

**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)

par Romain DAVID

thèse soutenue pour l'obtention d'un doctorat en océanographie



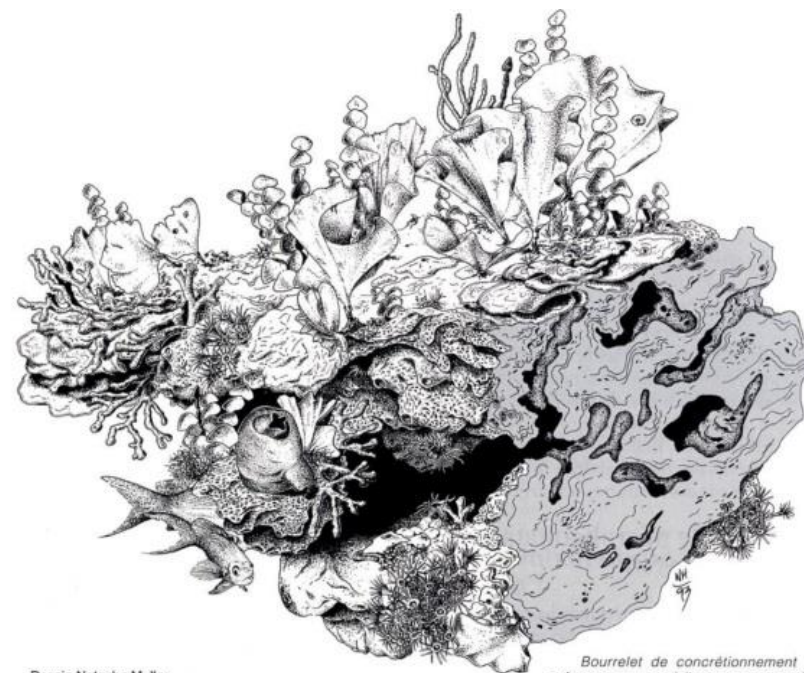
Introduction sur les inférences et les liens de causalité

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

Recommandations et perspectives



Dessin Natacha Muller
tiré de "Invitation sous l'écume"
J.G. Harmelin - Parc national de Port-Cros

Bourrelet de concrétionnement coralligène
anfractueux, vu partiellement en coupe. Construit par
des strates d'algues calcifiées. Il offre un habitat à une foule
d'invertébrés et d'algues.

Figure 1 Schéma d'un bloc de coralligène (d'après Laubier, 1966)

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'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...**

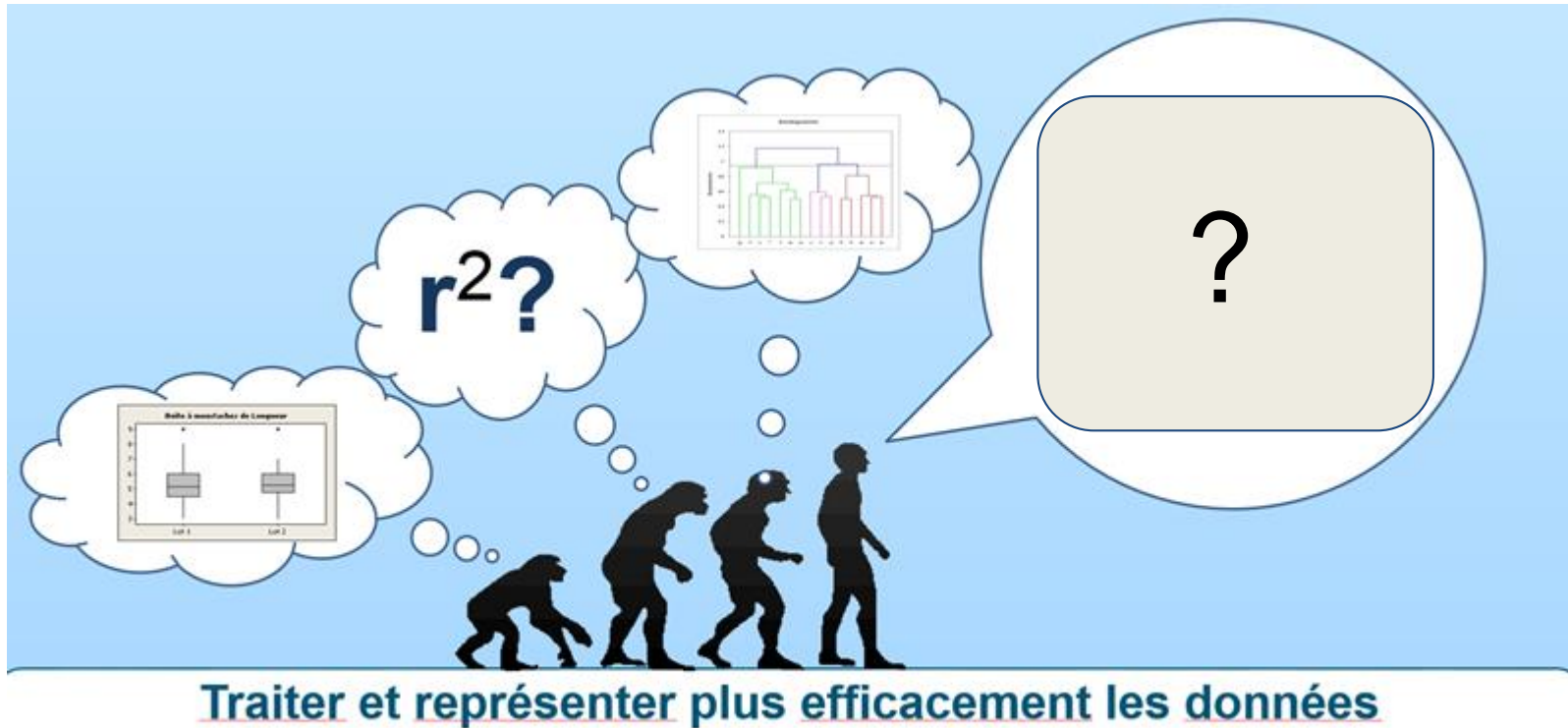
→ INTEROPERABLES et MULTIDISCIPLINAIRES

A large échelle : quels enjeux sur la biodiversité?



La dernière “une” de l’AFB

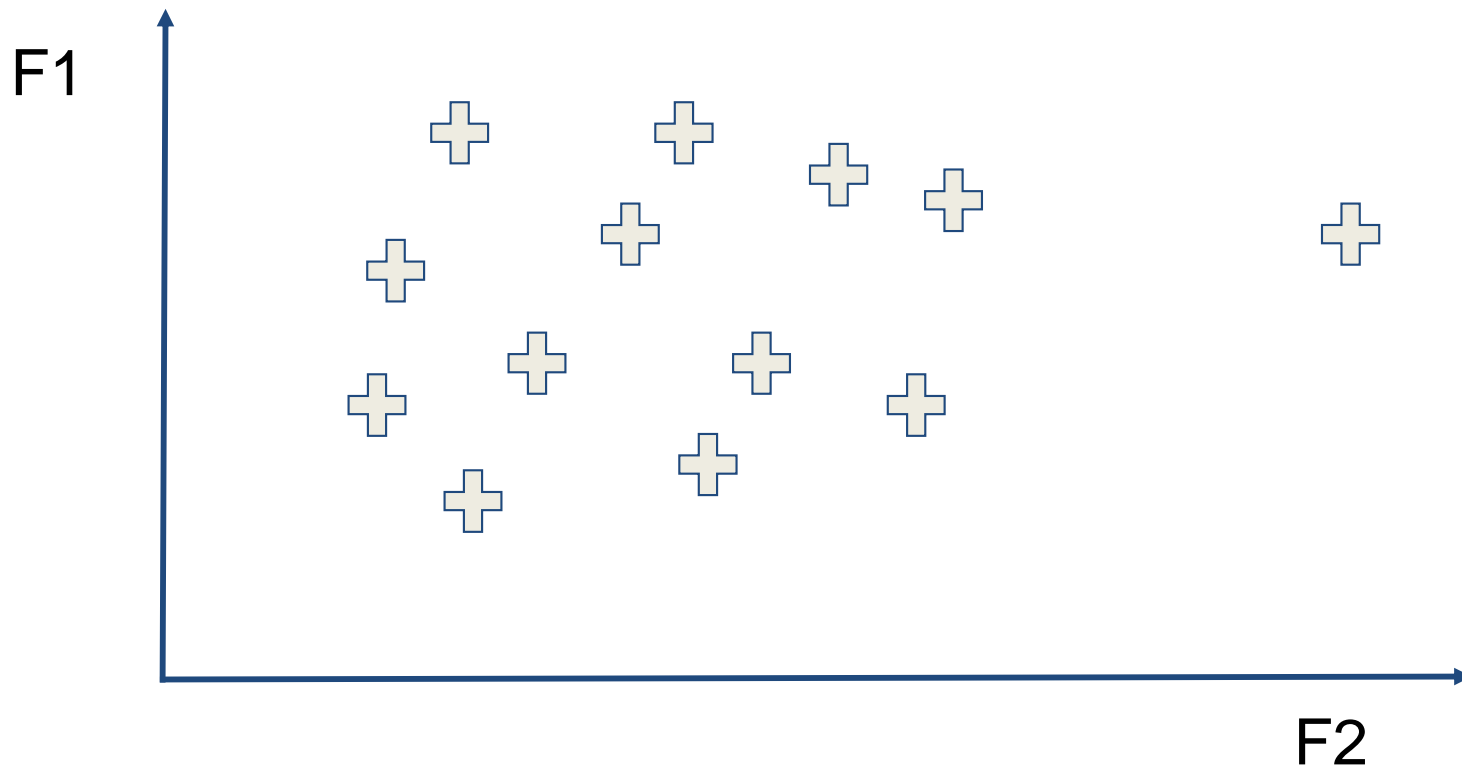
à titre d’exemple, en 10 ans
30 % d’oiseaux en moins
80 % d’insectes en moins



Par quoi commencer : les statistiques ou la visualisation ?

A large échelle : que cherche-t-on ?

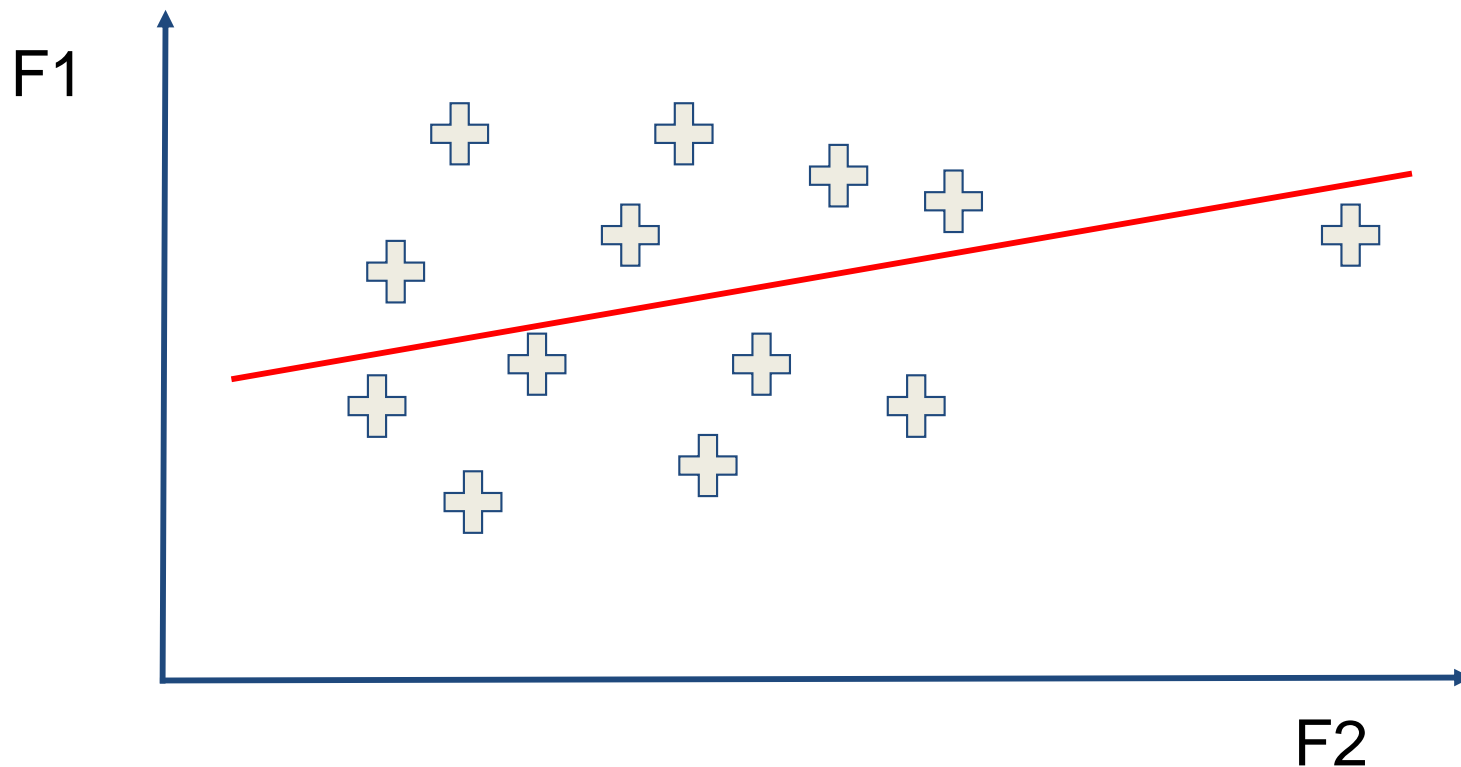
De manière classique, en science : recherche des “inférences”



A large échelle : que cherche-t-on ?

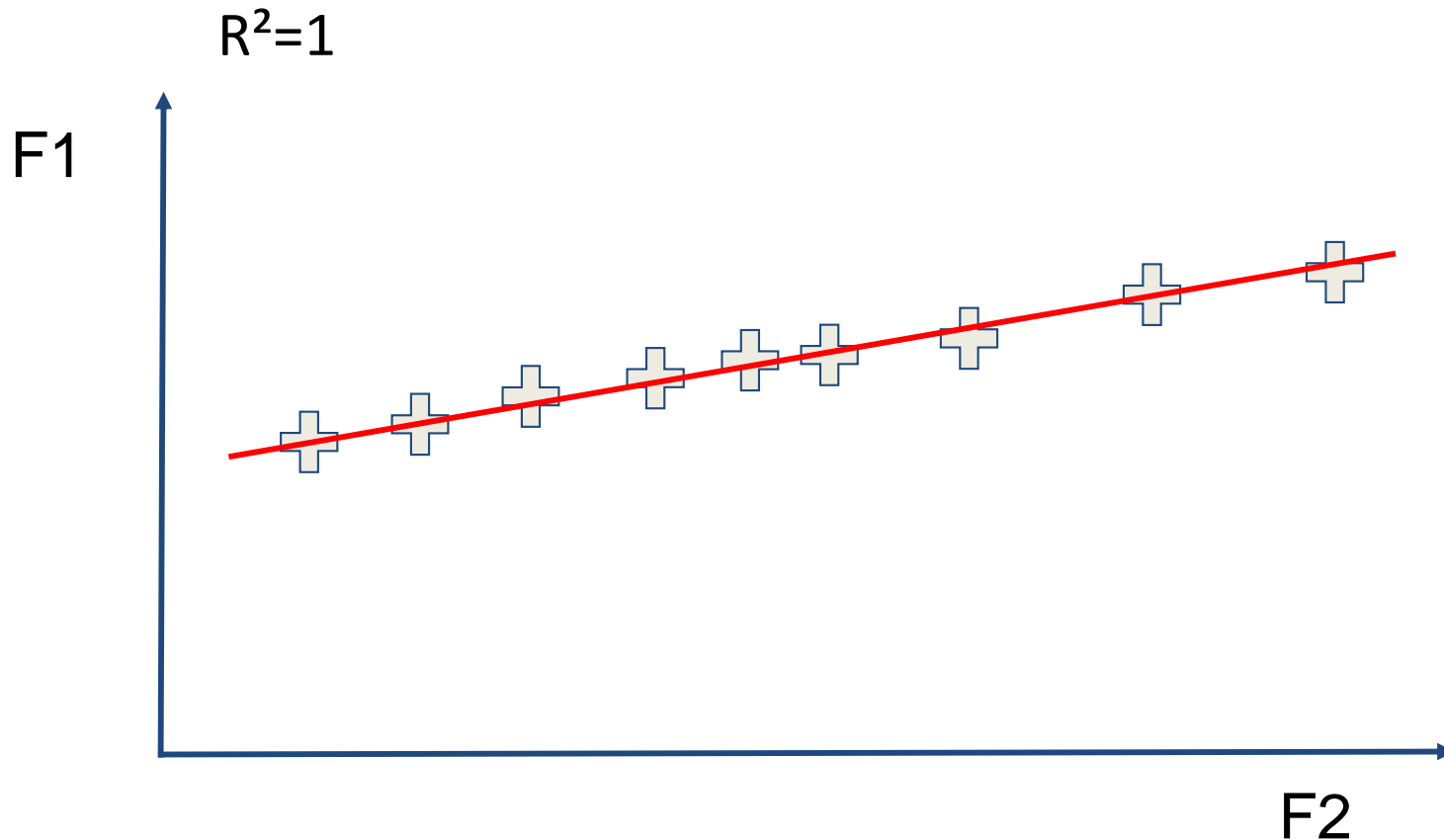
De manière classique, en science : recherche des “inférences”

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



A large échelle : que cherche-t-on ?

