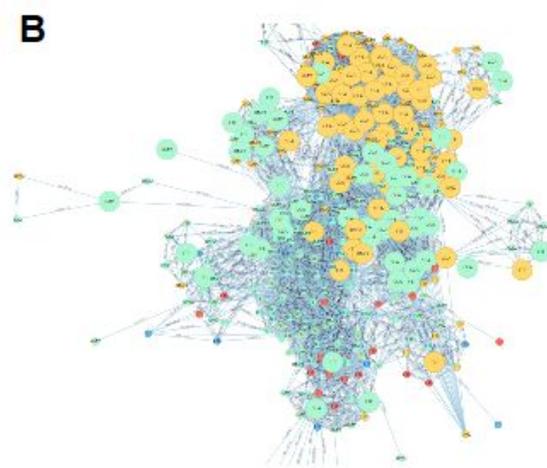
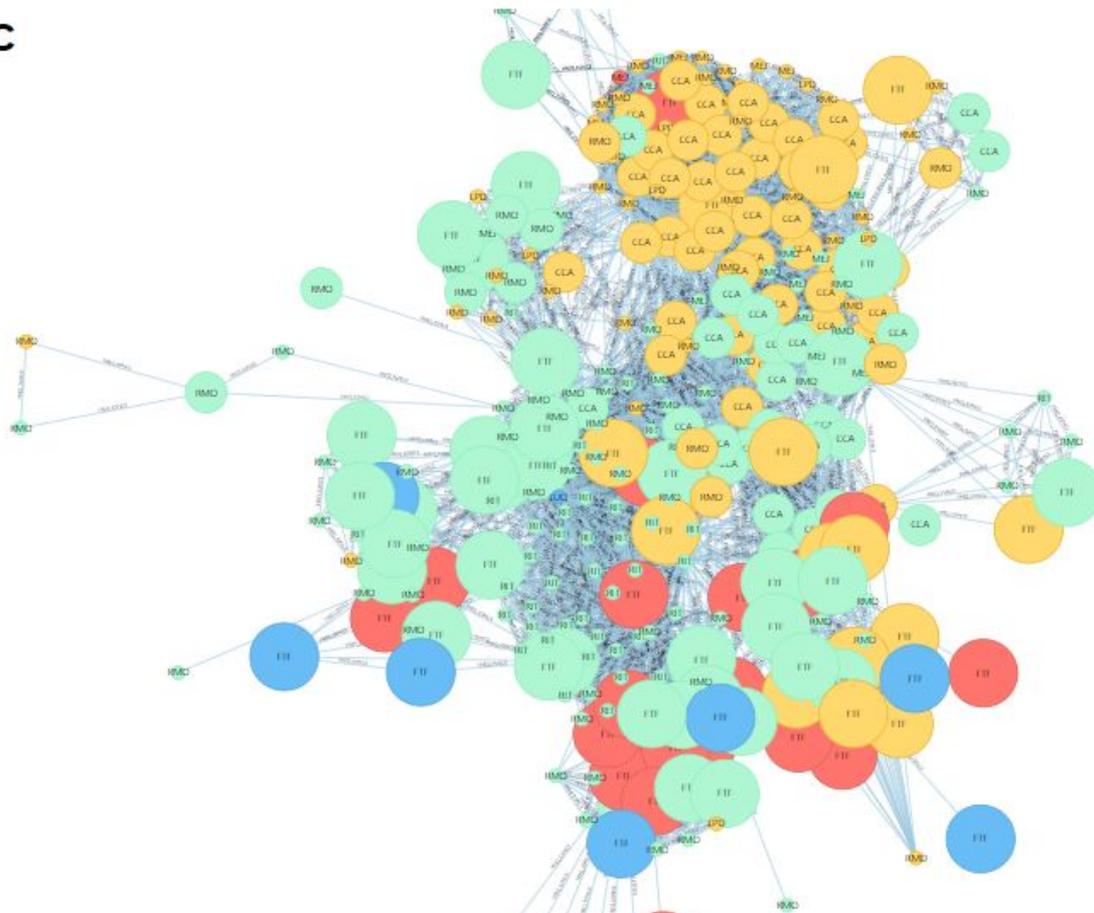
### Patterns recognition



## Sélection de concepts sur les systèmes d'information

Comment faire parler ces données hétérogènes et forcément imparfaites ?

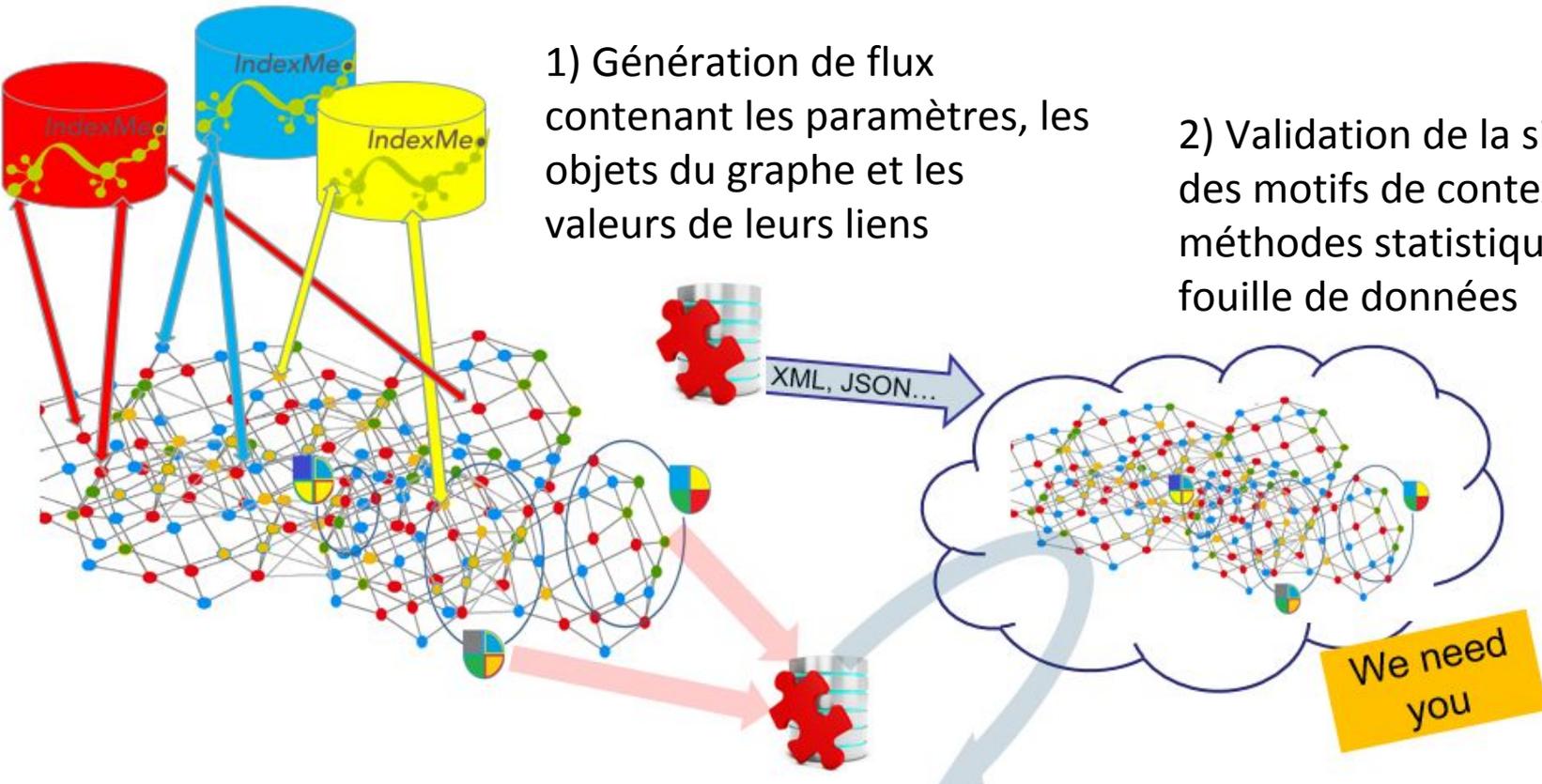


**A****B****C**



### 3) Discussion : recherche opérationnelle et perspectives

Quels systèmes prévisionnels d'aide à la décision ?