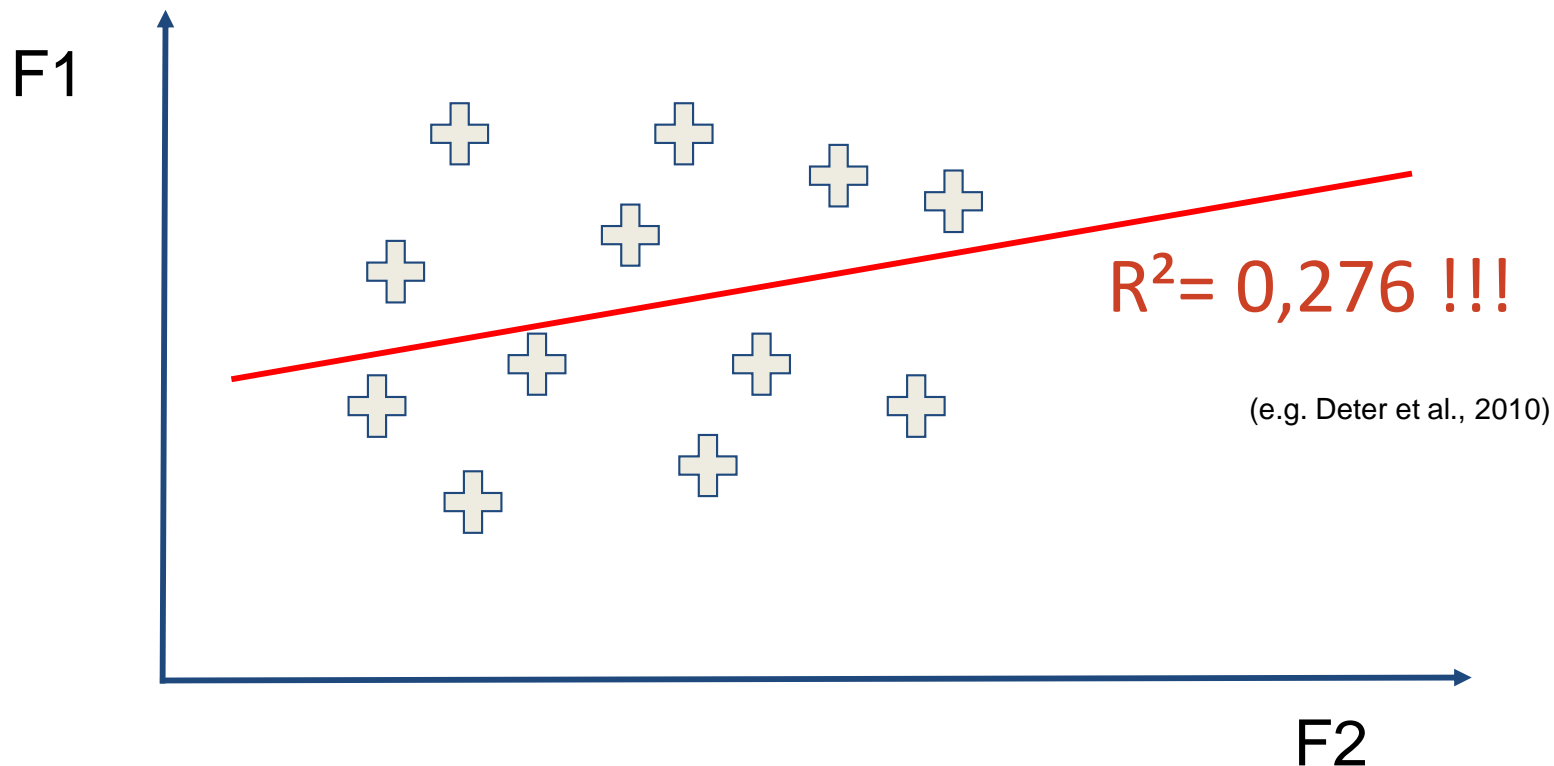
De manière classique, en science : recherche des “inférences”



A large échelle : que cherche-t-on ?

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

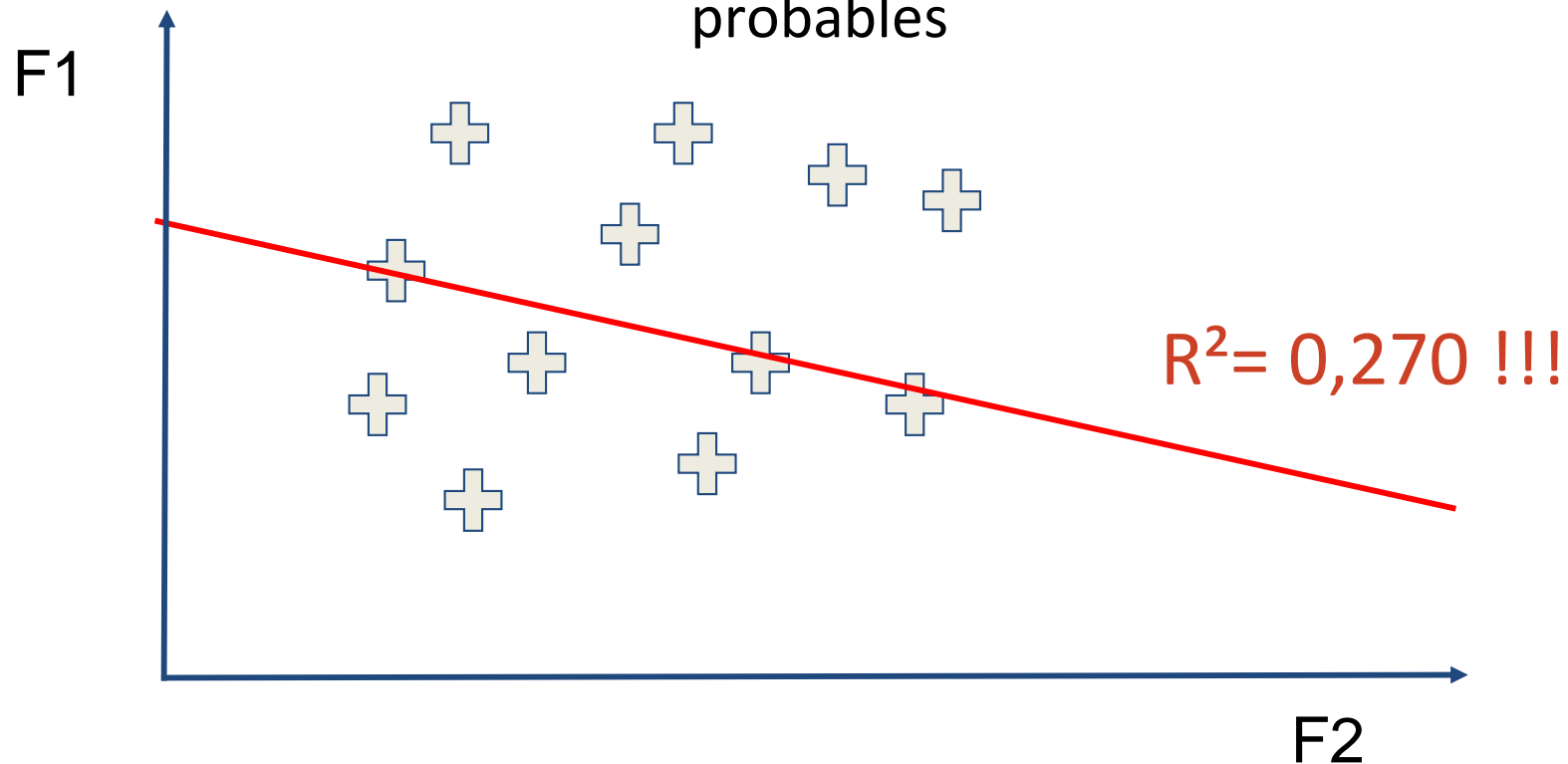
R^2 = parfois pas si proche de 1 pour les études en écologie



A large échelle : que cherche-t-on ?

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

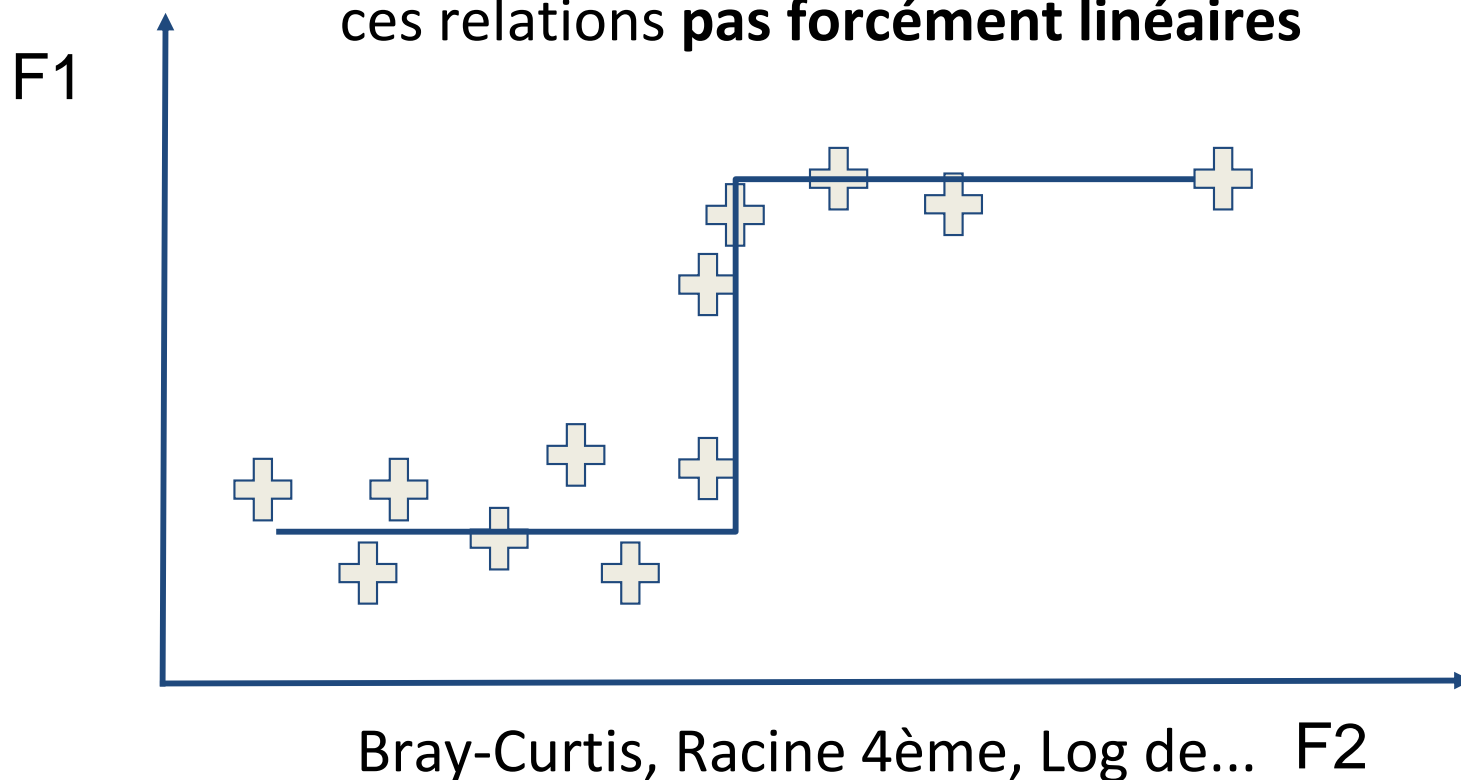
Avec parfois des explications contraires pratiquement aussi probables



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](#)

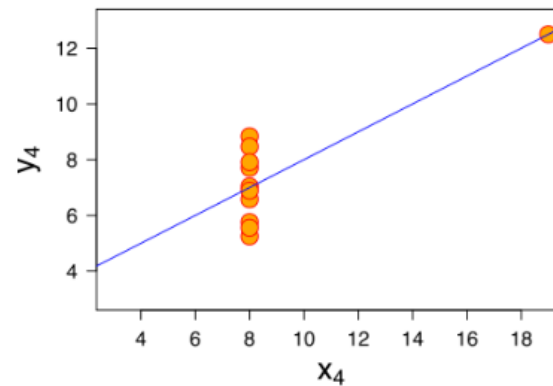
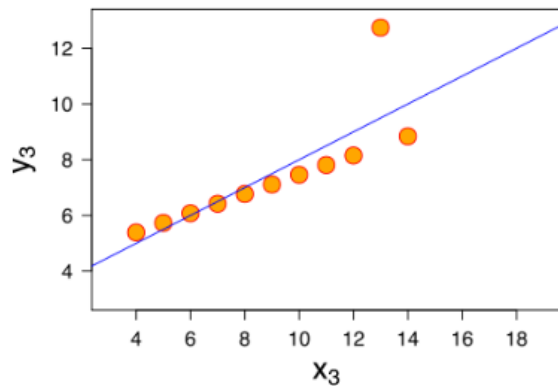
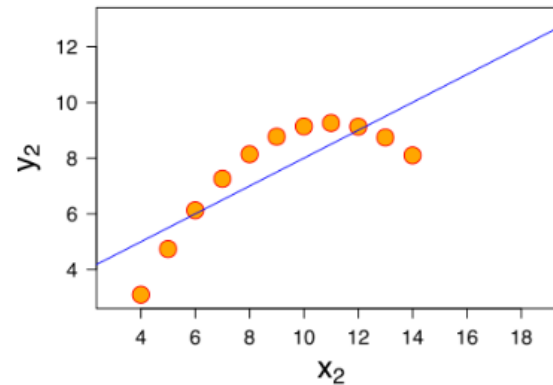
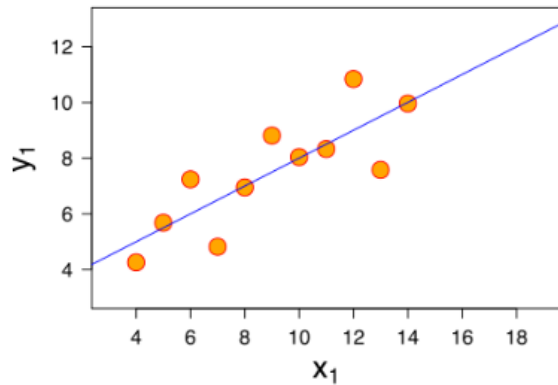
Visualisation et inférences statistiques

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

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 :

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?

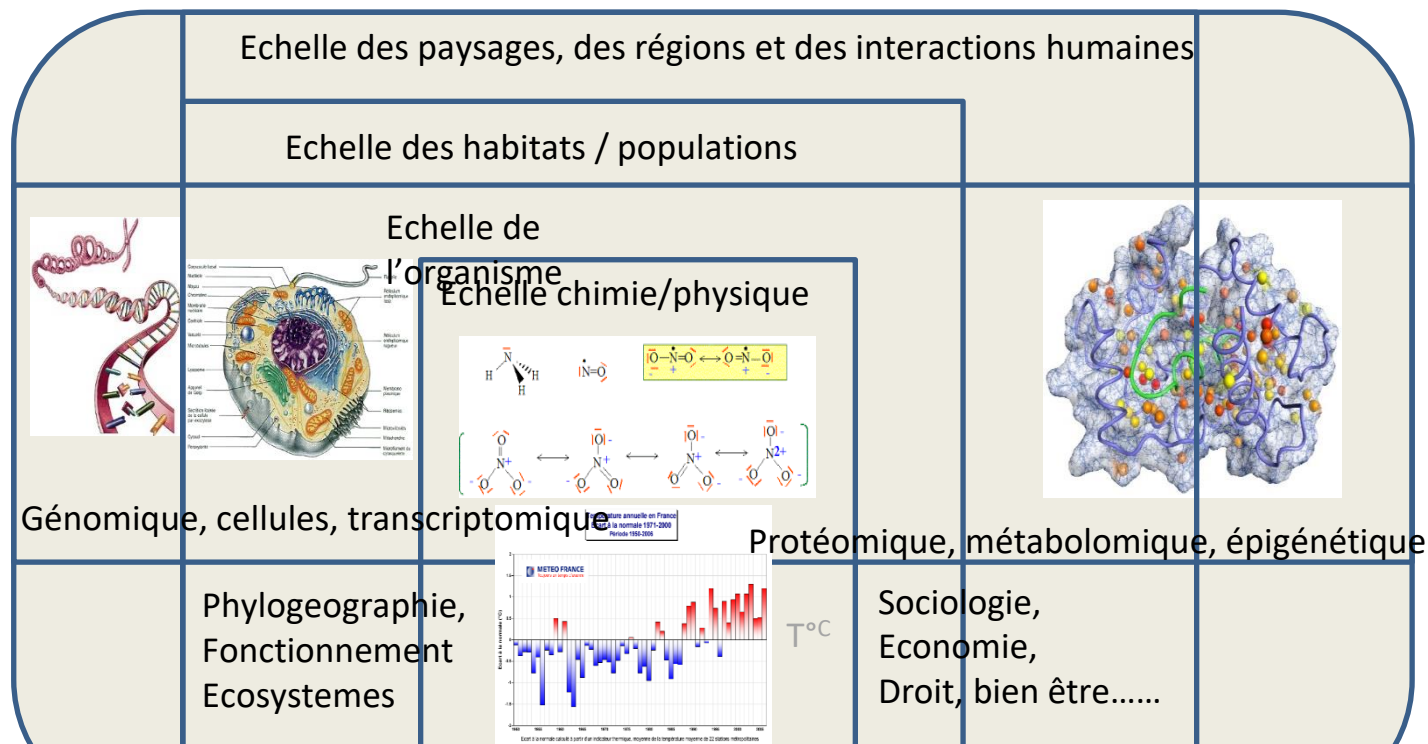
ou F3 a un effet sur F1 et F2

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...

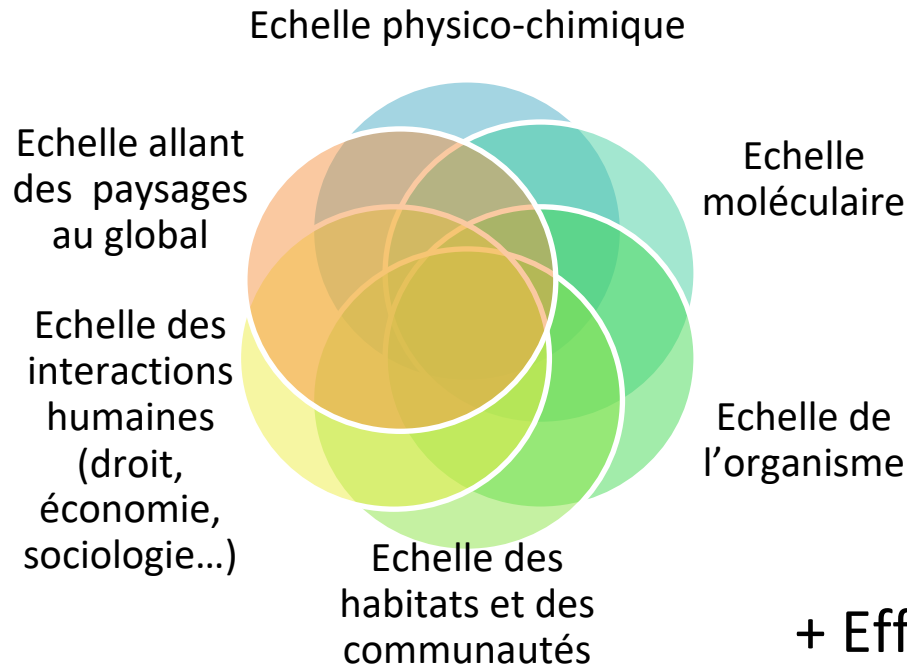


Quelle approche intégrative pour ce patchwork de contextes à chaque niveau d'organisation?

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

Le contexte : des données très « sectorielles »

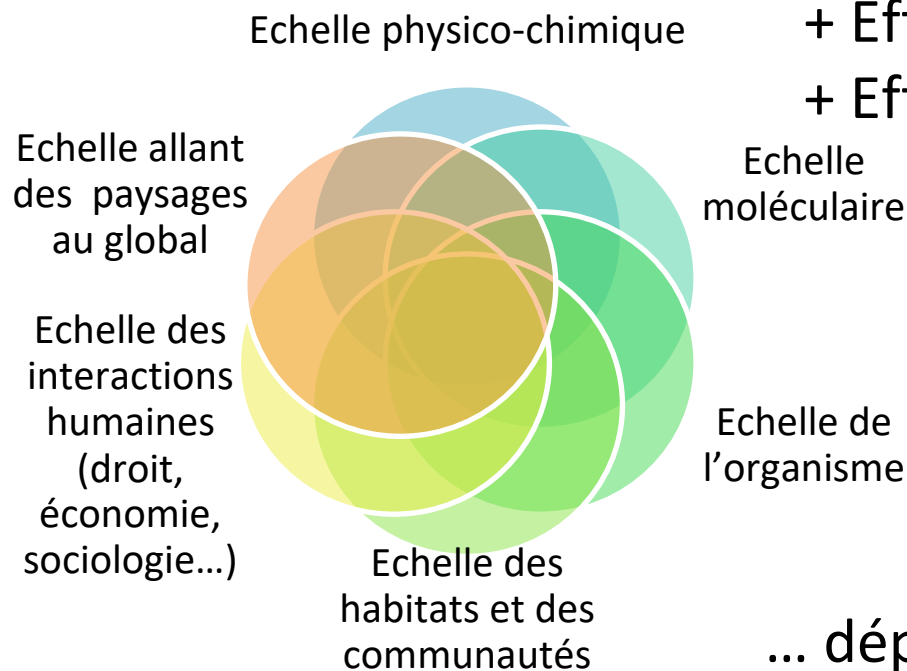
Pourtant, il faut décrire
un système fortement interconnecté et interdépendant



- + 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

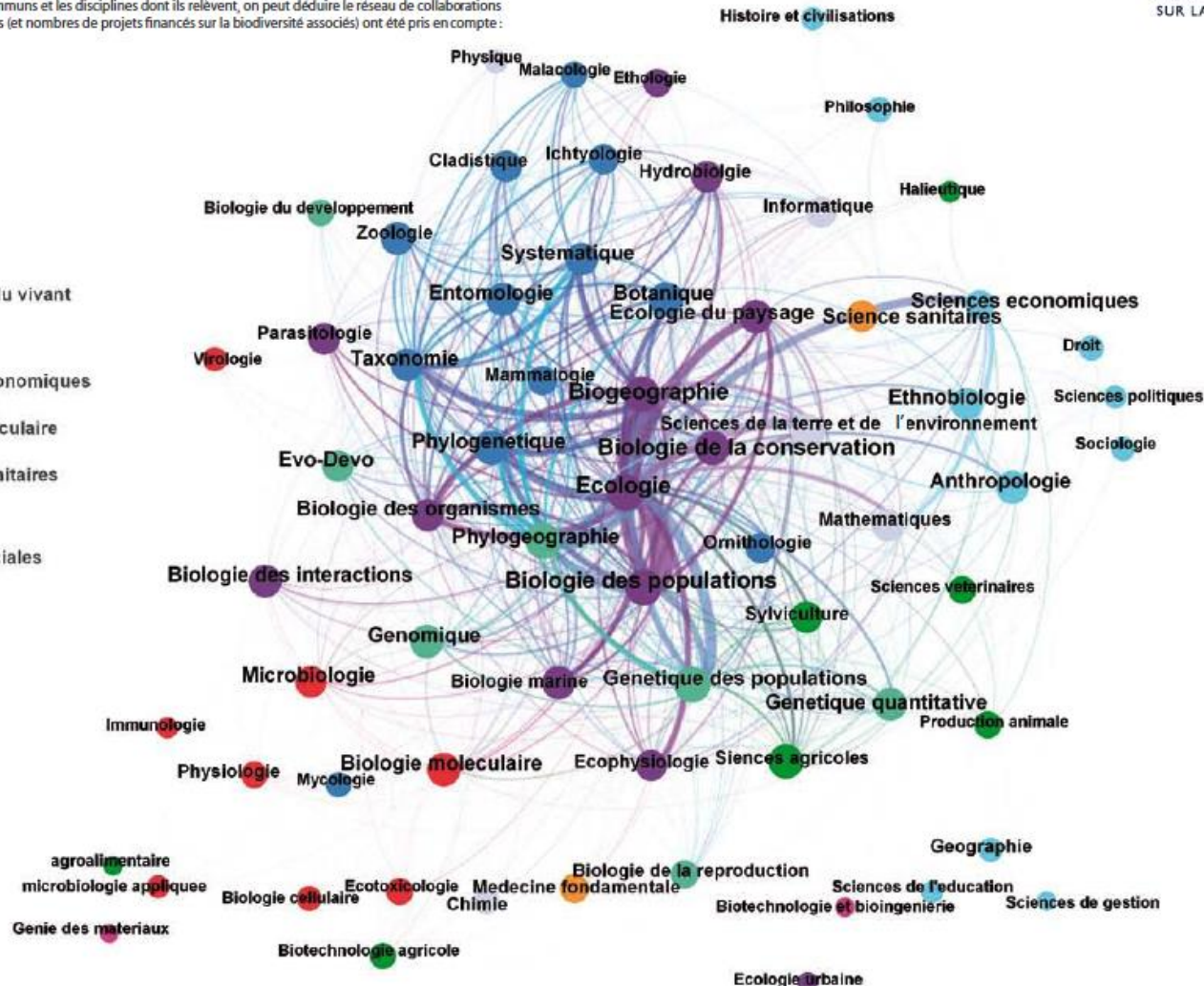
Interdisciplinarité : des disciplines encore très compartimentées

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

Appliqués aux habitats coralligènes et aux récifs artificiels

Questionnements liés aux systèmes d'observation abordés pendant la thèse

Quels protocoles d'observations nécessairement courts (en plongée) sont les plus efficaces (qualités et défauts) ?

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

Questionnements appliqués aux habitats coralligènes et aux ARMS

Questionnements liés aux systèmes d'observation abordés pendant la thèse

Quels protocoles d'observations nécessairement courts (en plongée) sont les plus efficaces (qualités et défauts) ?

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

Comment cerner la variabilité à large échelle liée à des différences de contexte et créer des typologies communes de descripteurs ?

Comment localement identifier la différence entre un changement durable et une variabilité naturelle ?

→ **Comment** mieux définir et décrire les facteurs de contextes à l'origine des changements ?

Questionnements appliqués aux habitats coralligènes et aux ARMS

3) Sélection de résultats sur les systèmes d'observation

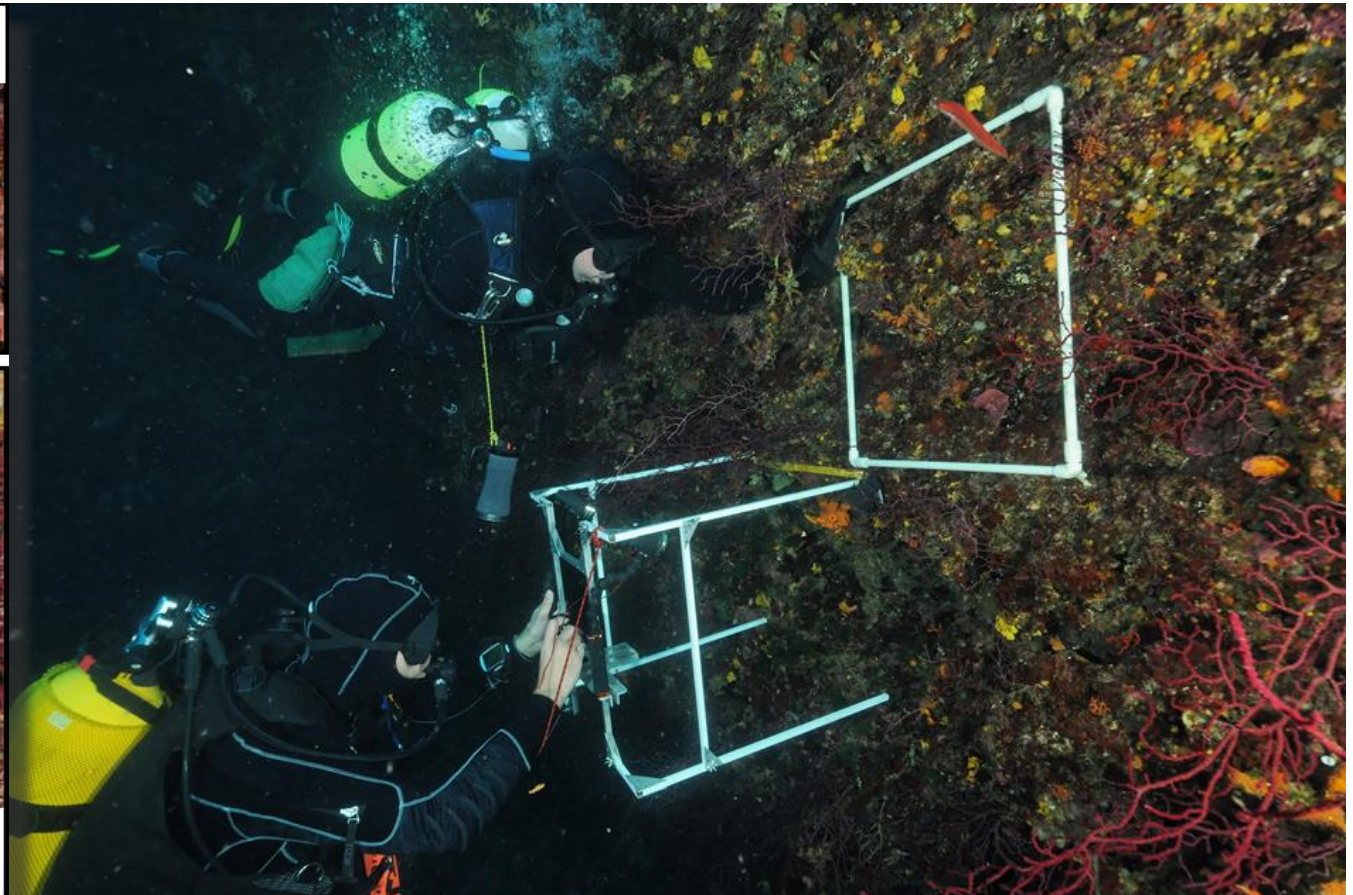
Comment appréhender et contrôler la variabilité liée aux changements d'opérateurs / d'observateurs / à l'évolution de leur expérience?

Rhodophytes calcaires



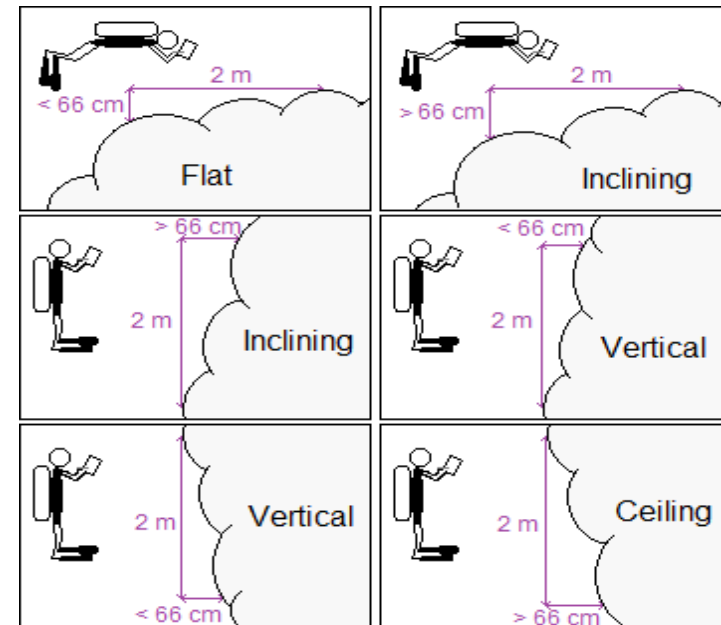
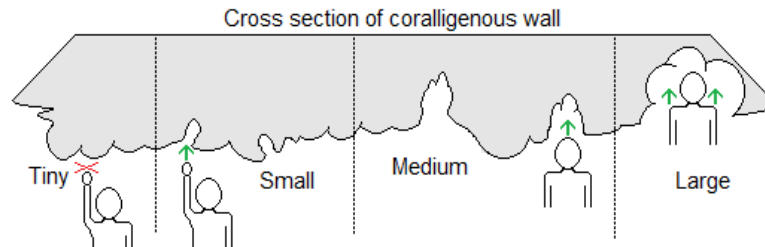
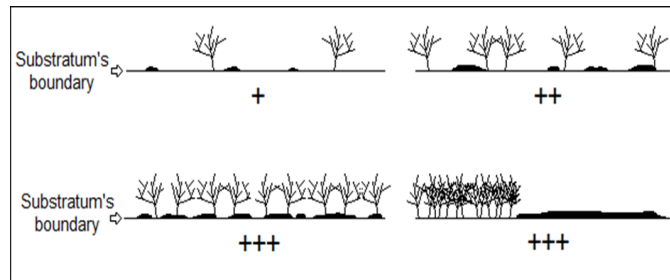
Frédéric Zubéren