1) Génération de flux contenant les paramètres, les objets du graphe et les valeurs de leurs liens

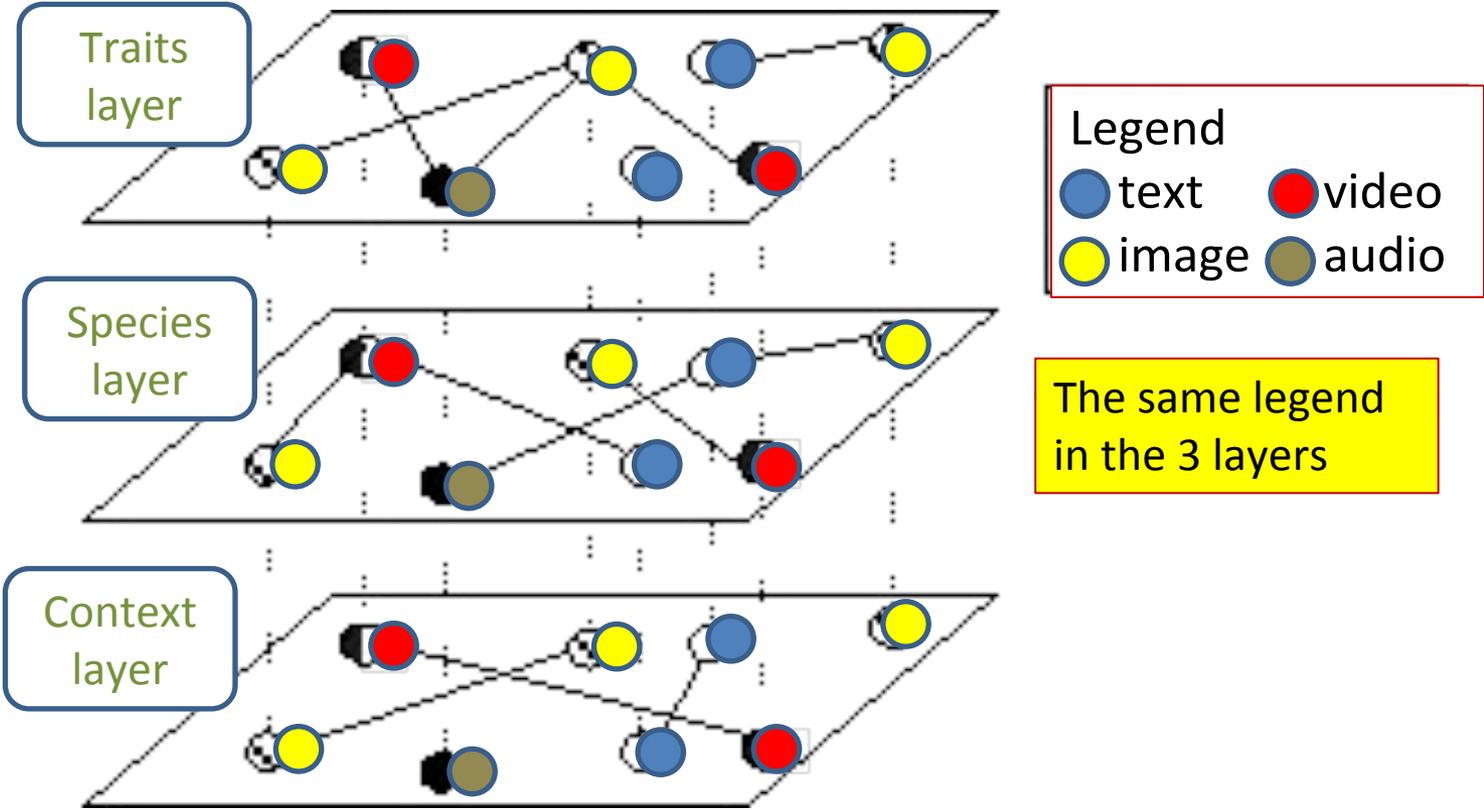
2) Validation de la significativité des motifs de contextes via des méthodes statistiques et de fouille de données

3) Conservation des motifs de contextes et de leur fréquence pour alimenter des outils d'aide à la décision/d'évaluation

### 3) Discussion : la fouille de données

Quels systèmes prévisionnels d'aide à la décision ?

### Multi Layer Graphs





# Bilan à chaud de l'atelier Visualisation de données sous forme de graphes

*Formation* : Romain Bourqui (LaBRI), David Auber (LaBRI)

*Organisation et animation scientifique* : Romain David (Aix Marseille Université, Université d'Avignon, CNRS, IRD, IMBE UMR 7263, Station Marine d'Endoume),  
Anna Cohen Nabeiro (FRB), Aurélie Delavaud (FRB)

*Participation* : Loup Bernard (Université de Strasbourg, UMR 7044 Archimède),  
Guillaume Body (ONCFS), Yvan Le Bras (MNHN, UMR CESCO, Station Biologique  
de Concarneau), Michelle Leydet (IMBE), Sophie Pamerlon (GBIF-France)

Paris, 16 et 17 novembre 2017

# Base de données APECS

## Description de la base de données :

- Données d'observation du requin pélerin dans l'Atlantique Nord-Est et en Bretagne
- Type de données : issues des sciences participatives et opportunistes
- Date de création : 1997
- Type de logiciel : MS Access organisée en trois tables
- Nombre d'utilisateurs : environ 1400
- Taille (Octet) et envergure spatio-temporelle – 5 MégaOctets
- Données accessibles en ligne : non

## Contenu de la base de données :

- Principaux champs porteurs d'information : Zone géographique, année, commune, type d'observation, classe de taille
- Nombre d'enregistrements dans les tables : 1402

# Description du graphe :

-noeuds : Signalement de requins + observateurs + Requins pèlerins

-liens : signalements vers observateurs / signalements vers requins / observateurs vers requins

Tulip 5.1 [Tulip] - C:/Users/ylebras/Documents/65MO\_MNHN/2016/PosteCAD/3\_data\_et\_indexmeed/use\_case\_APECS/Graph\_APECS\_Requin\_Pelerin.tlpx

File Edit Window Help

The screenshot shows the Tulip software interface. On the left is a sidebar with various tool categories: Favorites, Algorithm (Curve edges, Edge bundling, Squarified Tree, Map Helper), Clustering (Equal Value, Hierarchical), Misc (M3 Layout, M3), and Graphs. The main workspace displays a dense network graph with blue edges and red nodes. A table in the center shows data for 'ANWEE' and 'DECADE'.

	ANWEE	DECADE		
0	2002	1	1	
1	2002	1	2	
2	2002	1	2	
3	2002	3	2	
4	2002	3	2	
5	2002	3	2	
6	2002	1	2	
7	2002	1	2	
8	2002	1	3	
9	2002	1	3	
10	2002	1	3	
11	2002	2	3	
12	2002	2	3	
13	2002	2	3	

Algorithms

Filter name...

- Perfect aspect ratio
- Basic
  - Circular
  - Circular (OGDF)
  - Random layout
- Force Directed
  - Bertault (OGDF)
  - Davidson Harel (OGDF)
  - Fast Multipole Embedder (OGDF)
  - FM<sup>3</sup> (OGDF)
  - Frutcherman Reingold (OGDF)
  - GEM (Frick)
  - GEM Frick (OGDF)
  - GRIP
  - Kamada Kawai (OGDF)
  - LinLog

Graphs

Name	Id	Nodes	Edges
graph_0	0	3222	2964
clone sub-graph of grap...	1	3222	2964

Undo

Redo

Workspace

Python IDE

Export

CSV

Import

graph\_0

### Properties

Show all properties + Add new

Show visual properties (view\*)

Show properties matching give a pattern

Name	Type
viewBorderWidth	double
viewColor	color
viewFont	string
viewFontSize	int
viewIcon	string
viewLabel	string
viewLabelBorderColor	color
viewLabelBorderWidth	double
viewLabelColor	color
viewLabelPosition	int
viewLayout	layout
viewMetric	double
viewRotation	double
viewSelection	bool
viewShape	int

Navigate in graph

clone sub-graph of graph\_0

Layers

Scene

1 Add panel 2 Expose 3 Search Python REPL

1 / 1

# Base de données ECOSCOPE

## Description de la base de données :

- Métadonnées de jeux de données d'observation de la biodiversité à différents niveaux d'organisation du vivant
- Type de données : données descriptives des dispositifs et de leurs jeux de données
- Date de création : 2015
- Type de logiciel : PostGre pour le stockage des métadonnées ; application de saisie et export au format XML ; PostGIS pour le stockage des informations géoréférencées
- Nombre d'utilisateurs : 50 à 100 pour les fournisseurs de métadonnées avec codes d'accès, beaucoup plus pour les potentiels utilisateurs (chercheurs, associations, gestionnaires, institutionnels...)
- Taille (Octet) et envergure spatio-temporelle – N/A
- Données accessibles en ligne : Oui sur le site [ecoscope.fondationbiodiversite.fr/ecoscope-portal/](https://ecoscope.fondationbiodiversite.fr/ecoscope-portal/)

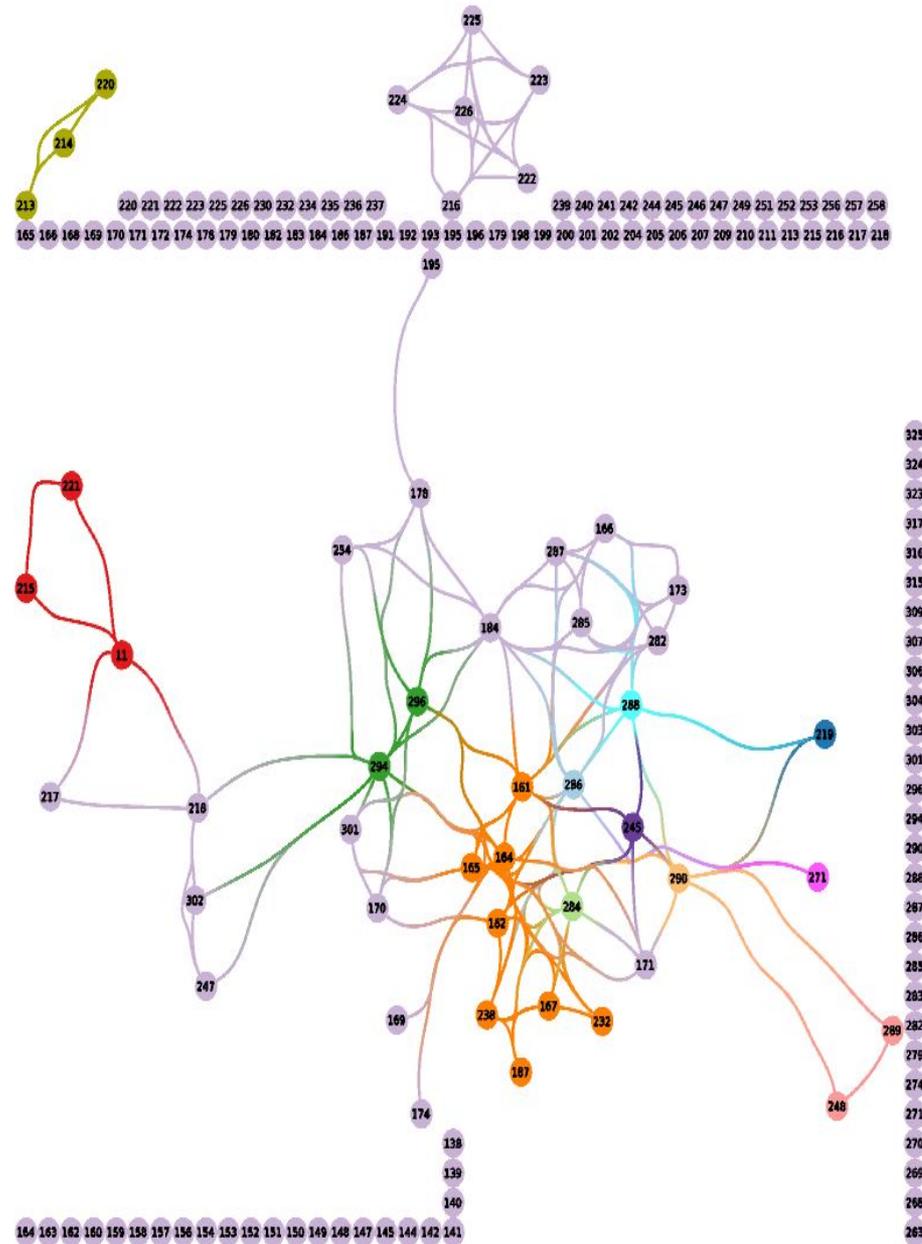
## Contenu de la base de données :