Coralligène à *Paramuricea clavata*, sur paroi verticale



3) Sélection de résultats sur les systèmes d'observation

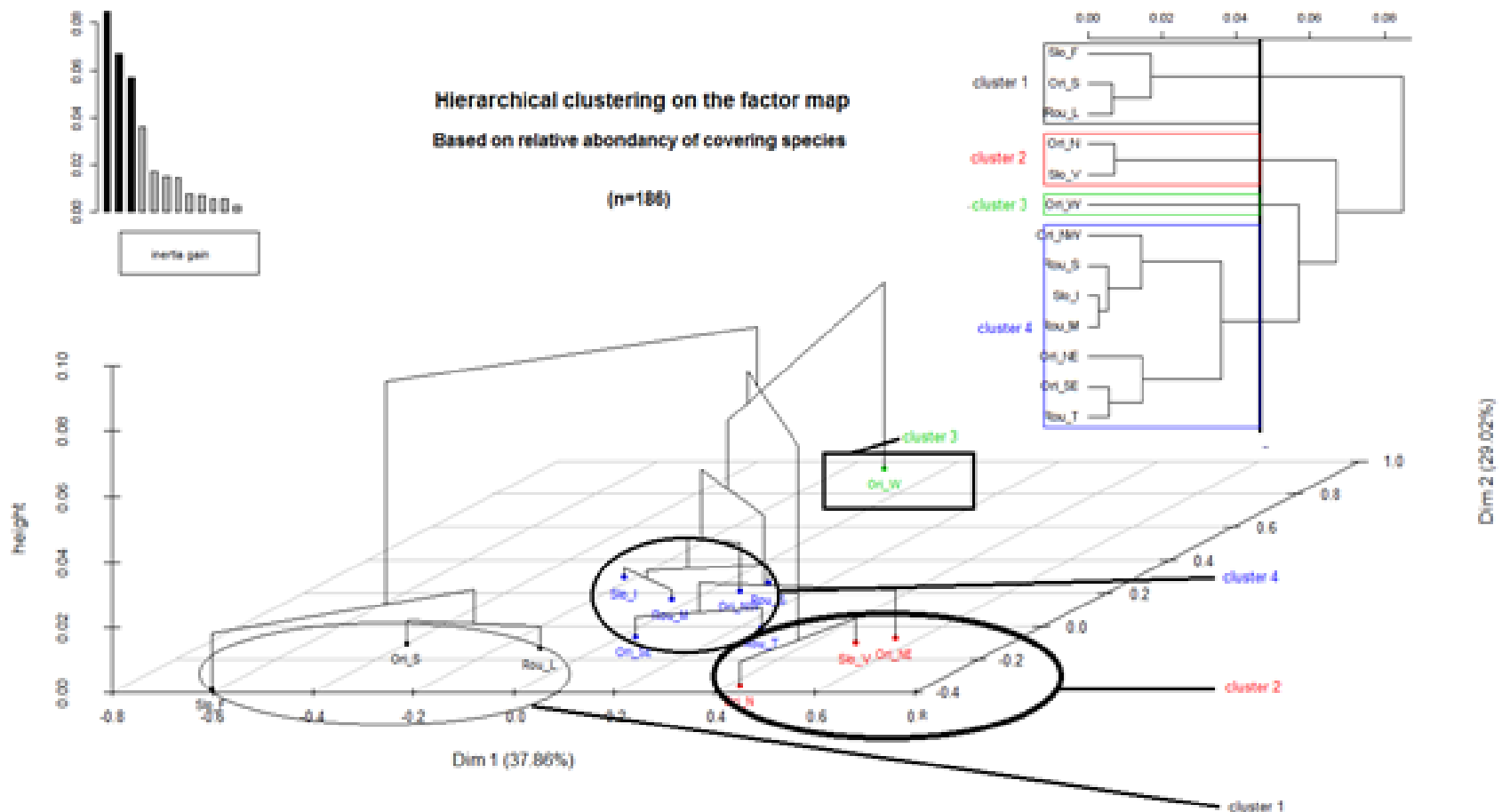
Les assemblages dominants sont localement expliqués par quelques variables très simples (pente, rugosité, orientation, profondeur)



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)

**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

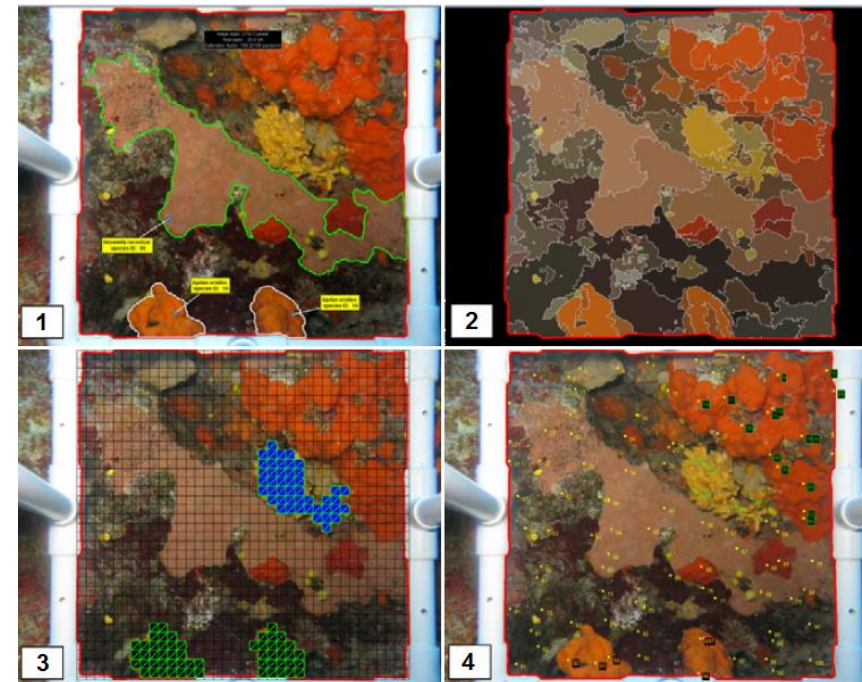
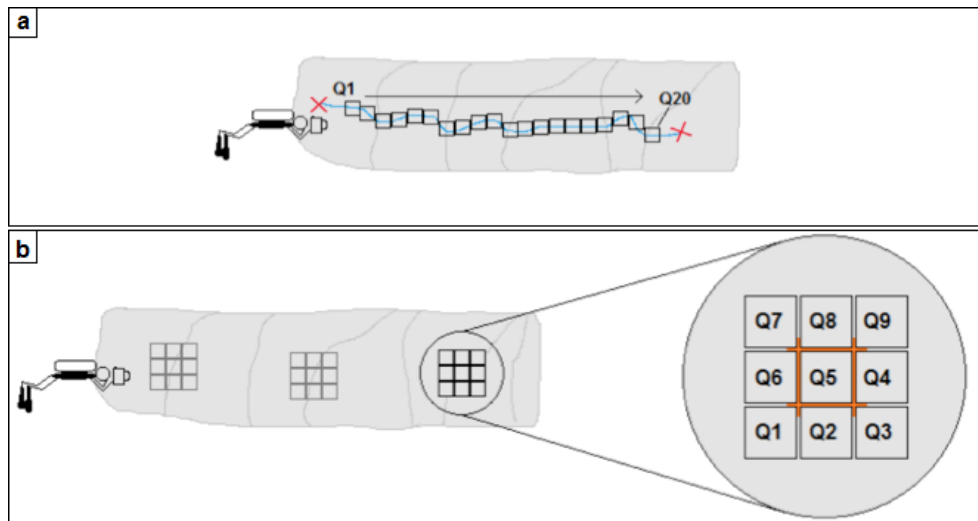
Site: Date:/..../..... Depth: m Observer:

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	Orientation																	
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?

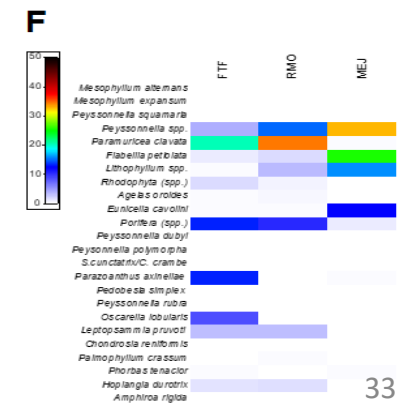
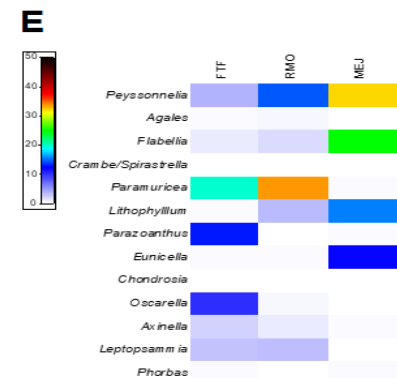
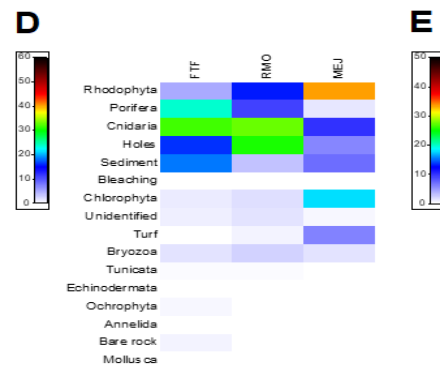
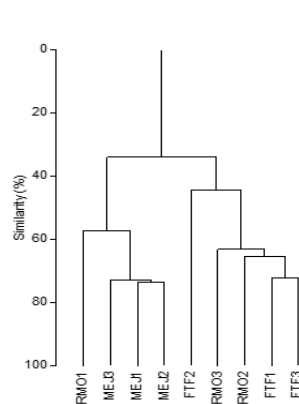
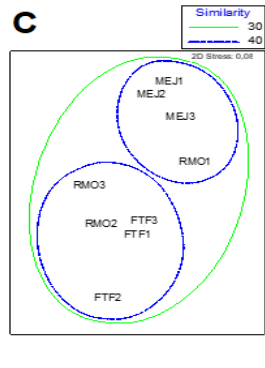
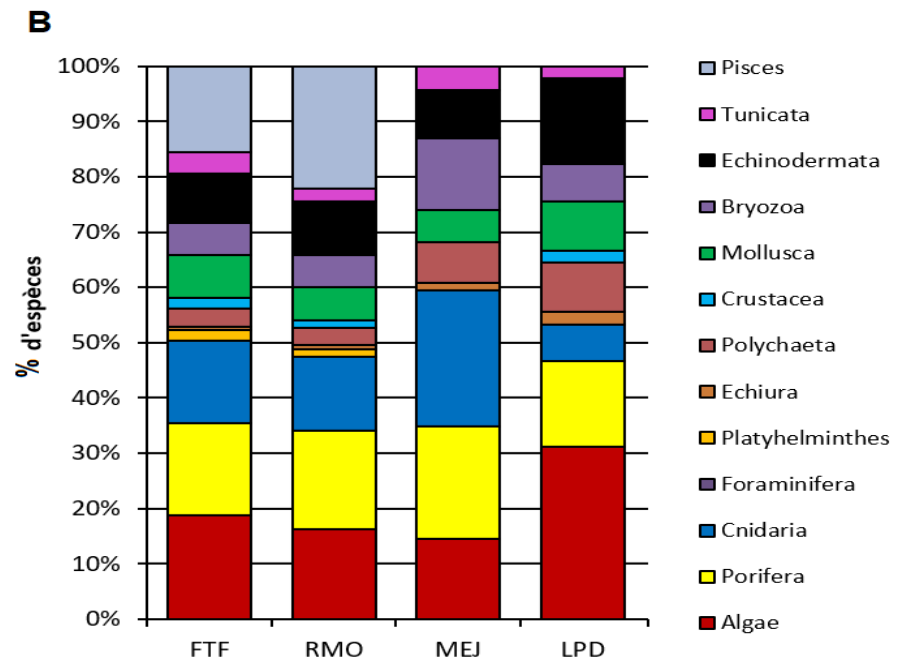
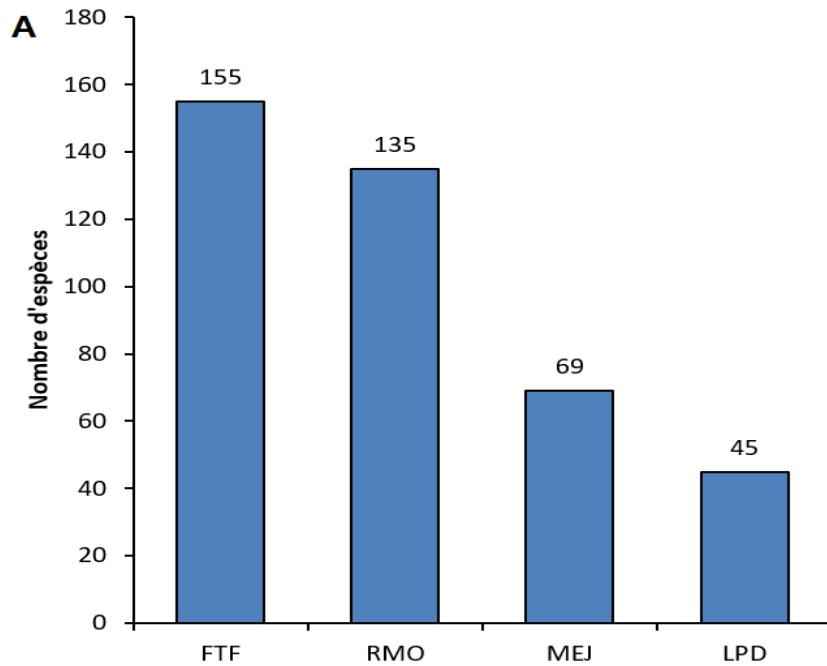
Sur la méthode d'échantillonnage



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

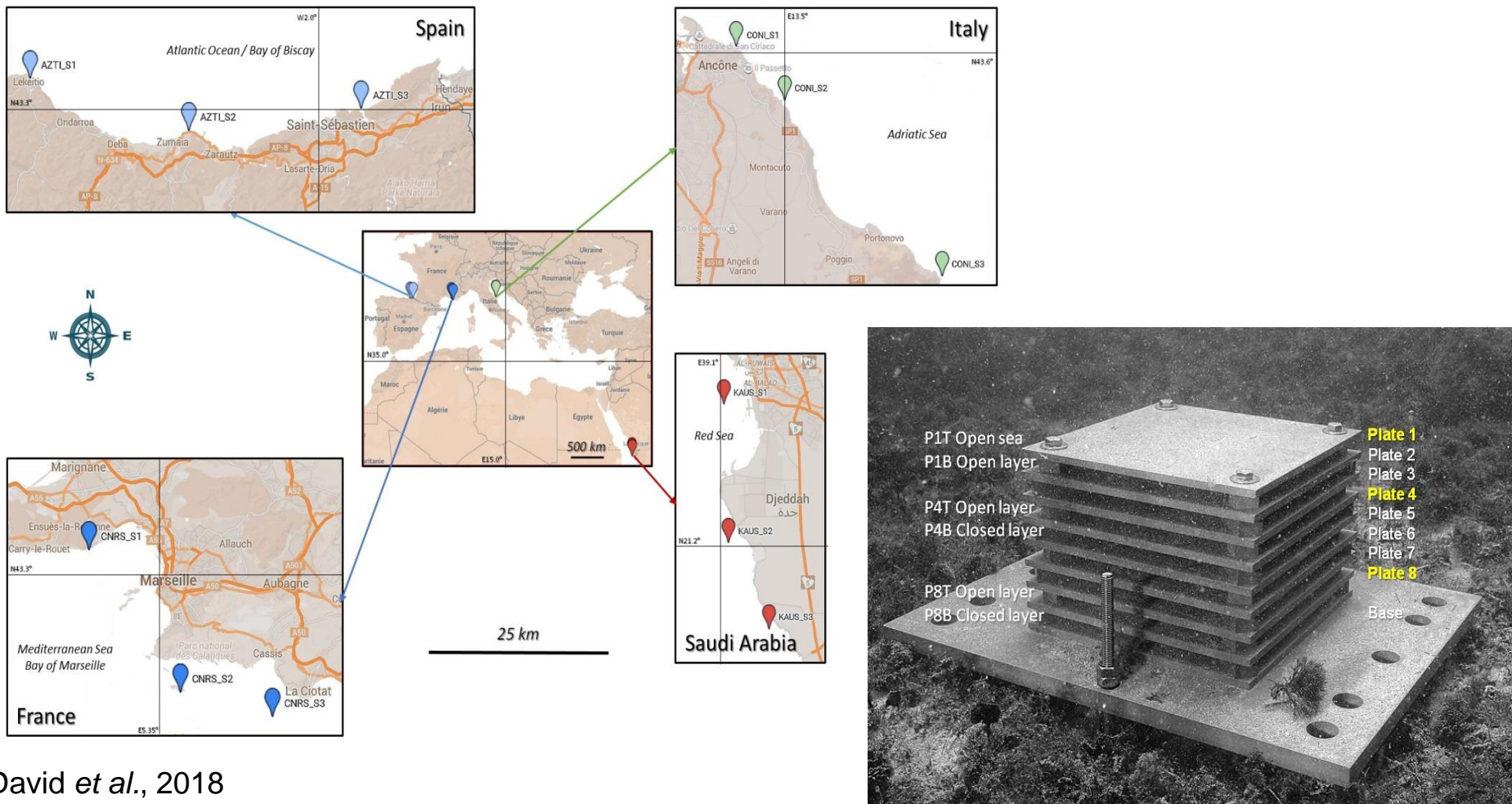
3) Sélection de résultats sur les systèmes d'observation

Résultats des analyses de quadrats CIGESMED autour de Marseille



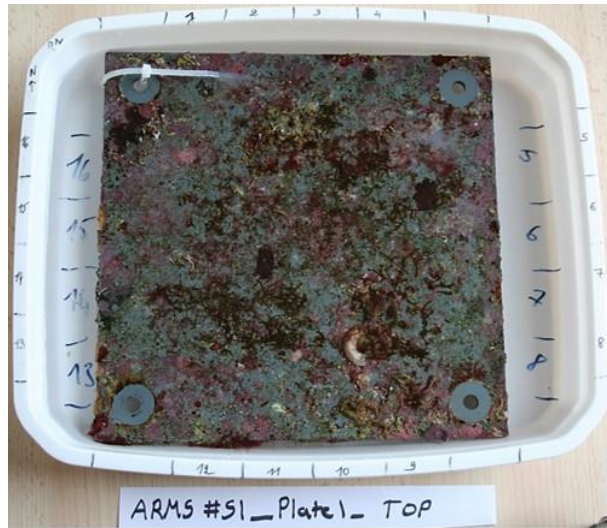
3) Sélection de résultats sur les systèmes d'observation

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

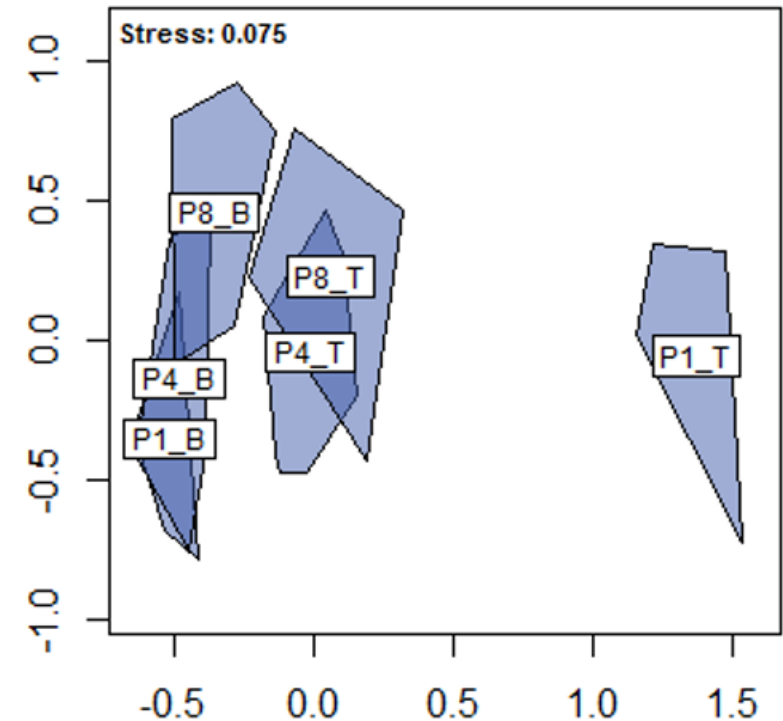
Comment cerner la variabilité à large échelle liée à des différences fondamentales / créer des typologies communes ?



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 ?

Plate face	Sea region (P-value)	Site (within Sea) (P-value)	
P1T (exposed)	$5. 10^{-4}$	$1. 10^{-4}$	
P4T	$1. 10^{-4}$	$43. 10^{-4}$	
P8T	$1. 10^{-4}$	$749. 10^{-4}$ (NS)	
P1B	$3. 10^{-4}$	$6. 10^{-4}$	
P4B (comp.)	$3. 10^{-4}$	$1. 10^{-4}$	
P8B (comp.)	$1. 10^{-4}$	$25. 10^{-4}$	



3) Quelques conclusions fondamentales sur les systèmes d'observation

- **L'inter-calibration et l'entraînement réguliers** sont incontournables
- La **formation de débutant** donne rapidement des résultats aussi performants que dans le cas d'observateurs plus expérimentés (sauf dans le cas de phylums plus complexes comme celui des éponges)
- Permettre une analyse comparative de données entre mers régionales nécessite de considérer **des niveaux taxonomiques supra spécifiques** (Ils restent pertinents pour différencier les sites).
- La **description de facteurs de contextes et les “micro habitats”** qui en dépendent est essentielle pour expliquer les différences de structure des principaux assemblages d'espèces

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) Concepts et questionnements scientifiques abordés pendant la thèse 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) Concepts et 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 ?

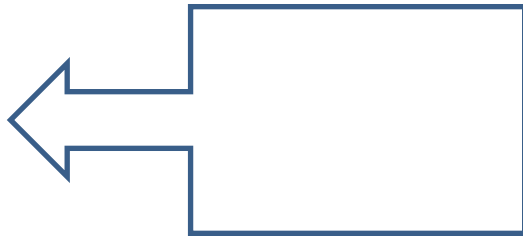
Quelle qualification des données a le meilleur impact sur le niveau d'usage (et donc sur la valeur des données) ?

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 ?



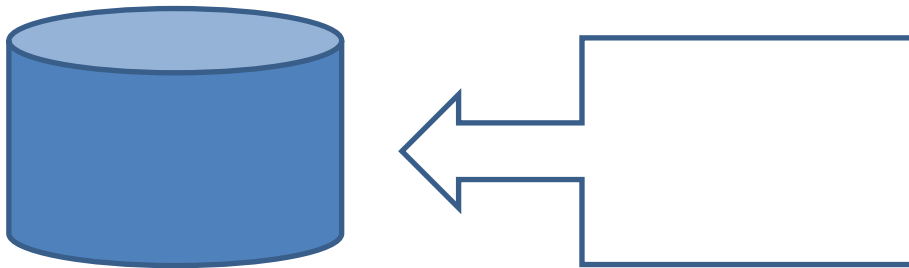
= un descripteur de données

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

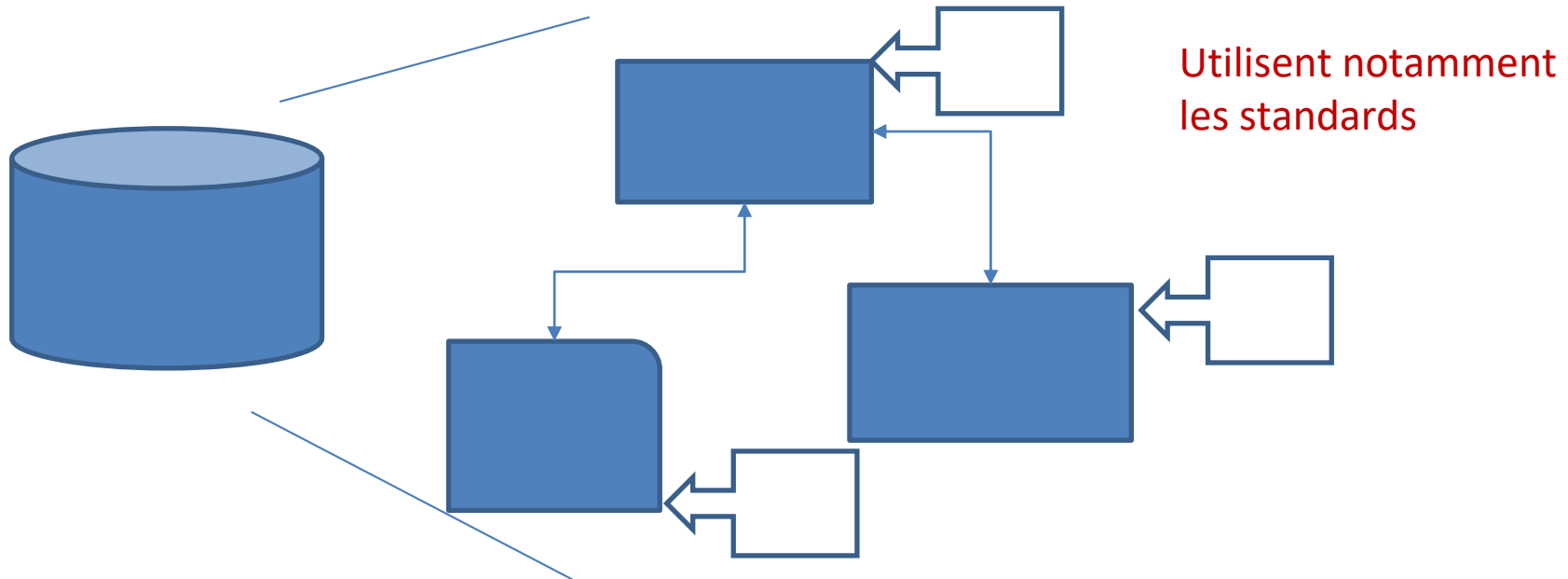


Notamment normes (ISO, INSPIRE...)

3) Sélection de résultats sur les systèmes d'information

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

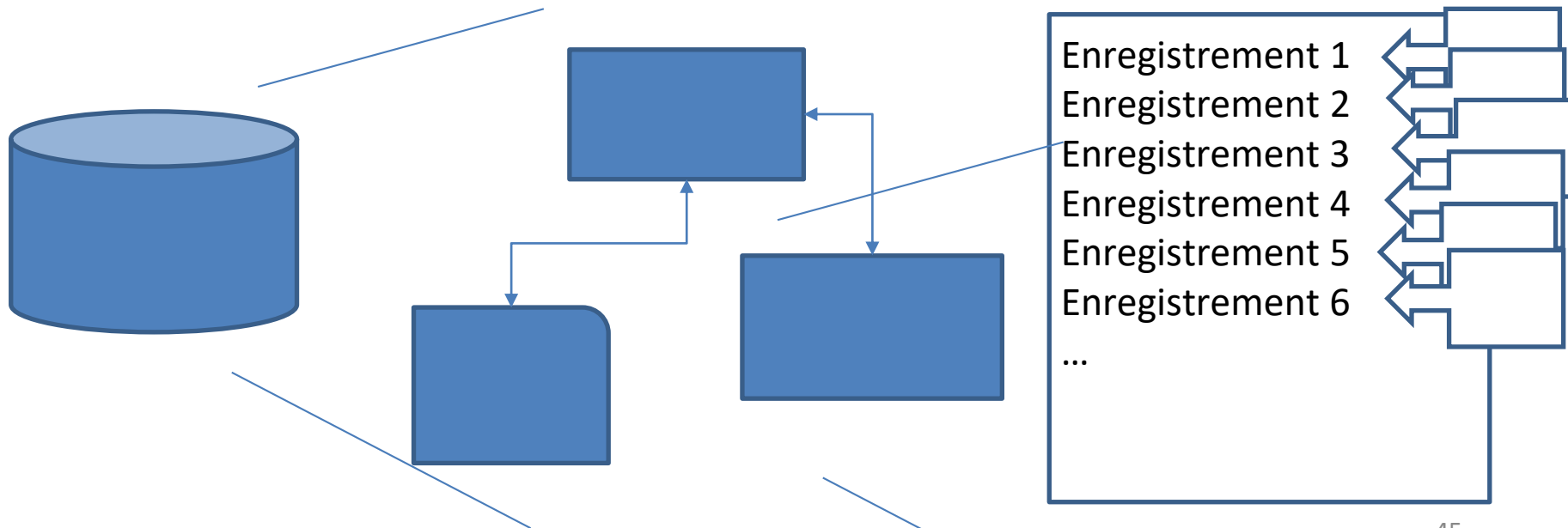
Niveau Tables: dictionnaire de données, mais pas uniquement (équivalences via thesaurus, règles d'intégrité)



3) Sélection de résultats sur les systèmes d'information

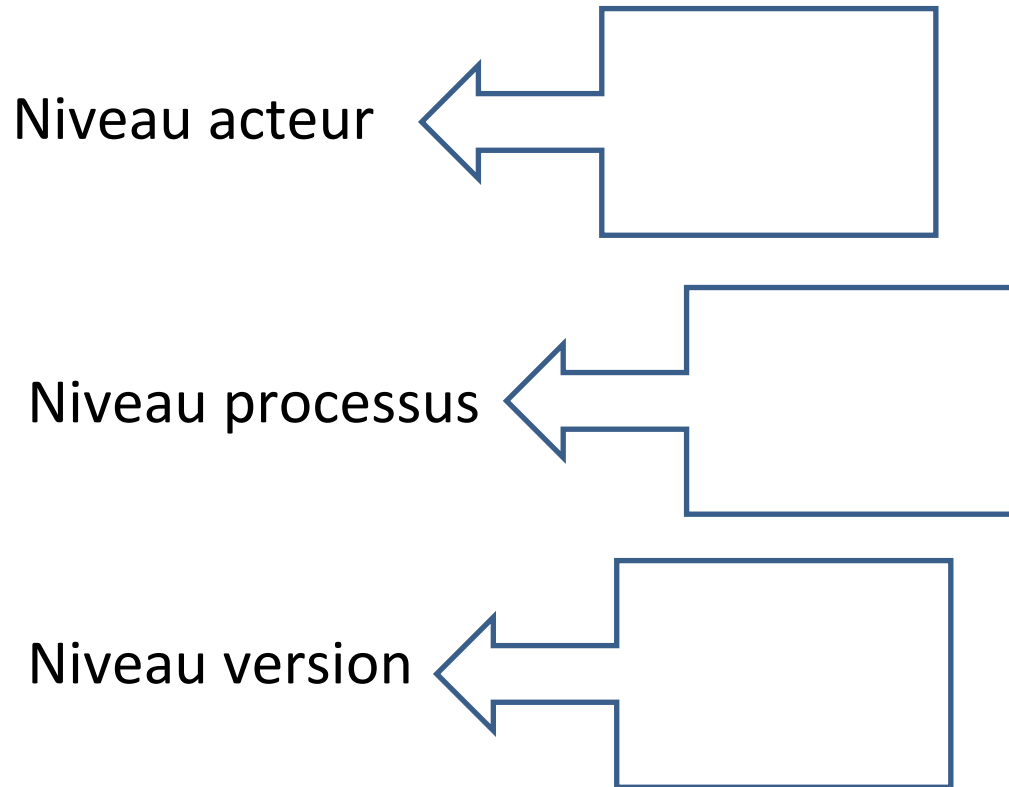
Comment qualifier / requalifier des données hétérogènes et réparties et permettre leur interopérabilité ?