- Principaux champs porteurs d'information :  
Titre, résumé, variables, couvertures temporelle et spatiale, taxons, matériel et méthodes, contacts, conditions d'accès et d'utilisation des données, collection, URL vers la base de données...
- Nombre d'enregistrements dans les tables principales : 50 lignes pour la table dispositif, 150 lignes pour la table jeu de données
- Combien/quel % de champs/descripteurs doivent obligatoirement avoir une valeur : title et starting date sont obligatoires à compléter dans l'application

Nœuds : fiches jeux de données

Liens : en fonction de 4 propriétés (si au moins l'une d'elle est partagée par deux nœuds, il y a création de liens)

Propriétés utilisées : Habitat étudié, Ecosystème étudié, Conditions et contraintes pour l'accès aux ressources génétiques



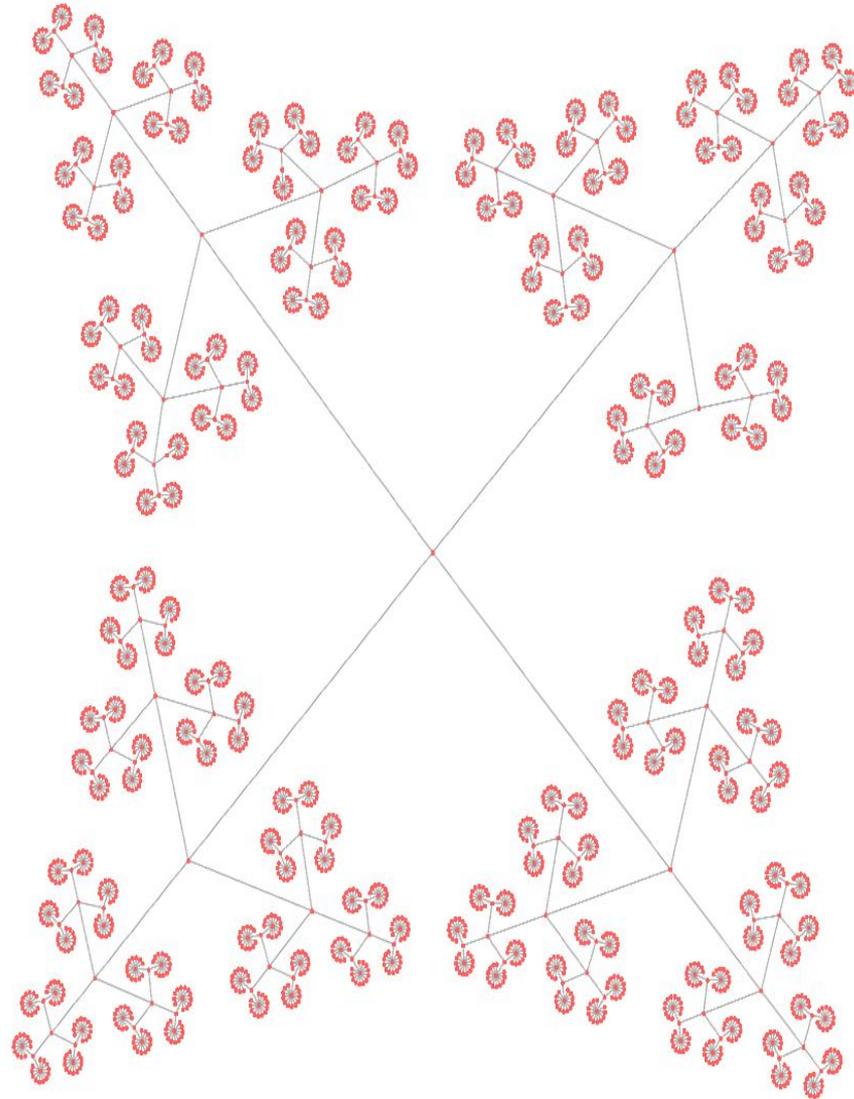
# Base de données DEVOTES

- Description : Depuis 2013, le projet européen DEVOTES (Développement D'Outils innovants pour la compréhension de la biodiversité marine et l'évaluation du bon état environnemental) a utilisé les ARMS (Artificial Reef Monitoring System) pour normaliser le suivi des communautés benthiques dans les substrats durs. Ce programme traite principalement du descripteur 1 de la DCSMM (MSFD en anglais), qui concerne la diversité biologique.

Ce jeu de données contient les premiers résultats concernant la biodiversité macroscopique La colonisation des plaques contenues dans ces récifs après plus d'un an d'immersion dans différentes conditions environnementales dans trois mers d'Europe (mer atlantique, près de la baie de Biscaye, mer Méditerranée du nord-ouest et de la mer Adriatique), ainsi que la mer Rouge. Les compositions communautaires, déduites de photographies, sont décrites par régions marines, par sites et par surfaces distinctes des dispositifs ARMS; Divers facteurs et variables environnementales reflétant à la fois le niveau d'anthropisation et la diversité de l'habitat local décrit chaque site.

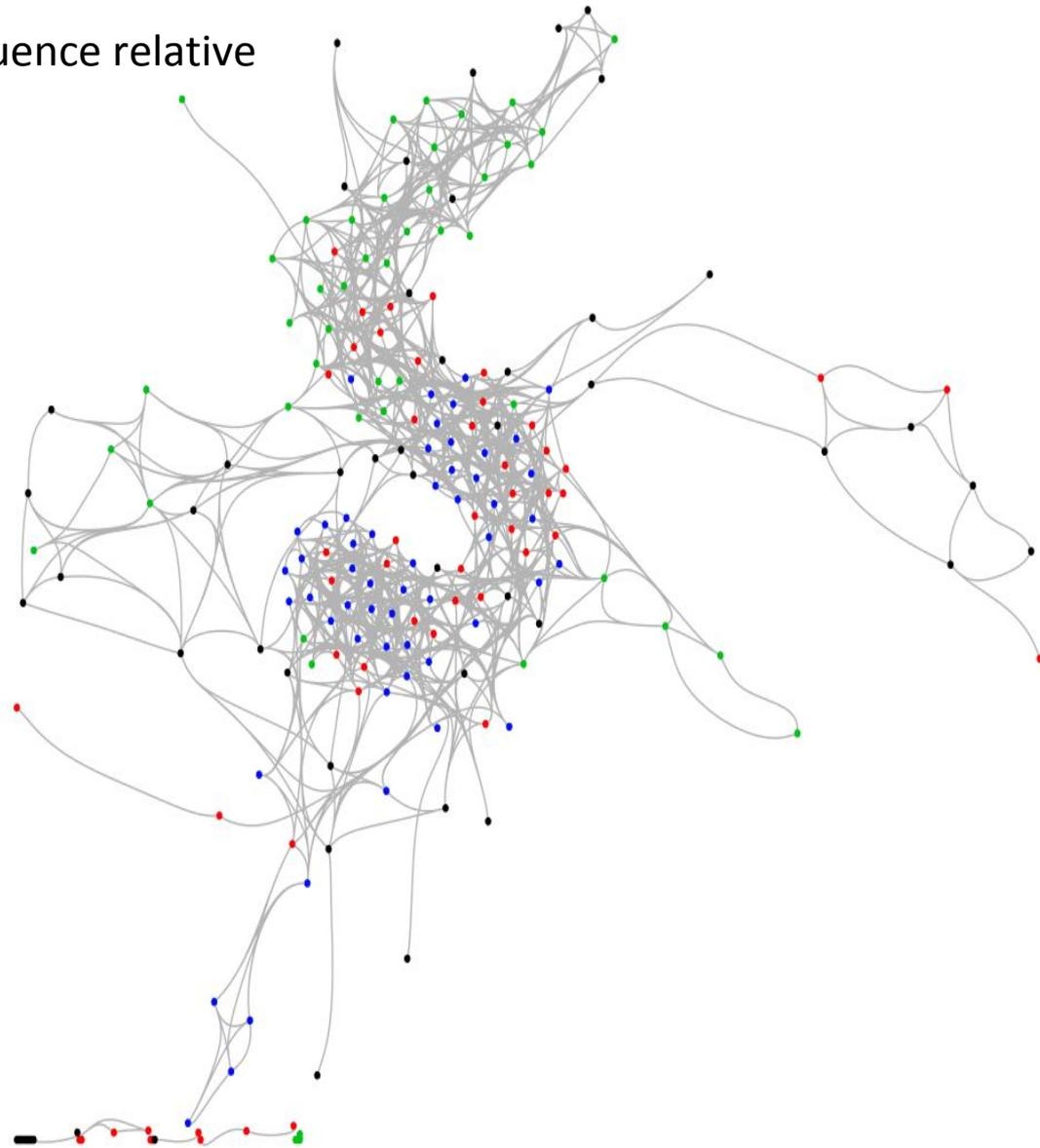
- Type de données : Fréquence de taxons reconnaissable sur photo
- Date de création : 2015
- Type de logiciel : Excel, csv
- Nombre d'utilisateurs : 10
- Taille (Octet) et envergure spatio-temporelle : 1 installation, d'autres installations sont prévues à l'avenir, avec un protocole de suivi temporel, 1,4 MO
- Données accessibles en ligne : Mise en place dans un entrepôt de données en open access dès la parution de la publication scientifique