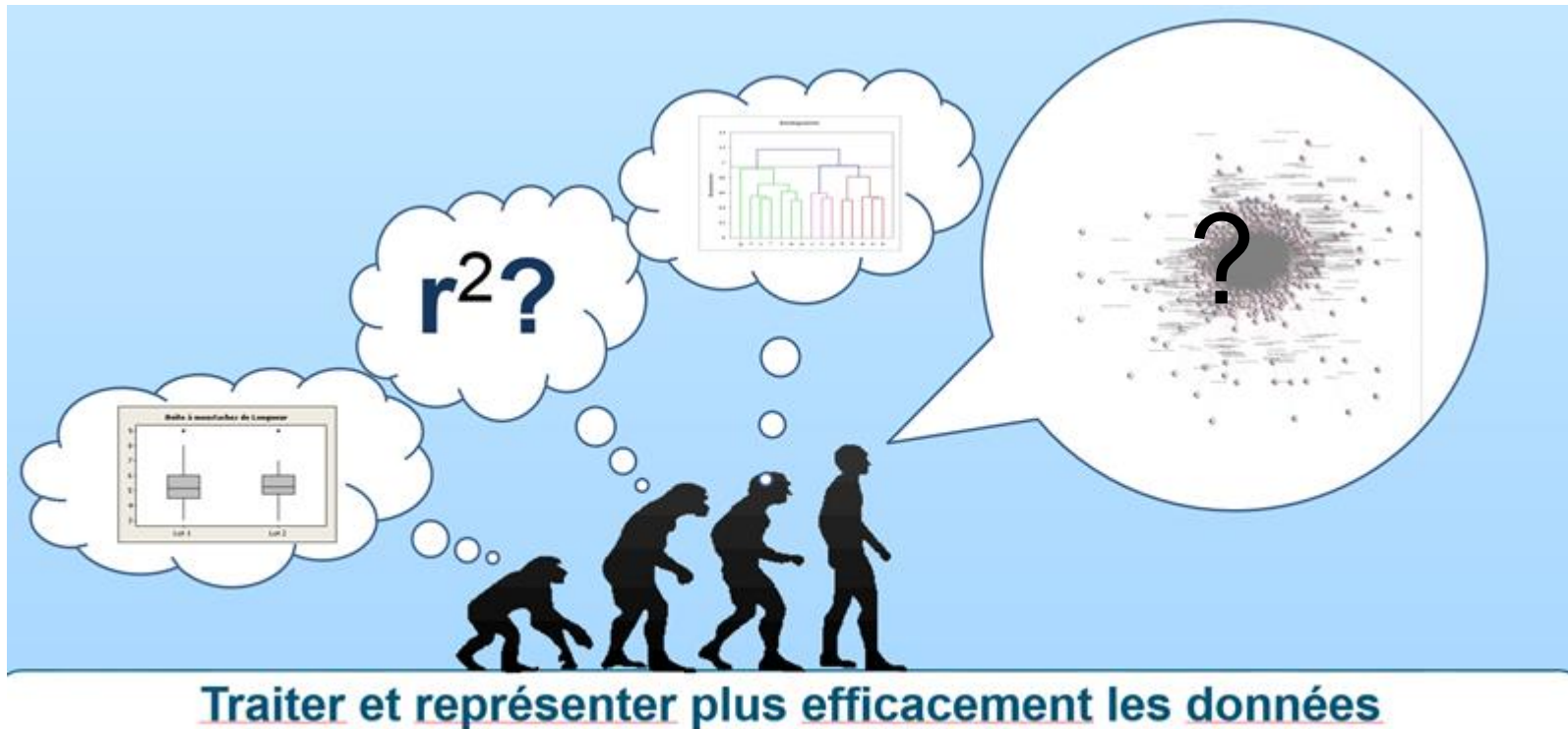
Niveau Enregistrement



3) Sélection de résultats sur les systèmes d'information

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





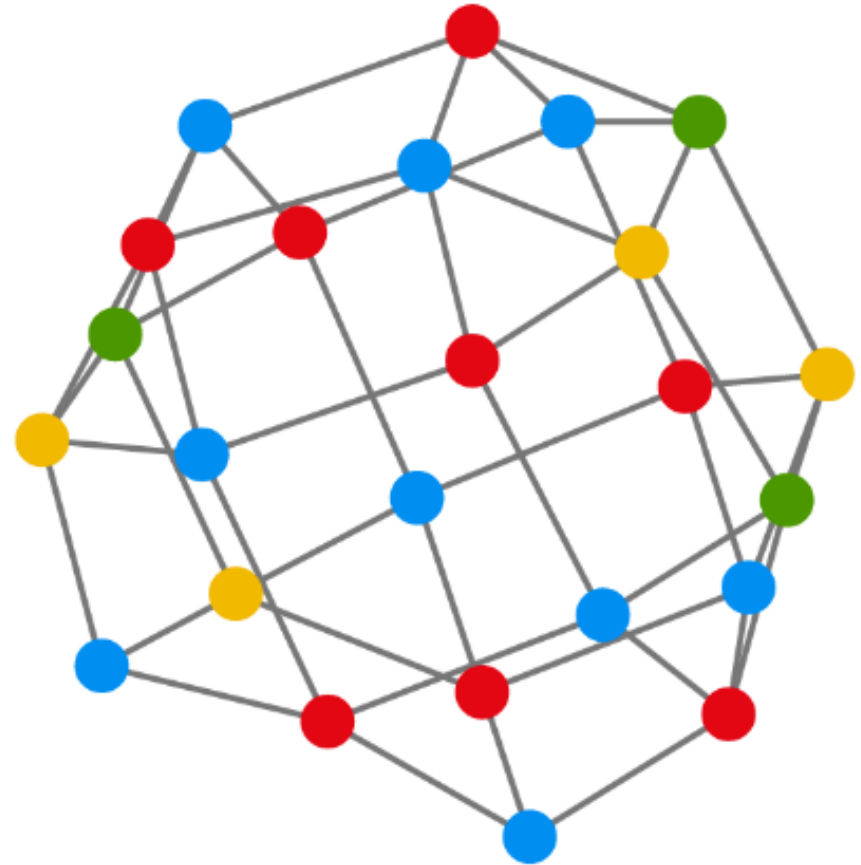
Proposition : utiliser les graphes pour explorer les données distantes et hétérogènes

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

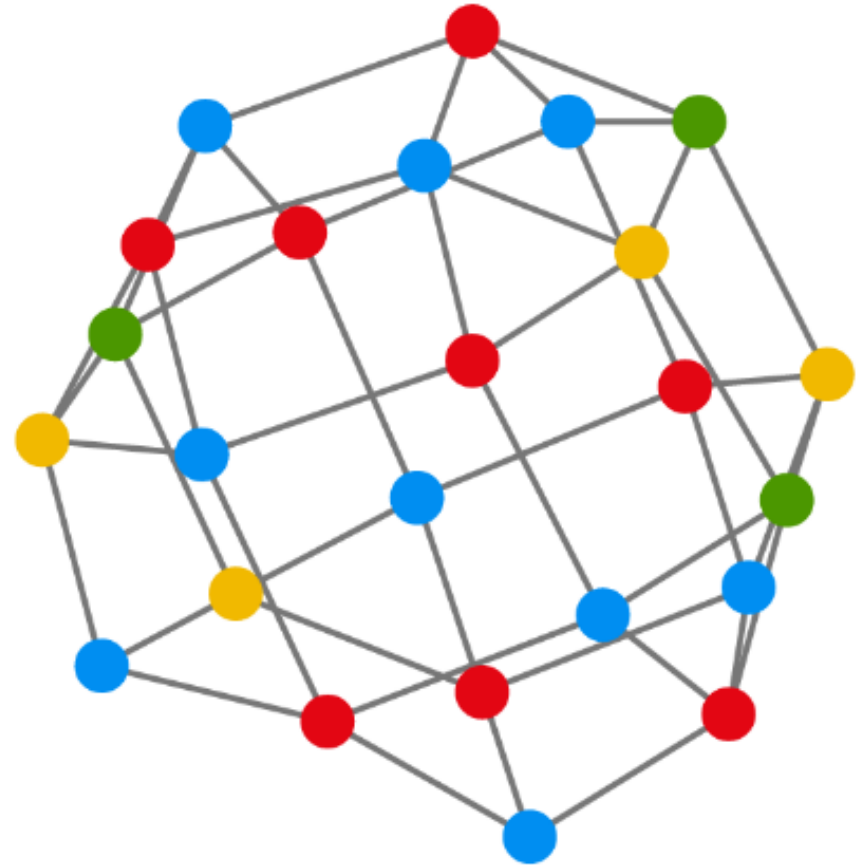
Quadrats photo

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

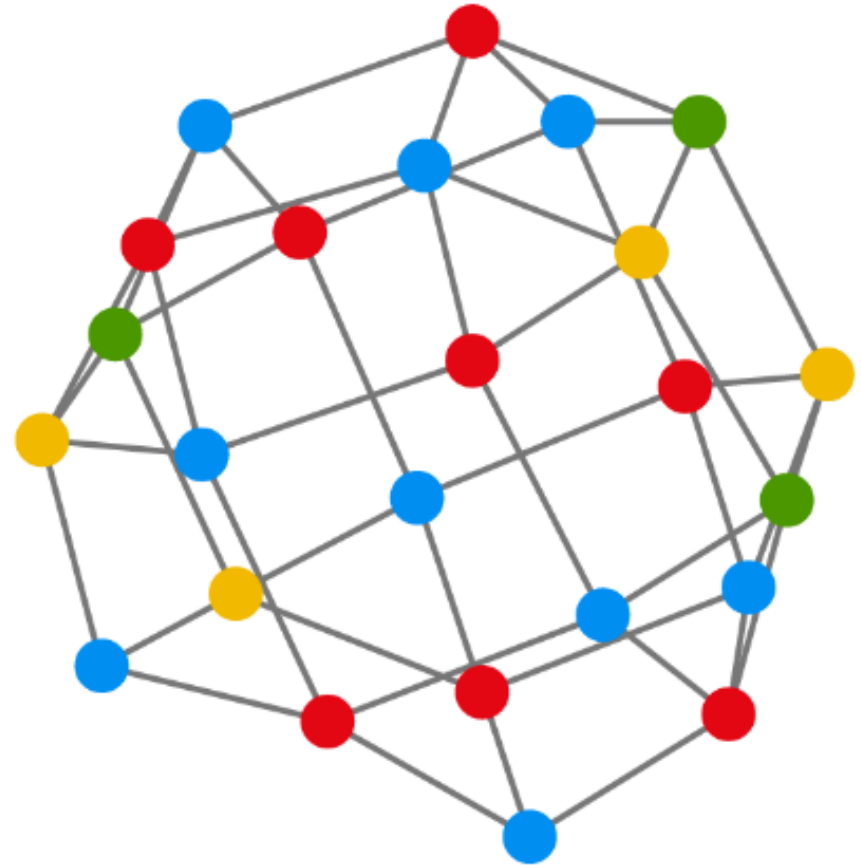
Echantillons

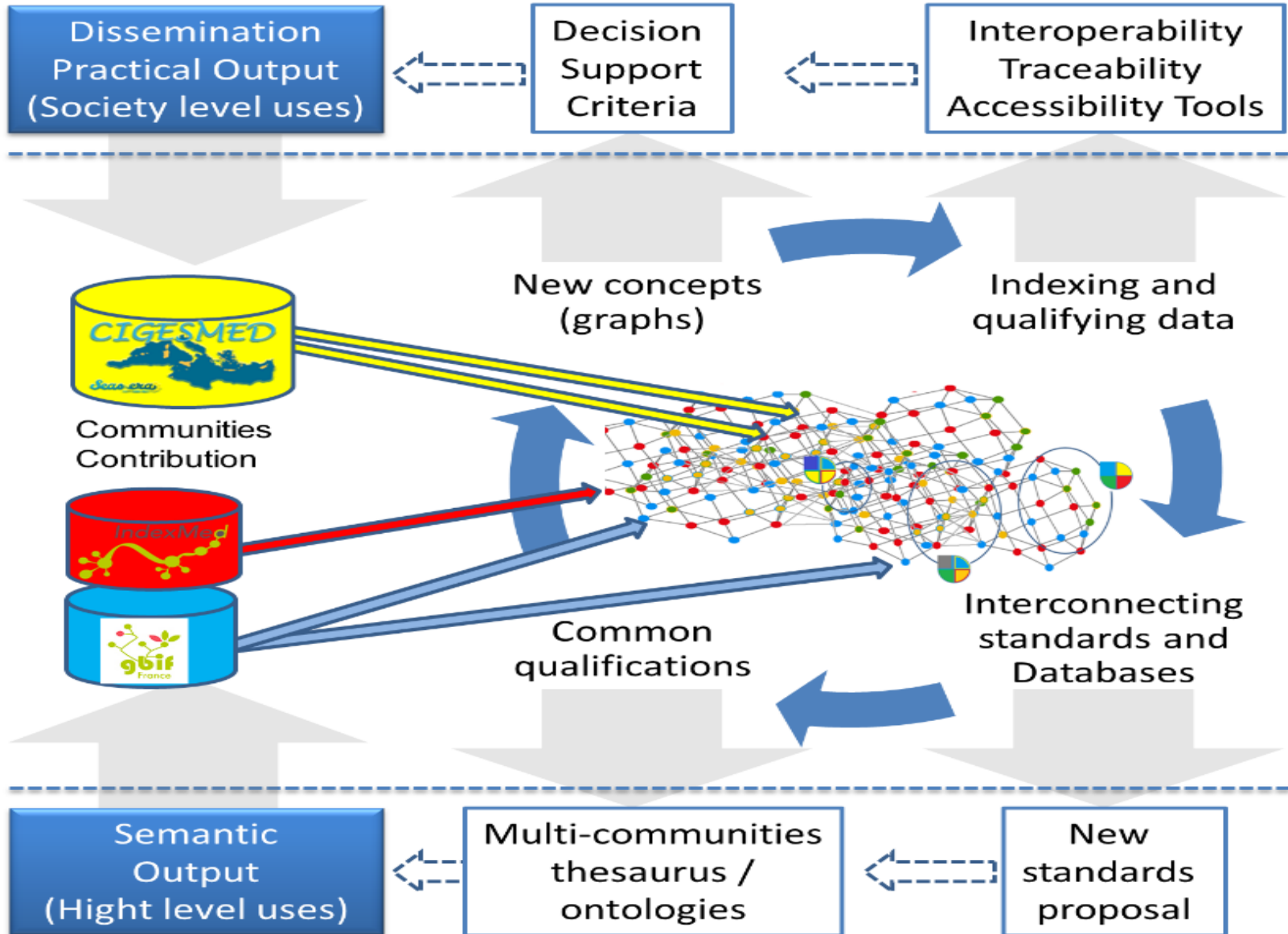
Objets "moléculaires"

Photos

Groupes citoyens

...





3) Sélection de résultats sur les systèmes d'information

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

Le prototype

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 blue node labeled 'Sp 1' and several pink nodes labeled 'SS17'. A green node labeled 'Marseille' is highlighted with a yellow circle and a red arrow. The interface includes a toolbar with icons for menu, link, upload, play, share, code, database, cloud, and URL. Below the toolbar, there are two dropdown menus for 'TraitementF' and a list of Cypher commands. A text box at the top right displays metadata for a specific photo observation. At the bottom, there are filters for 'Lieu', 'EspècesAssignees', and 'TraitementPhotos', along with a list of properties like 'SE_TROUVE DANS CETTE PHOTO' and 'A_ETE PRISE A'.

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

3) Sélection de résultats sur les systèmes d'information

Une proposition innovante : l'utilisation des graphes

```
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
```

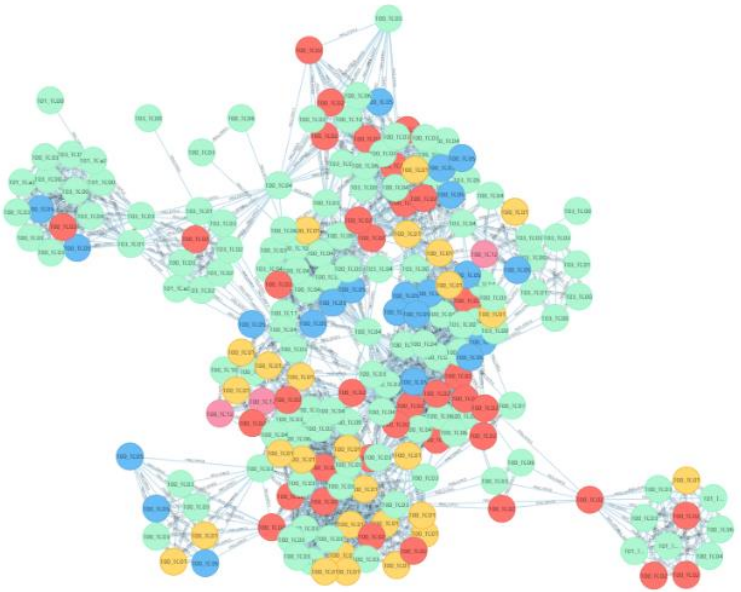
```
MATCH ()-[r:FREQ_ESPECE]->() WHERE
toInt(r.target1)>50 RETURN r ORDER BY
toInt(r.target1) LIMIT 300
```

```
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
```

Label : 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



Nodes properties Attributes properties

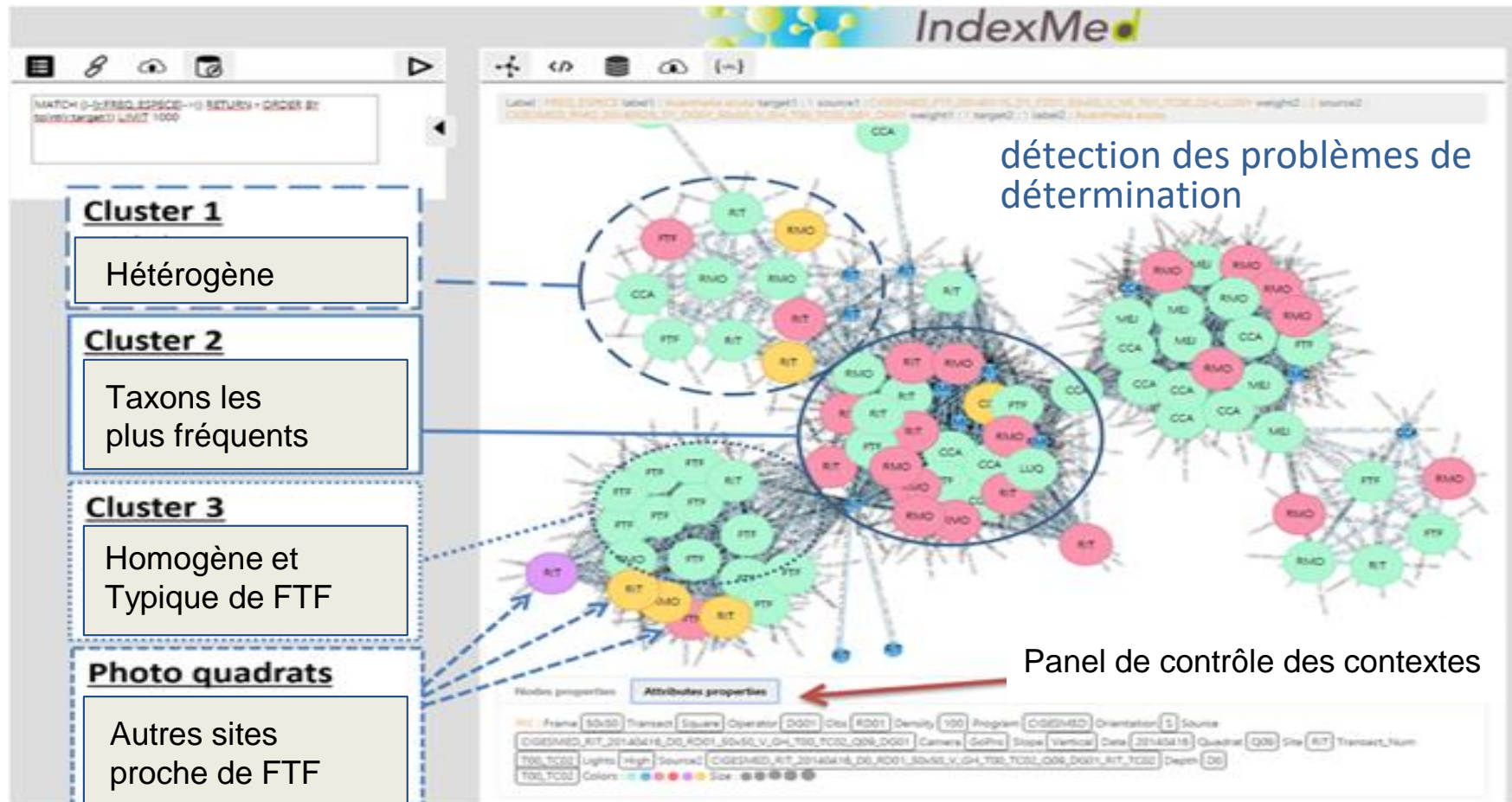
PIC : Frame Transect Operator Obs Density Program Orientation Source Camera Slope Date Quadrat Site Transect_Num Lights Source2 Depth