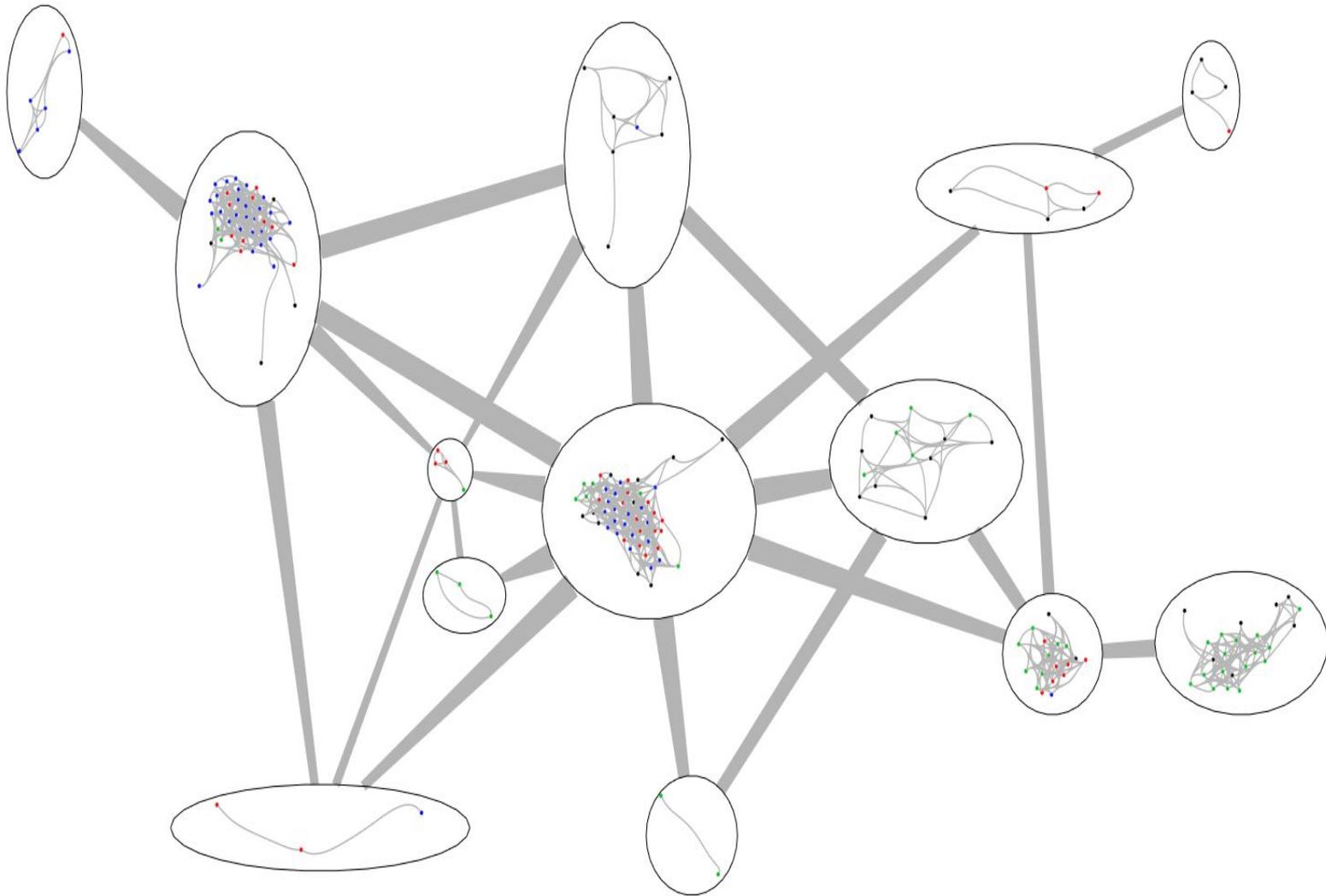
Plan d'échantillonnage qui permet de mettre en évidence des éléments manquants (une ARMS et une plaque c'est-à-dire 2 f0000)



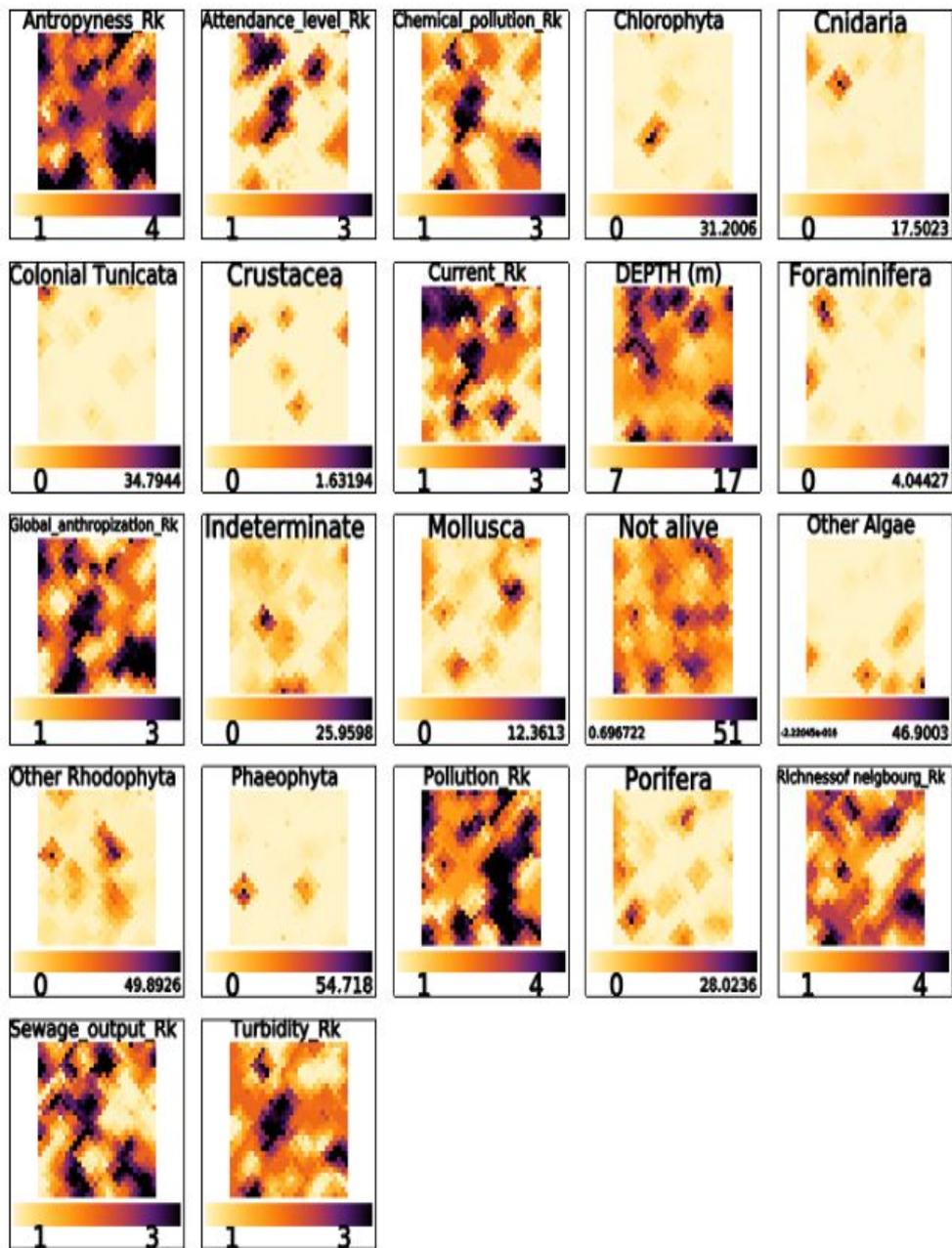
Nœuds : faces de plaque d'ARMS dans 4 mers

Lien : similarité de fréquence relative









# Base de données ArkéoGIS

Données d'archéologie funéraire en vallée du Rhin supérieur en provenance de plusieurs bases de données.

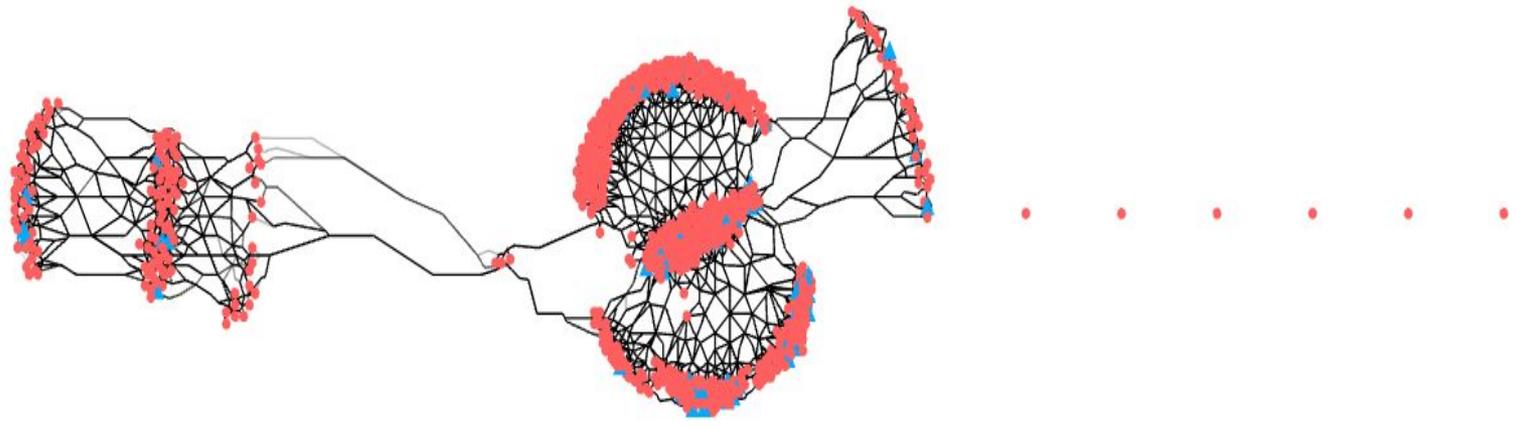
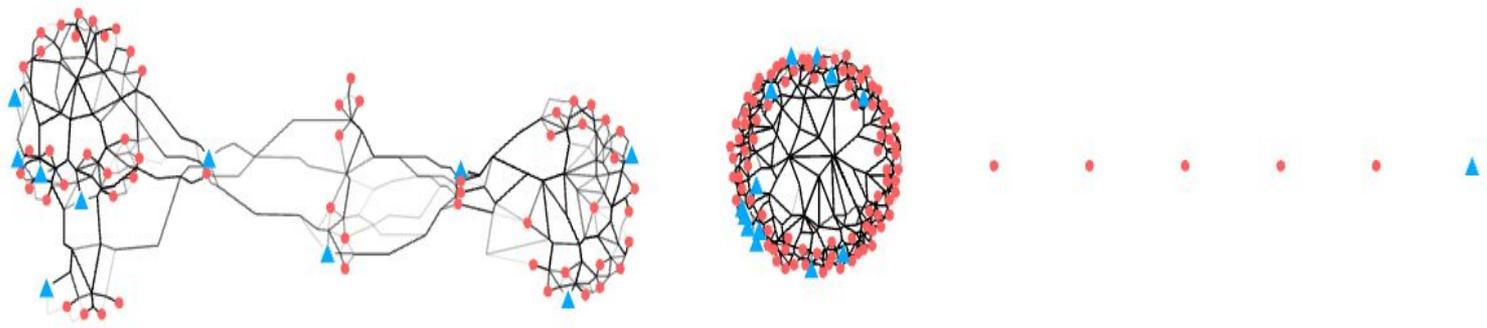
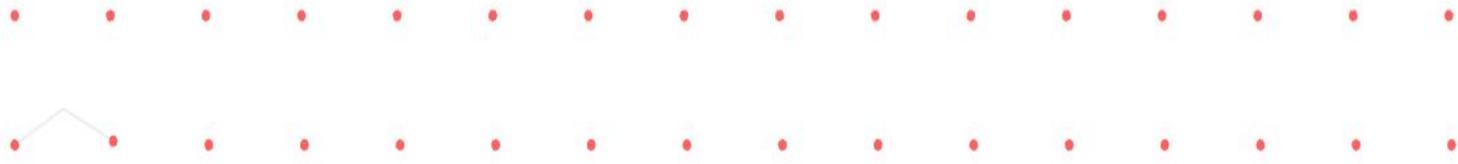
La question était de savoir si la pratique de l'inhumation est variable dans le temps.

Nœuds : tombes

Arêtes : définis sur les chronologies par rapport à l'ensemble.

Deux tombes sont reliées si les fourchettes de dates de fin (champ `ENDING_PERIOD`), coupé en deux dates numériques via la console python correspondent.

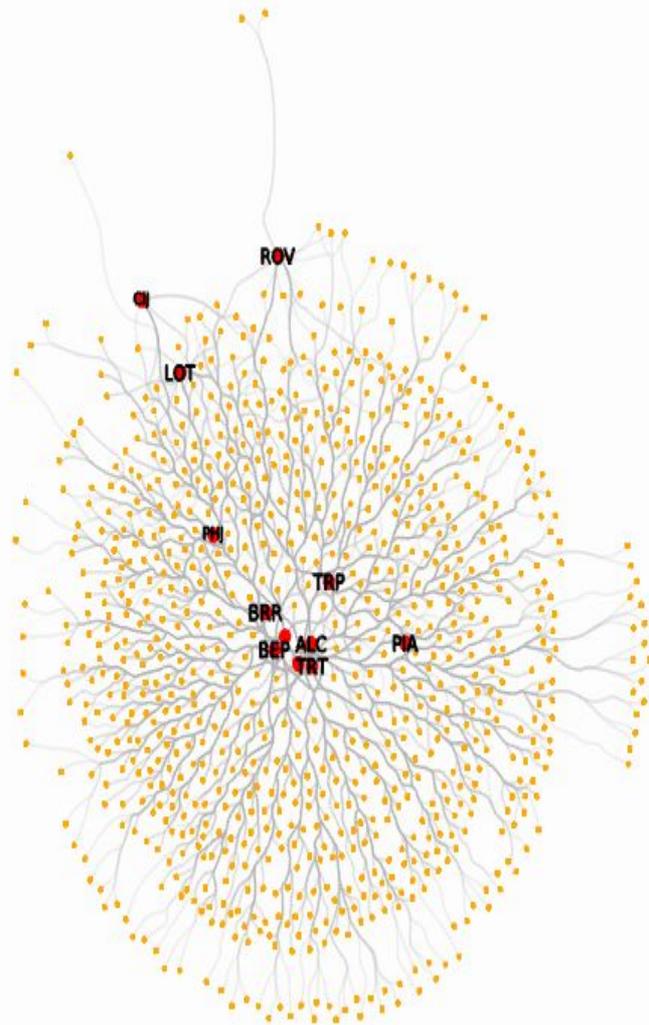
Le graphe permet de faire apparaître par période et par base de données les tombes à inhumation, il fait apparaître plusieurs groupes et des bases liens, plus large dans leur définition (chronologique ou spatiale)