Colors : Size :

Journal de
requêtes

3) Discussion : la fouille de données

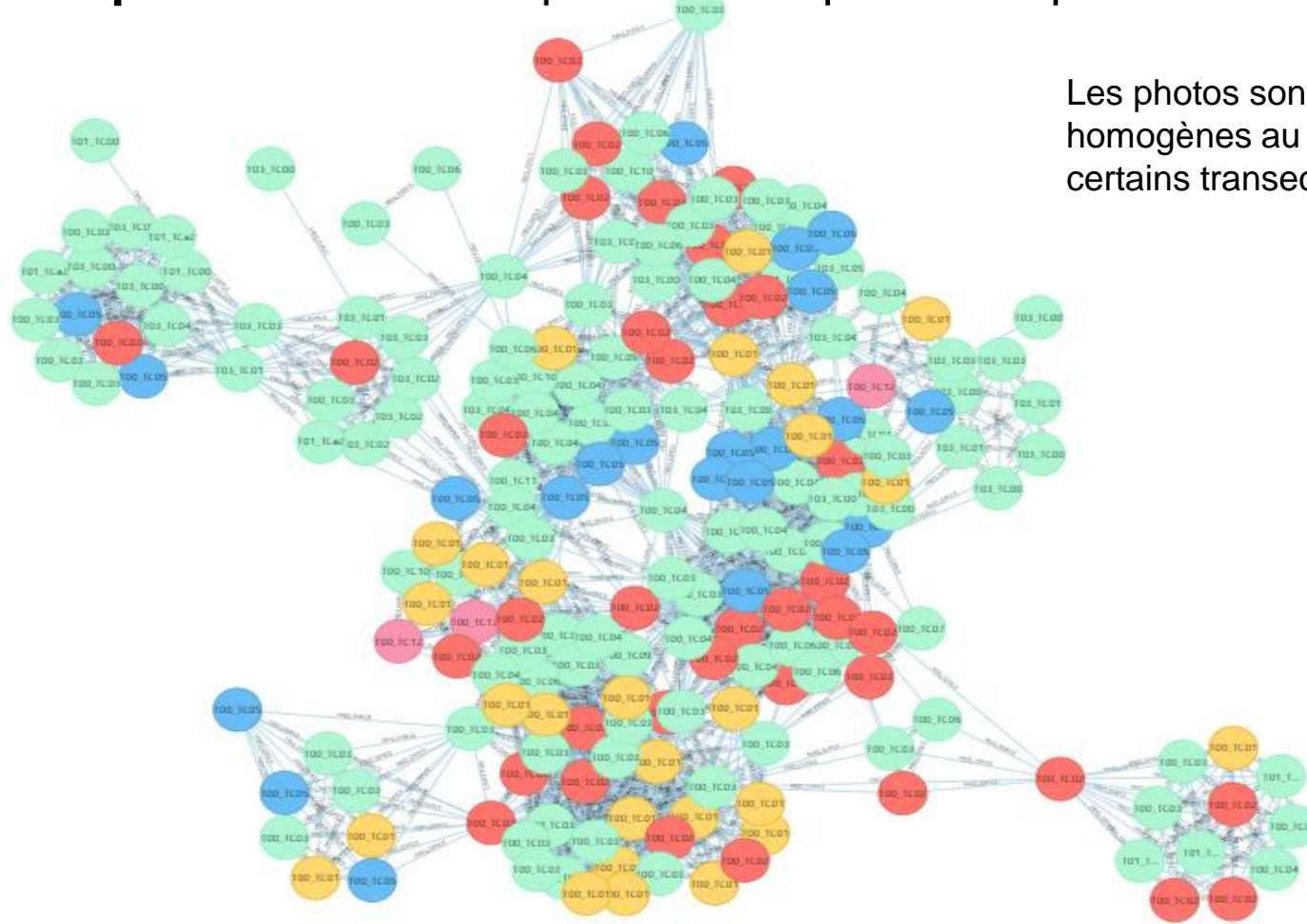
La fouille de données basée sur les graphes, utilisables pour étudier les Compétences des opérateurs :



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

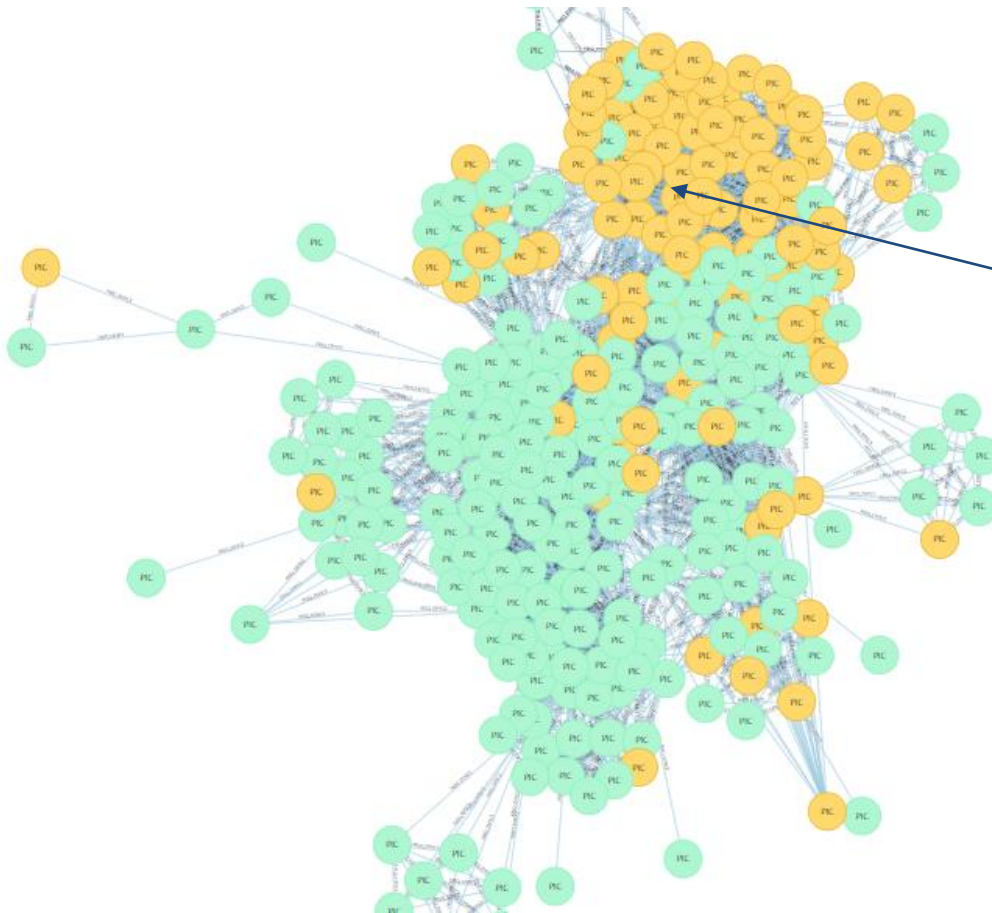


Les photos sont elles
homogènes au sein de
certains transects ?

3) Discussion : la fouille de données

les noeuds : des photos

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



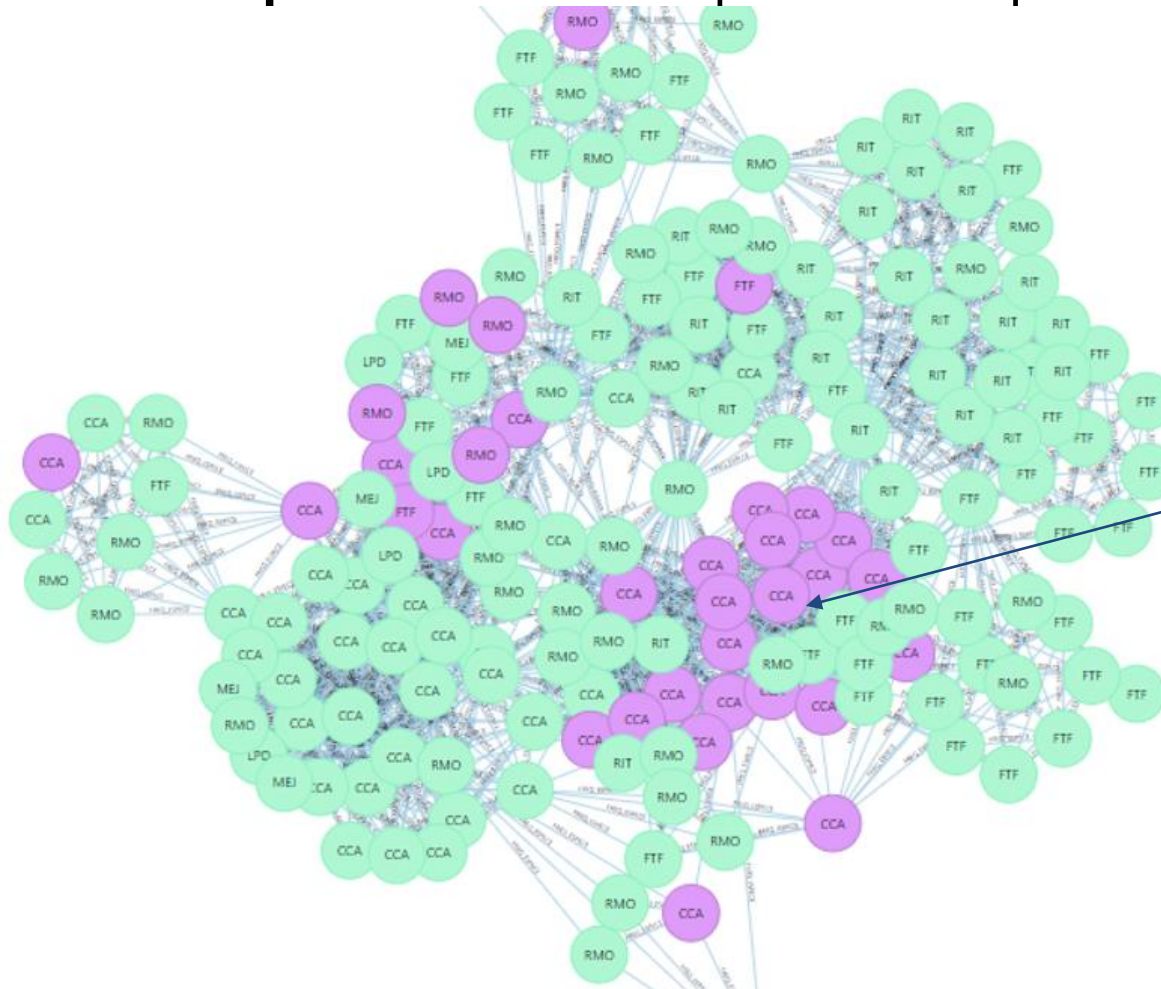
L'inclinaison est elle un facteur explicatif de la structure des assemblages d'espèces ?

Pente Intermédiaire

3) Discussion : la fouille de données

les noeuds : des photos

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



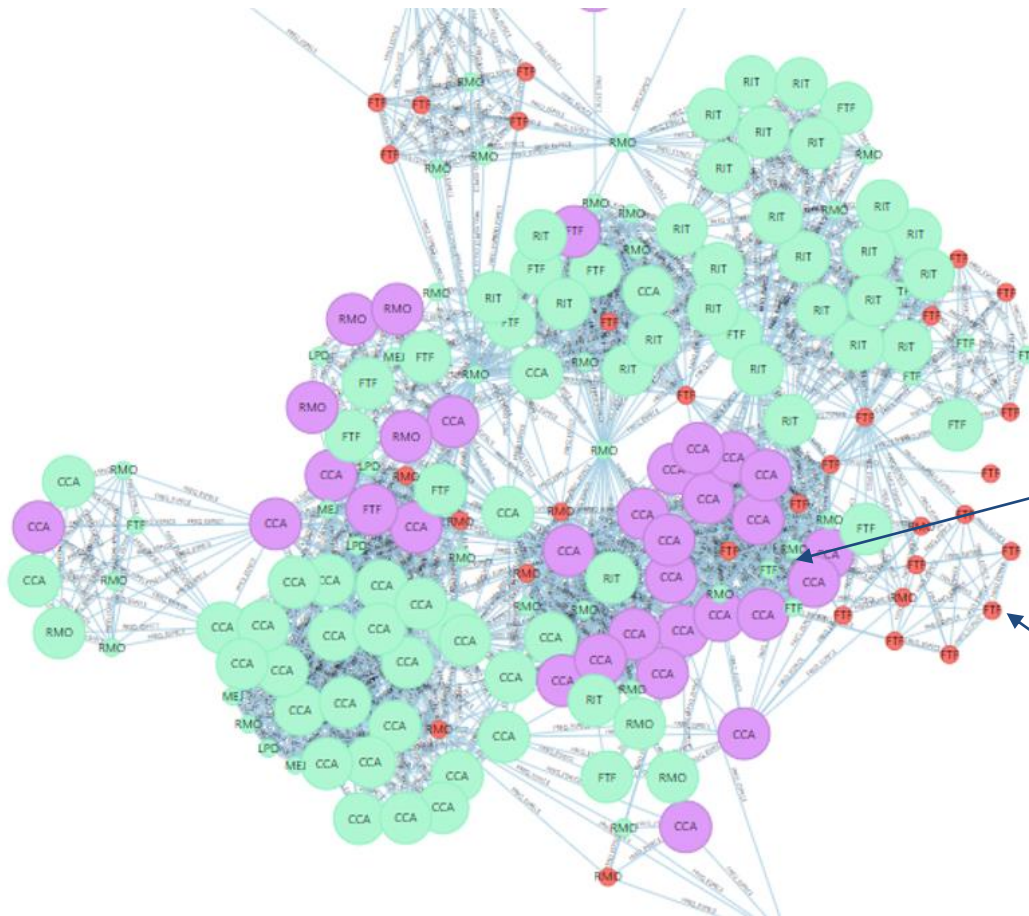
L'orientation est-elle un facteur possiblement explicatif de la structure des assemblages d'espèces ?

Orientation Est

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 ?