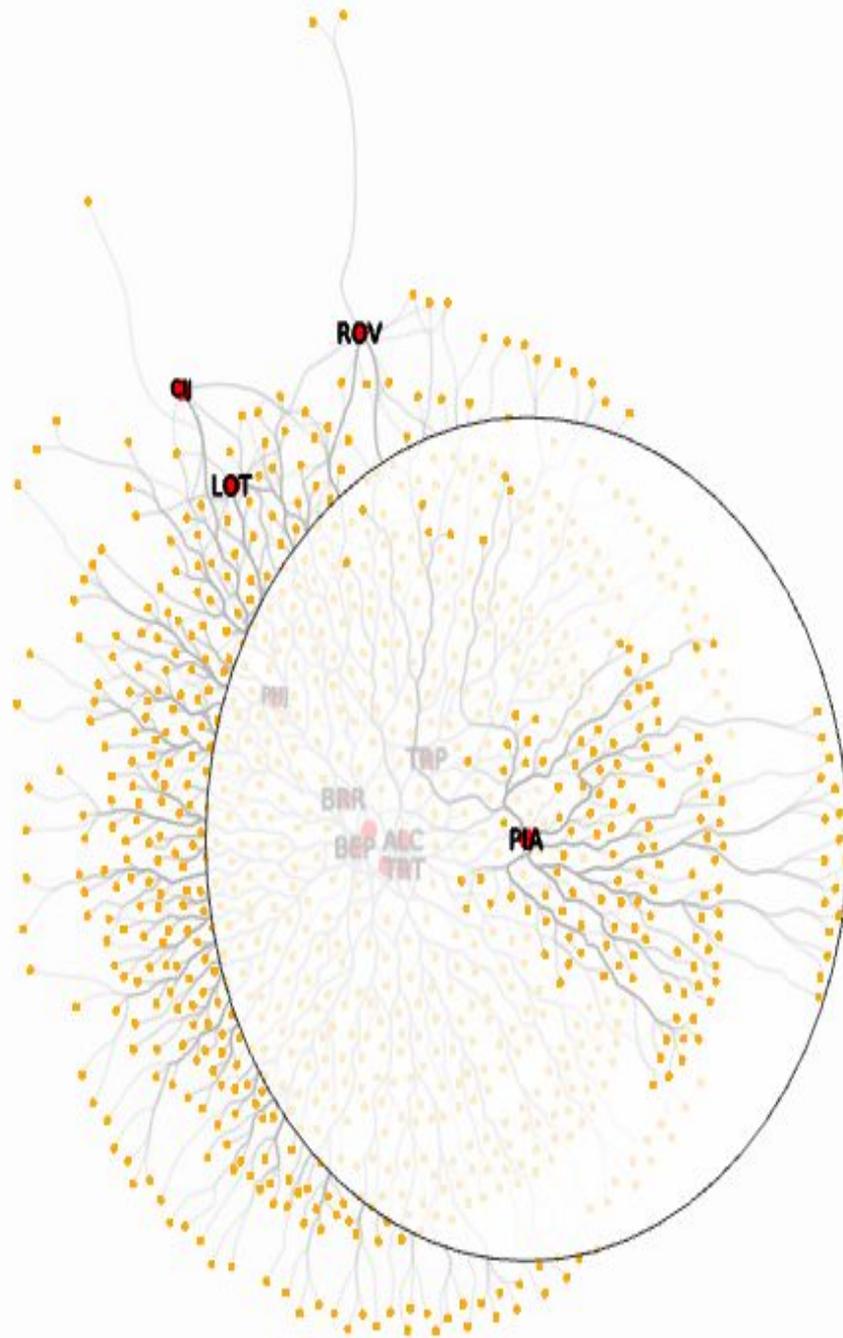


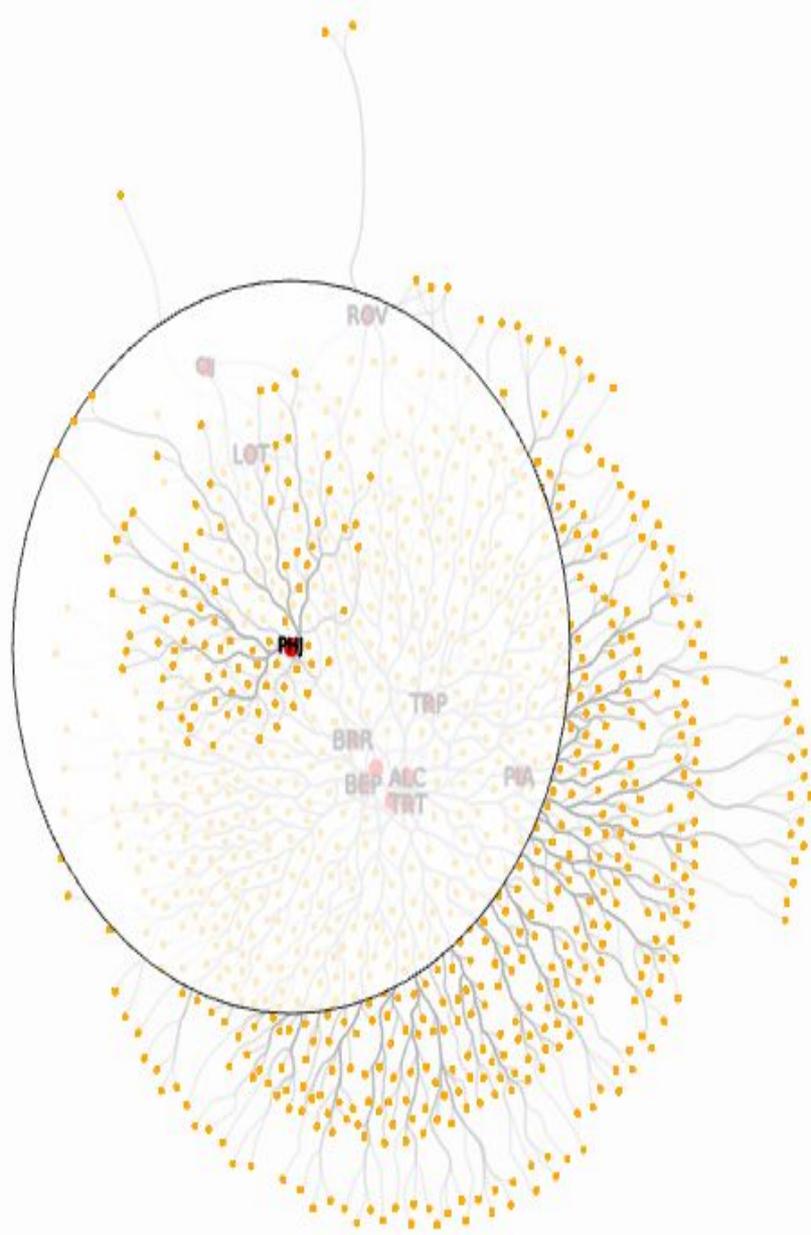
# Base de données ONEPF

Les données de bases sont des points de comptages d'avifaune.

Les points oranges sont les sites, les rouges sont les espèces. Les sous graphes démontrent que certaines espèces ne se retrouvent pas dans les mêmes sites (spécialistes) et d'autres sont plus généralistes.

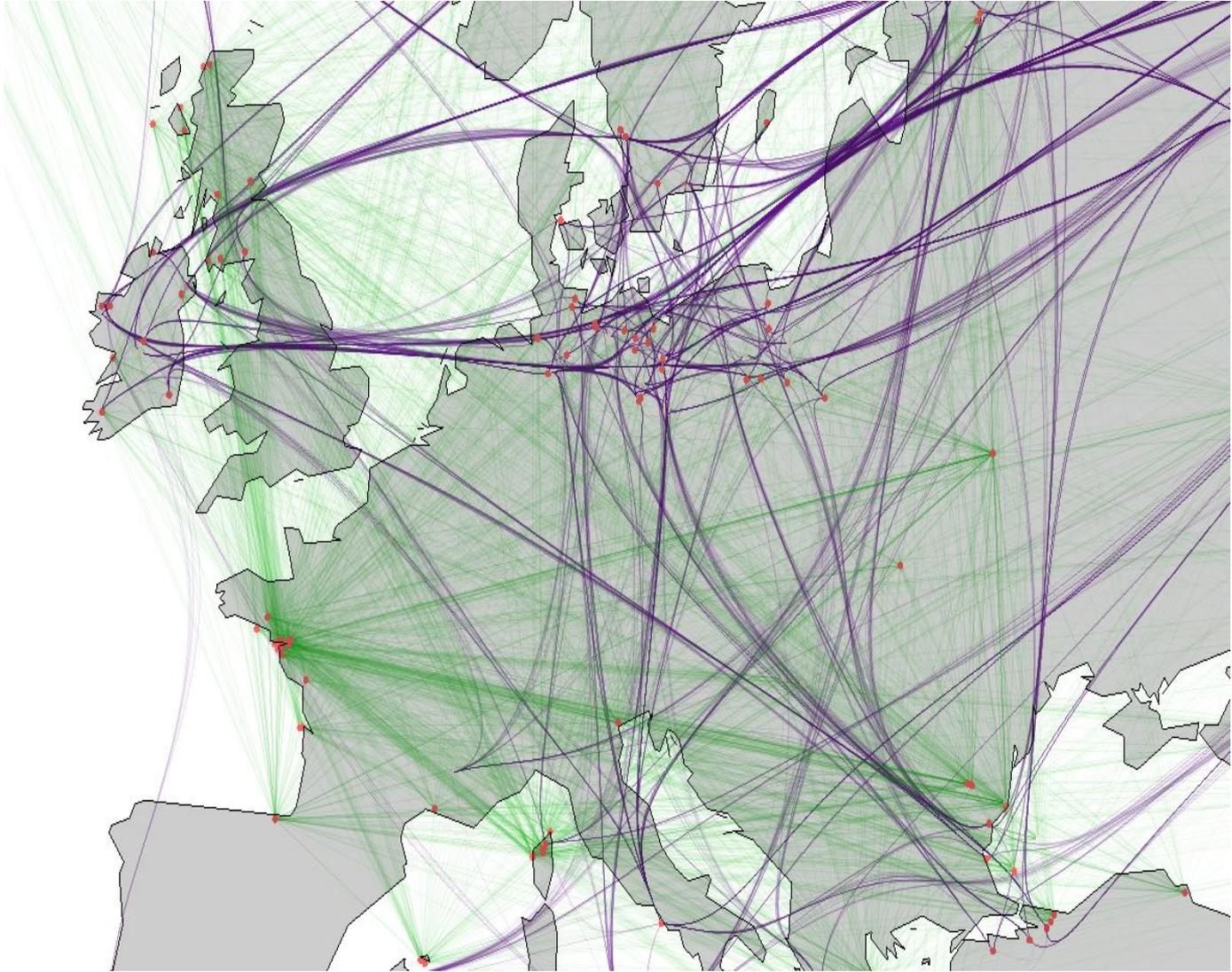




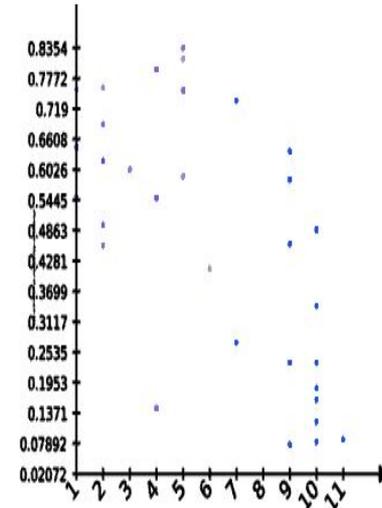
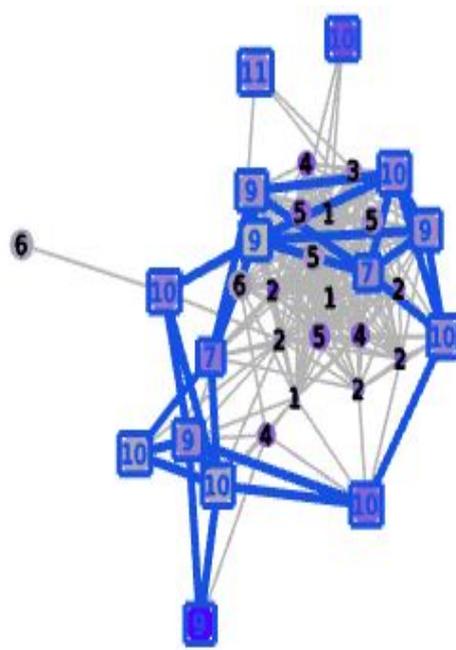
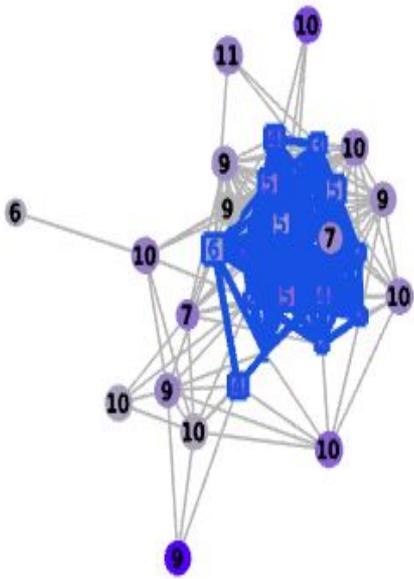


# Base de données European Pollen Database

- **Principaux champs porteurs d'information :**
- palynologic sites
- entity by site
- location by country
- elevation
- type of site
- palynologist
- contact
- entity dated [age] or undated
- 
- Nombre d'enregistrements dans les tables : dizaines de milliers d'entrées
- 1 table: 1812 entities dated and 973 entities undated
- 
- Combien/quel % de champs/descripteurs doivent obligatoirement avoir une valeur :
- tous les champs sont remplis sauf pour les enties non datées



# Base de données Rennes



# 8 fiches cas d'étude en cours d'amélioration



## NOM DU CAS D'ETUDE

Situid tamen quidem alim propinquant



### DESCRIPTION DU CAS

Situid tamen clausos vehementer angebat, quod captis navigis, quae frumenta vehebant per flumen, Isauri quidem alimentorum copis adfluebant, ipsi vero solitarum rerum cibos iam consumendo inediae propinquantis aerumnas exitialis horrebant.

Situid tamen clausos vehementer angebat quod captis navigis, quae frumenta vehebant per flumen, Isauri quidem alimentorum copis adfluebant, ipsi vero solitarum rerum cibos iam consumendo inediae propinquantis aerumnas exitialis horrebant.

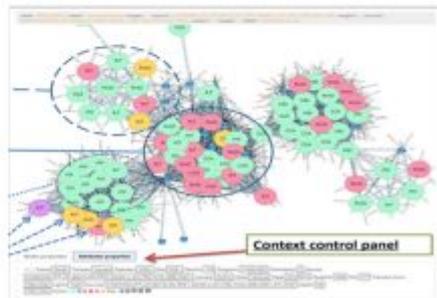
Isauri quidem alimentorum copis adfluebant, ipsi vero solitarum rerum cibos iam consumendo inediae propinquantis aerumnas exitialis horrebant.

Contact: ---- at ----.

### QUESTION 1 OU CONCLUSION EN QUELQUES MOTS ATTIRANT L'ATTENTION

Novo denique perniciosoque exemplo idem Gallus ausus est inire flagitium grave, quod Romae cum ultimo dedecore temptasse aliquando dicitur Gallienus, et adhibitis paucis clam ferro succinctis vesperi per tabernas palabatur et conpita quaeritando Graeco sermone, cuius erat inpendio gnarus, quid de Caesare quisque sentiret.

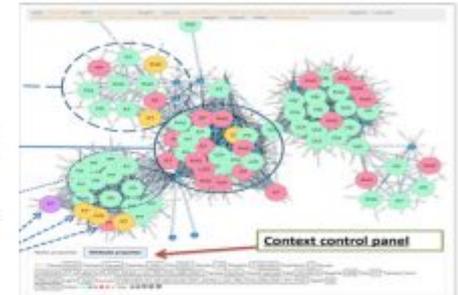
- et haec confidenter agebat in urbe ubi pemoctantium luminum claritudo dierum solet imitari fulgorem.
- postremo agnitus saepe iamque, si prodisset, conspicuum se fore contemplans, non nisi luce palam egrediens ad agenda quae putabat seria cerebatur.



### QUESTION 2 OU CONCLUSION EN QUELQUES MOTS ATTIRANT L'ATTENTION

Novo denique perniciosoque exemplo idem Gallus ausus est inire flagitium grave, quod Romae cum ultimo dedecore temptasse aliquando dicitur Gallienus, et adhibitis paucis clam ferro succinctis vesperi per tabernas palabatur et conpita quaeritando Graeco sermone, cuius erat inpendio gnarus, quid de Caesare quisque sentiret.

- et haec confidenter agebat in urbe ubi pemoctantium luminum claritudo dierum solet imitari fulgorem.
- postremo agnitus saepe iamque, si prodisset, conspicuum se fore contemplans, non nisi luce palam egrediens ad agenda quae putabat seria cerebatur.



### Les prochaines étapes...

Novo denique perniciosoque exemplo idem Gallus ausus est inire flagitium grave, quod Romae cum ultimo dedecore temptasse aliquando dicitur Gallienus, et adhibitis paucis clam ferro succinctis vesperi per tabernas palabatur et conpita quaeritando Graeco sermone, cuius erat inpendio gnarus, quid de Caesare quisque sentiret.

### CE QUE LES GRAPHES APPORTENT...

Pandente itaque viam fatorum sorte tristissima, qua praestitutum erat eum vita et imperio spoliari, itineribus interiectis permutatione iumentorum emensis venit Pelobionem oppidum Noricorum, ubi reseratae sunt insidarum latebrae omnes, et Barballo repente apparuit comes, qui sub eo domesticis praefuit, cum Apodemio agente in rebus milites ducens, quos beneficiis suis oppigheratos elegerat imperator certus nec praemis nec miseratione ulla posse deflecti.

#### Le consortium IndexMEED

Pandente itaque viam fatorum sorte tristissima, qua praestitutum erat eum vita et imperio spoliari, itineribus interiectis permutatione iumentorum emensis venit Pelobionem oppidum Noricorum, ubi reseratae sunt insidarum latebrae omnes.

Site web: www.  
Contact: --- at ---

Le programme de travail 2017 a été soutenu par

# Perspectives pour GRAMINEES 2018

Organisation d'ateliers sur des compétences complémentaires s'appuyant sur les cas d'études développés dans le cadre des ateliers de 2017.

Les thèmes proposés

- « la sémantique et les ontologies interdisciplinaires dans le domaine environnemental »
- « Algorithmes et calculs sur les graphes basés sur des données en sciences de l'environnement »

Etape importante : la mise en place d'une démonstration de fouilles de graphes basés sur des données environnementales utilisant la grille de calcul (soutien France Grilles) avec les problèmes de passage à l'échelle.

(Dino IENCO ->Luc HOGIE)

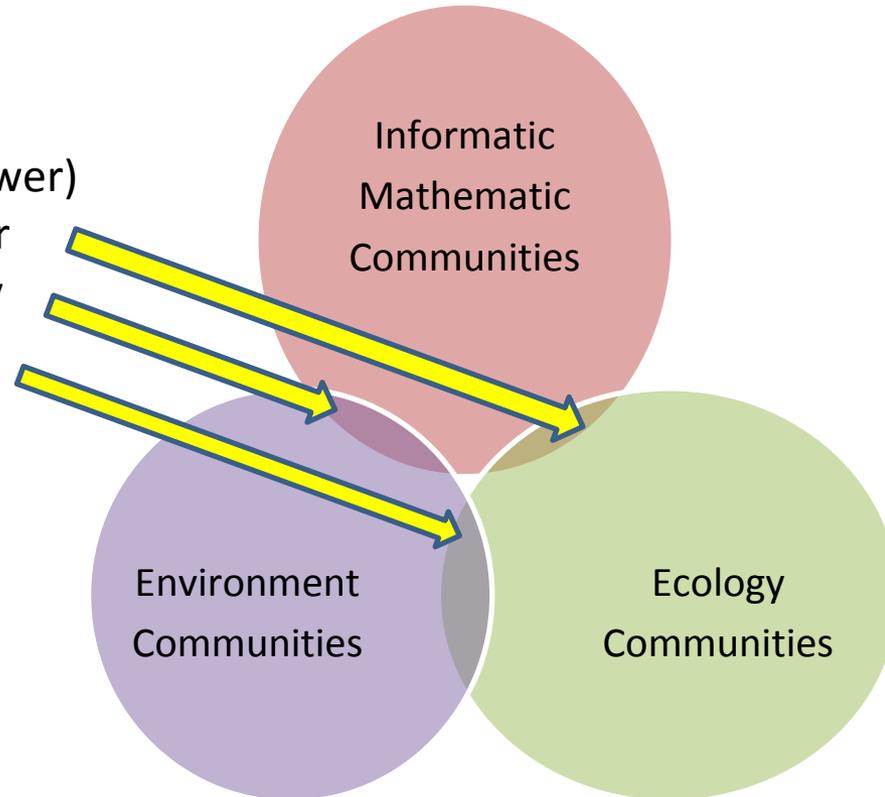
Ce que l'on va (aussi) améliorer en 2018: mises en synergies avec d'autres actions de MaDICS notamment :

- Préservation des données
- Reproductibilité des données
- Qualité des données
- ...?

## Carte des compétences: quels domaines de compétence?

Quality of communities links : more or less strong

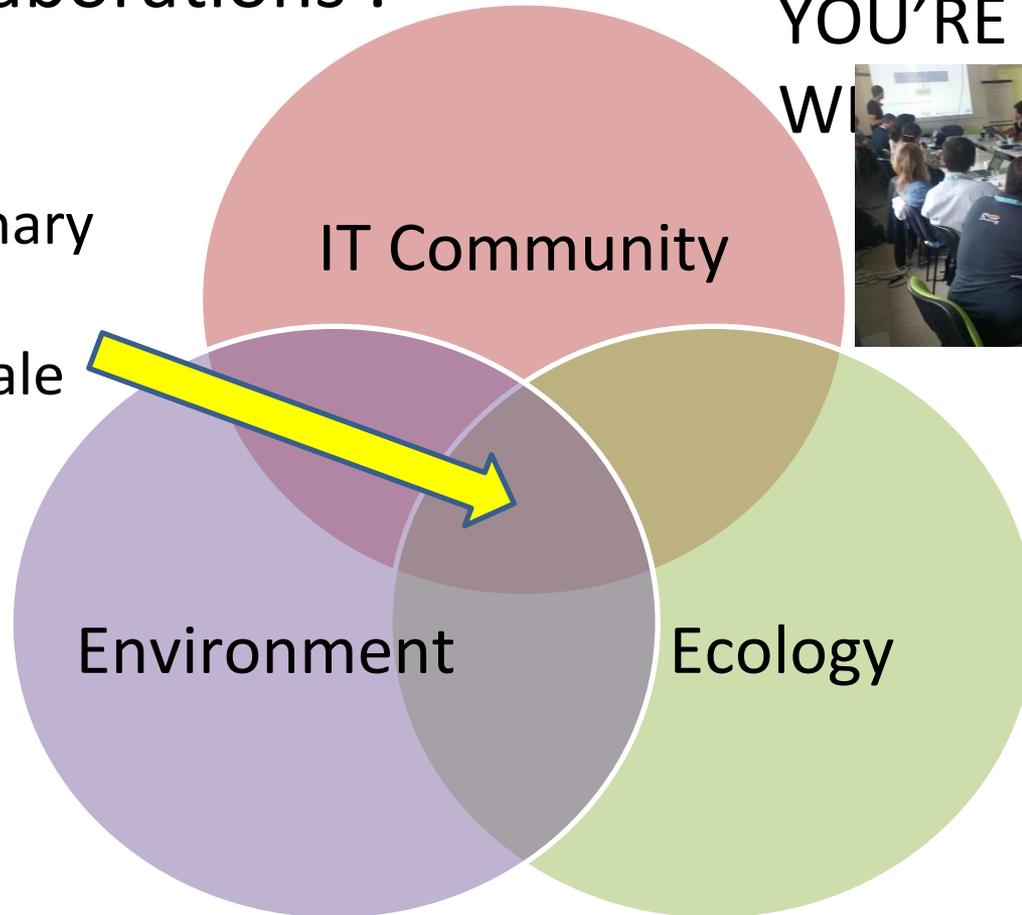
Only a few part  
of time (man power)  
is really spent for  
Inter-disciplinary  
approach



Merci de votre attention

STRONGER collaborations :

**2018 Goal :**  
Transdisciplinary  
Workshops  
at a larger scale



YOU'RE  
WI



# Graminées: quels domaines de compétence?

## Proposition de structure du projet

1. Web sémantique

- Normalisation
- Qualification
- Interopération (open data?)

2. Cycle de vie et qualité

- Indexation
- Traçabilité des transformations
- Résolution de l'index

3. Concepts à partir des données

- **Prototype de visualisation**
- Modélisation de Graphes
- **Re-qualification et alignements (thésaurus, ontologies)**

4. Mise à disposition des matrices

- Bancarisation des concepts
- Mixe en place des FLUX « concepts »
- Architecture de requête sur les FLUX

5. Calcul (Grille et Mesocentres)

- Dépôt de projets de calculs
- Parcours de graphes, **apprentissage**
- Vérification statistique et algorithmique

6. Démarche opérationnelle

- Bancarisation des motifs
- Mixe en place des FLUX « opérationnels »
- Opération par les outils d'aide à la décision

2018

(§9) Romain David, Maria C. Uyarra, Susana Carvalho, Holger Anlauf, Angel Borja, et al. Lessons from photo analyses of Autonomous Reef Monitoring Structures, as tools to detect geographical, spatial, and environmental effects" 2018 submitted, Marine Pollution Bulletin.

(§8) Romain David, Anna Cohen Nabeiro, Jean-Pierre Féral, Aurélie Delavaud, Anne-Sophie Archambeau, et al., 2018 accepted, Bilan des journées du GRAAL 2016 et avenir de l'utilisation des graphes en écologie, Nature, Science et Société, 18-3.

2017

(§7) Víctor Méndez Muñoz, Anna Cohen-Nabeiro, Romain David, Vicente Ivars Camáñez, Alfons Nonell-Canals, et al.. Analysis on the Graph Techniques for Data-mining and Visualization of Heterogeneous Biodiversity Data Sets. Complexis 2017, Apr 2017, Porto, Portugal. pp.144 - 151, 2017, Proceedings of the 2nd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2017). <<http://www.complexis.org/?y=2017>>. <10.5220/0006379701440151>. <hal-01541140>

(§6) Romain David, Loup Bernard, Cyrille Blanpain, Alrick Dias, Jean-Pierre Feral, et al.. Visualisation de données sous forme de graphes en archéologie. Rencontre opérationnelle des archéologues d'ArkeoGIS et des écologues d'IndexMed . Digital Archaeology, iste open science, 2017, 17-1 (1), <<https://www.openscience.fr/Data-visualisation-in-archaeology-based-on-graph-approach-Operational-meeting>>. <hal-01617580>

2016

(§5) Romain David, Jean-Pierre Feral, Anne Archambeau, Nicolas Bailly, Cyrille Blanpain, et al.. IndexMed projects: new tools using the CIGESMED DataBase on Coralligenous for indexing, visualizing and data mining based on graphs. Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.). 8th International Congress on Environmental Modelling and Software, Jul 2016, Toulouse, France. Brigham Young University BYU Scholars Archive, International Environmental Modelling and Software Society (iEMs) 8th International Congress on Environmental Modelling and Software Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.) <http://www.iemss.org/society/index.php/iemss-2016-proceedings>, 8th, pp.32, 2016, International Congress on Environmental Modelling and Software. <[http://scholarsarchive.byu.edu/iemssconference/2016/?utm\\_source=scholarsarchive.byu.edu%2Fiemssconference%2F2016%2FStream-C%2F32&utm\\_medium=PDF&utm\\_campaign=PDFCoverPages](http://scholarsarchive.byu.edu/iemssconference/2016/?utm_source=scholarsarchive.byu.edu%2Fiemssconference%2F2016%2FStream-C%2F32&utm_medium=PDF&utm_campaign=PDFCoverPages)>. <hal-01425559>

2015

(§4) Romain David, Jean-Pierre Feral, Sophie Gachet, Alrick Dias, Cyrille Blanpain, et al.. A first prototype for indexing, visualizing and mining heterogeneous data in Mediterranean ecology: within the IndexMed consortium interdisciplinary framework. 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Nov 2015, Bangkok, Thailand. IEEE Explore, 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 232-239, 2015, 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). <<http://ieeexplore.ieee.org/document/7400571/>>. <10.1109/SITIS.2015.119>. <hal-01433600>

(§3) Romain David, Sophie Dubois, Zinovia Erga, Dorian Guillemain, Laure Thierry de Ville d'Avray, et al., (2015) "CIGESMED\*'s protocol and network (Coralligenous based Indicators to Evaluate and Monitor the « good ecological status » of the MEDiterranean coastal waters)", CIGESMED Project. Pages 828-843 ; CNR-IBIMET Florence (Italy), December 2014, ISBN : 978-88-95597-19-5 (Fifth Symposium Monitoring of Mediterranean coastal areas : problems and measurement techniques Livorno (Italy) 17-19 June 2014)

2014

(§2) Romain David, Christos Arvanitidis, Melih Ertan Çinar, Stéphane Sartoretto, A Doğan, et al.. CIGESMED habitat's characterization: a simple and reusable typology at the Mediterranean scale. 2nd Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions, Oct 2014, Portorož, Slovenia. Proceedings of the 2nd Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions. <hal-01620541>

(§1) Romain David, Christos Arvanitidis, Melih Ertan Çinar, Stéphane Sartoretto, Sophie Dubois, et al.. CIGESMED protocols: CIGESMED Protocols : how to implement a multidisciplinary approach on a large scale for coralligenous habitats surveys. RAC/SPA 2nd Mediterranean Symp. on the Conservation of coralligenous and other calcareous bio-concretions, Oct 2014, Portorož, Slovenia. <10.13140/2.1.1895.0086>. <hal-01620550>

Jean-Pierre Féral, Christos Arvanitidis, Anne Chenuil-Maurel, Melih Ertan Çinar, Romain David, et al.. CIGESMED: Coralligenous based indicators to evaluate and monitor the "good environmental statut" of the Mediterranean coastal waters, a SeasEra project. 2nd Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions, Oct 2014, Portorož, Slovenia. Proceedings of the 2nd Mediterranean Symposium on the conservation of Coralligenous & other Calcareous Bio-Concretions. <hal-01620607>

## Challenge du cycle de vie des données : quelle complexité?