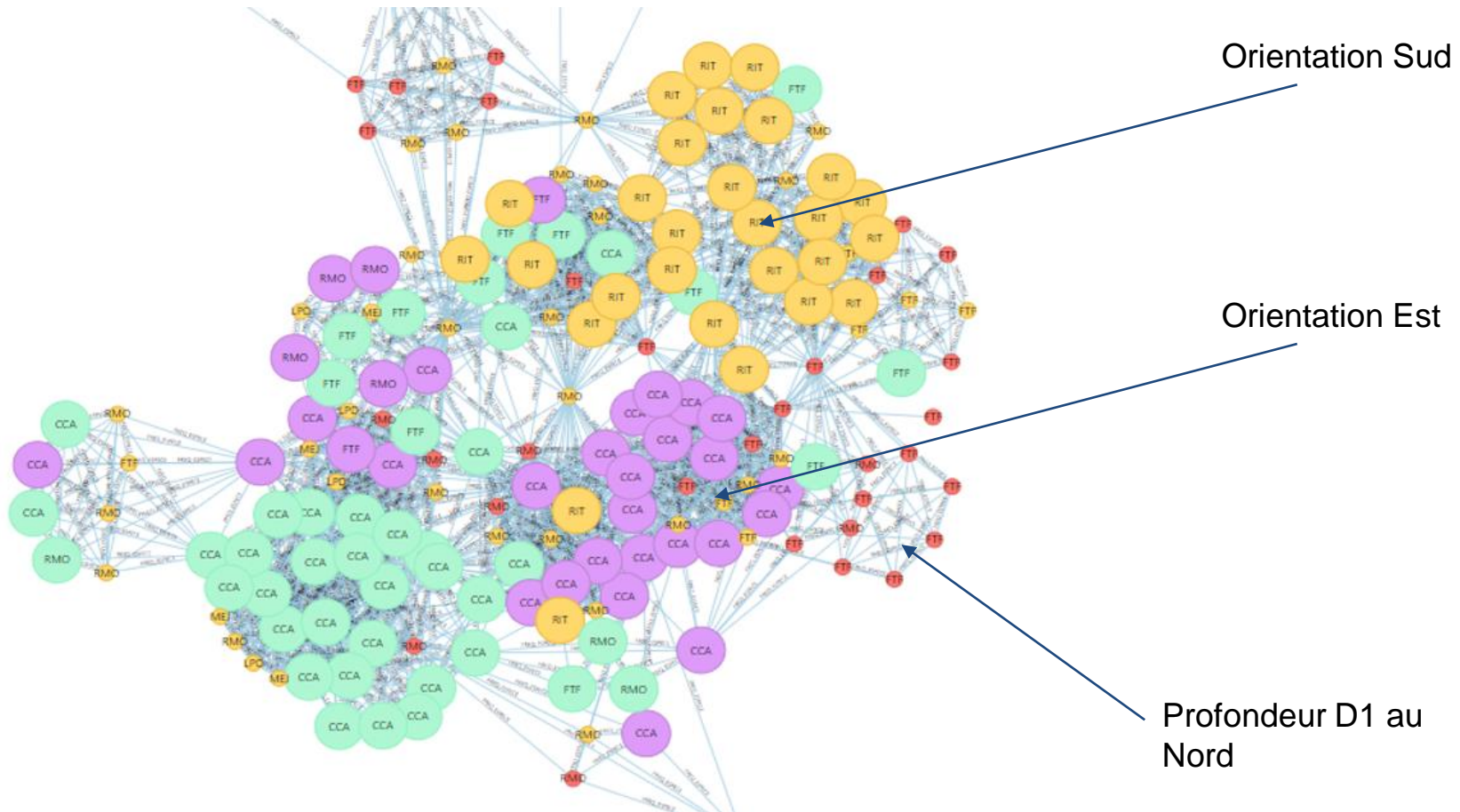
Orientation Est

Profondeur D1 au Nord

3) Discussion : la fouille de données

les noeuds : des photos

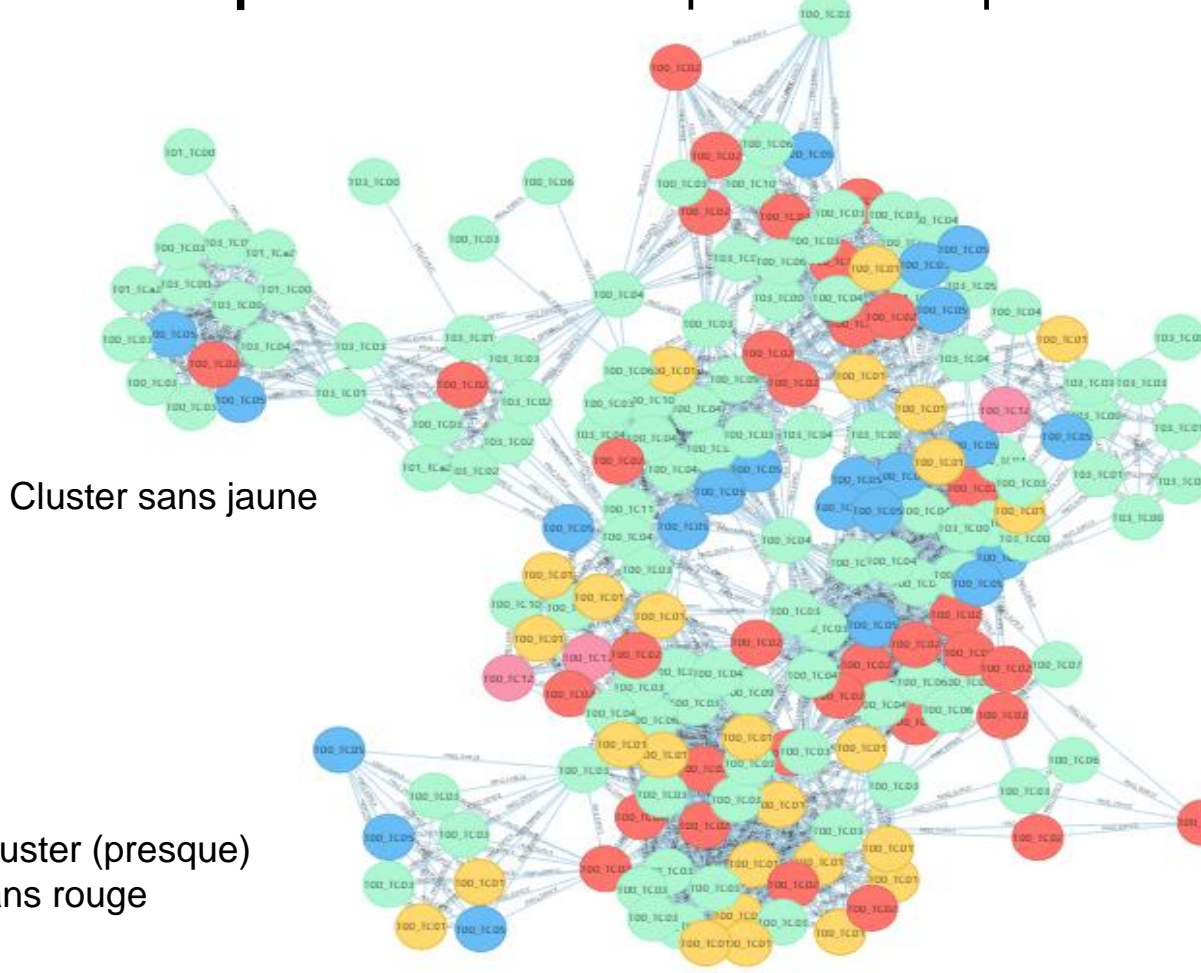
les liens pondérés : les fréquences d'espèces les plus abondantes



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



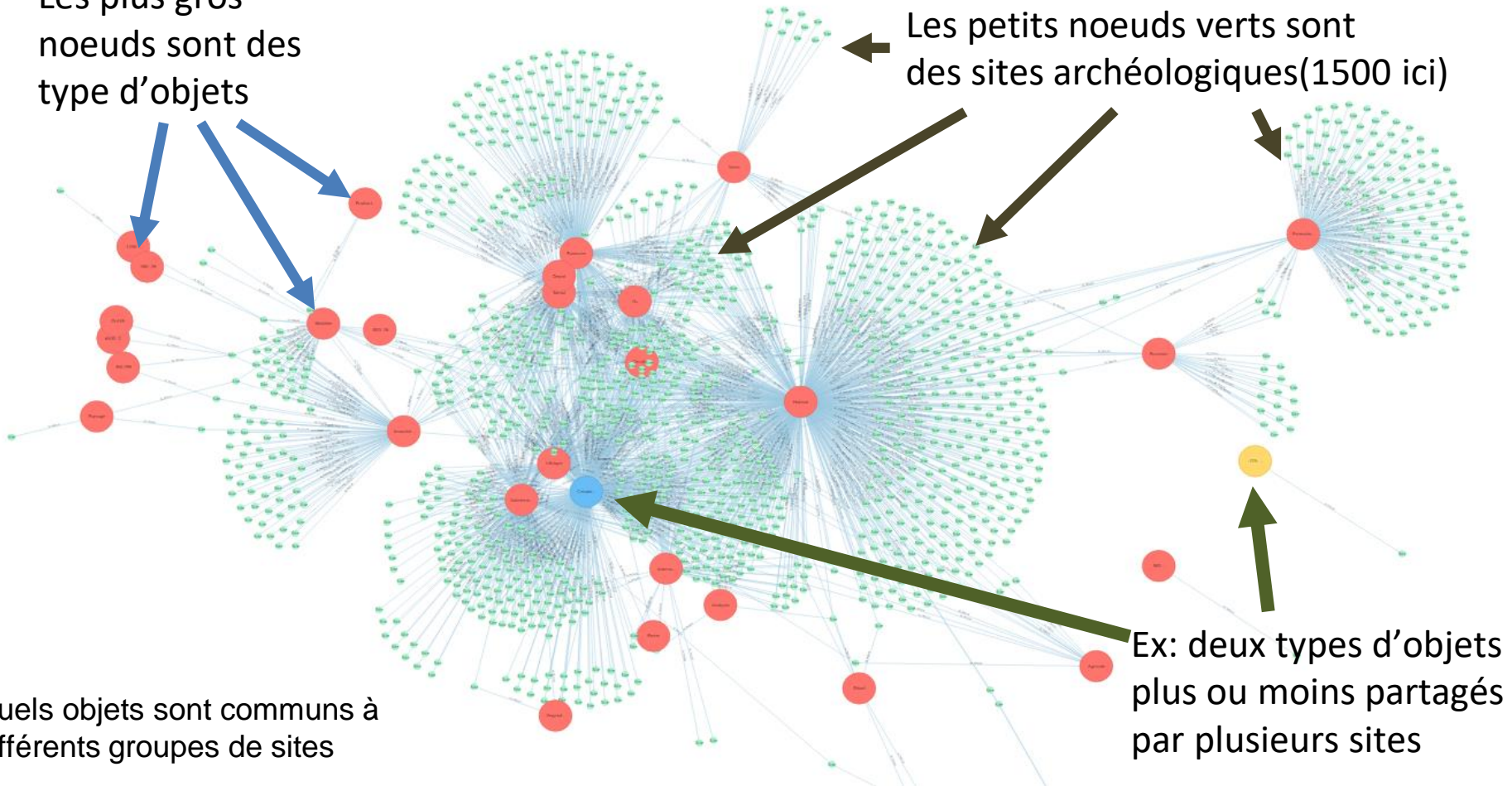
Peut on visualiser l'absence de certains transects dans certains clusters représentant une structure particulière des assemblages d'espèces ?

3) Discussion : la fouille de données

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

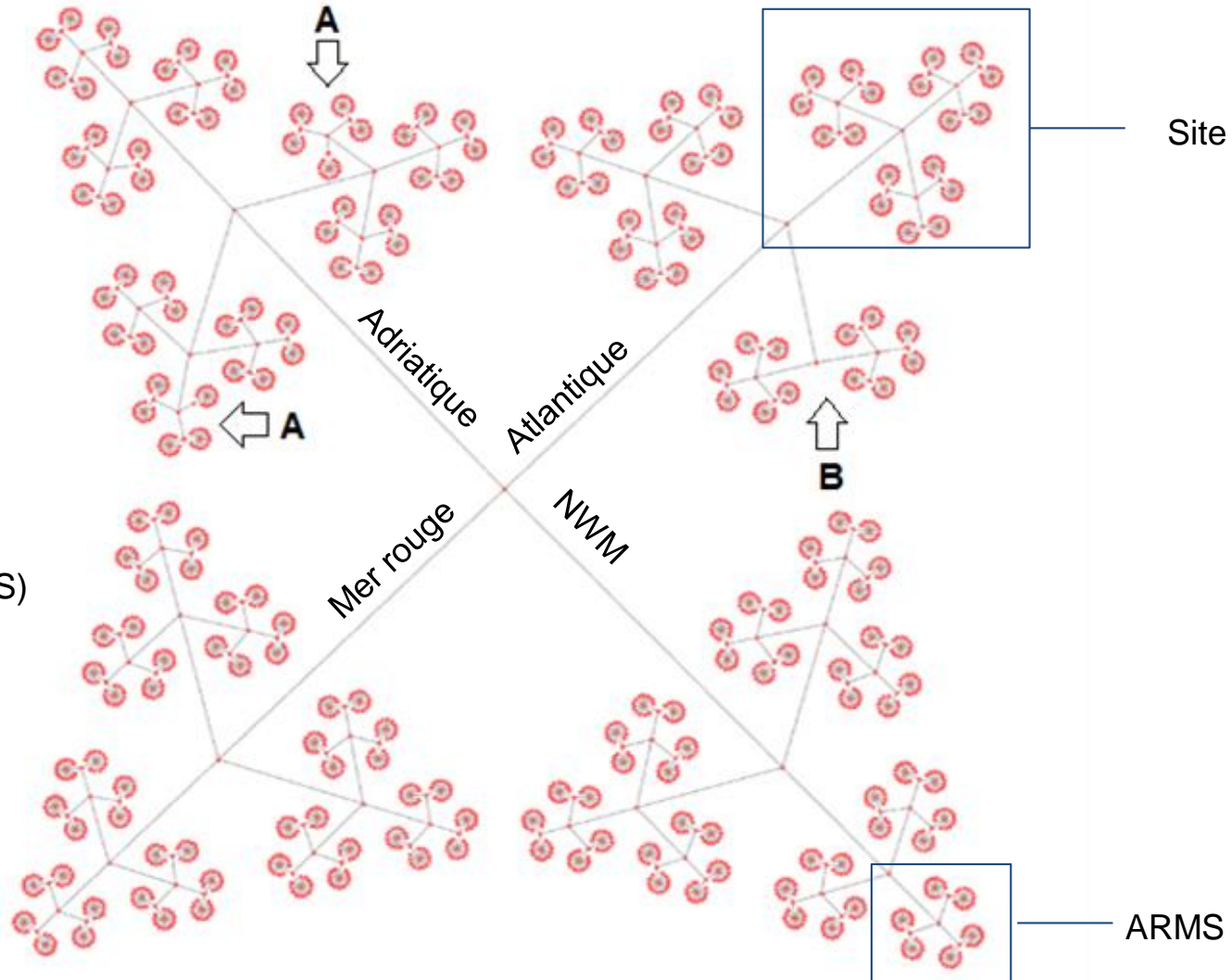
Les plus gros
noeuds sont des
type d'objets

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



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

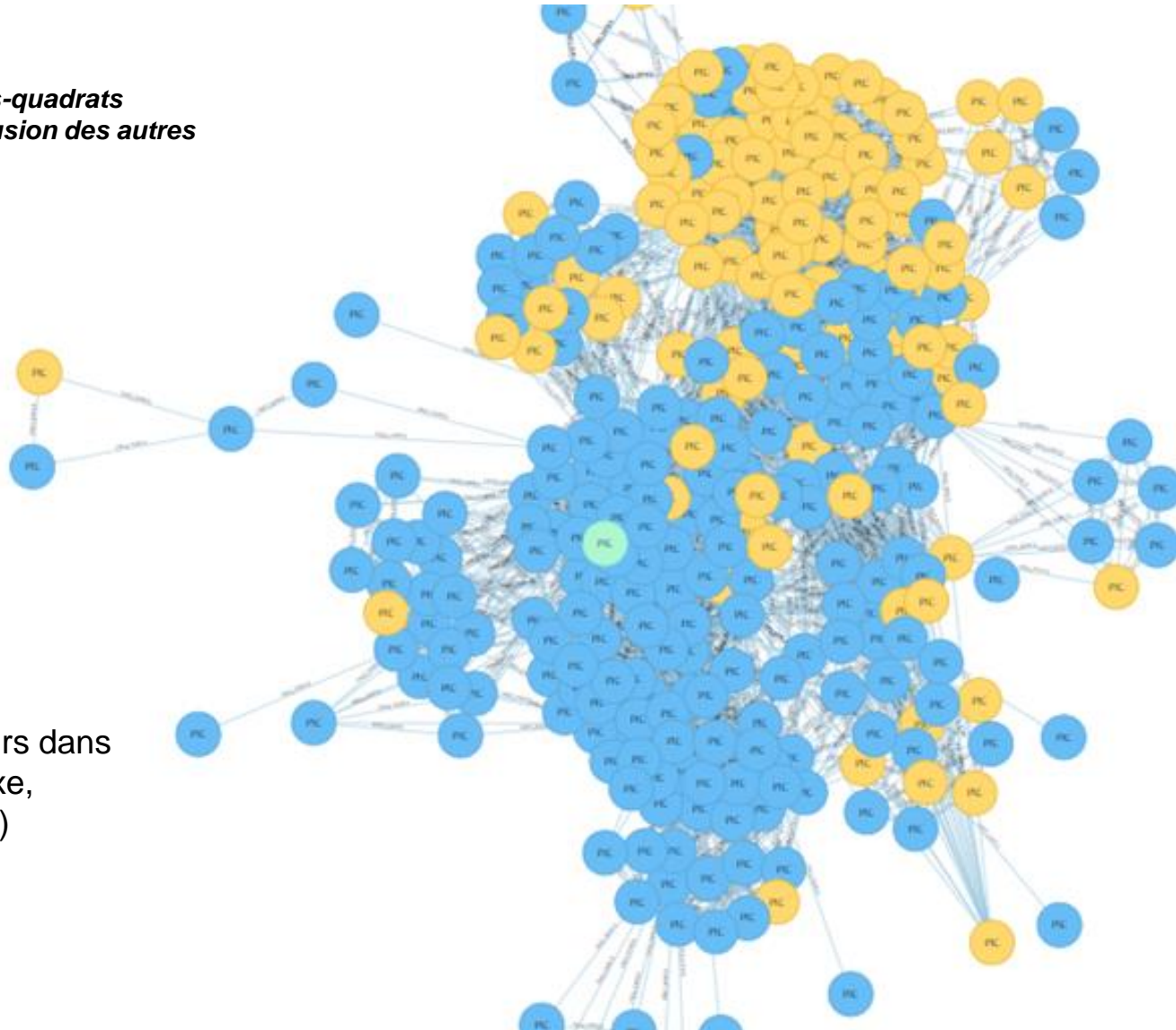
Visualisation du plan
d'échantillonnage
des faces de plaques
d'ARMS
(programme DEVOTES)



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

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

*En orange : incliné
En bleu : vertical*

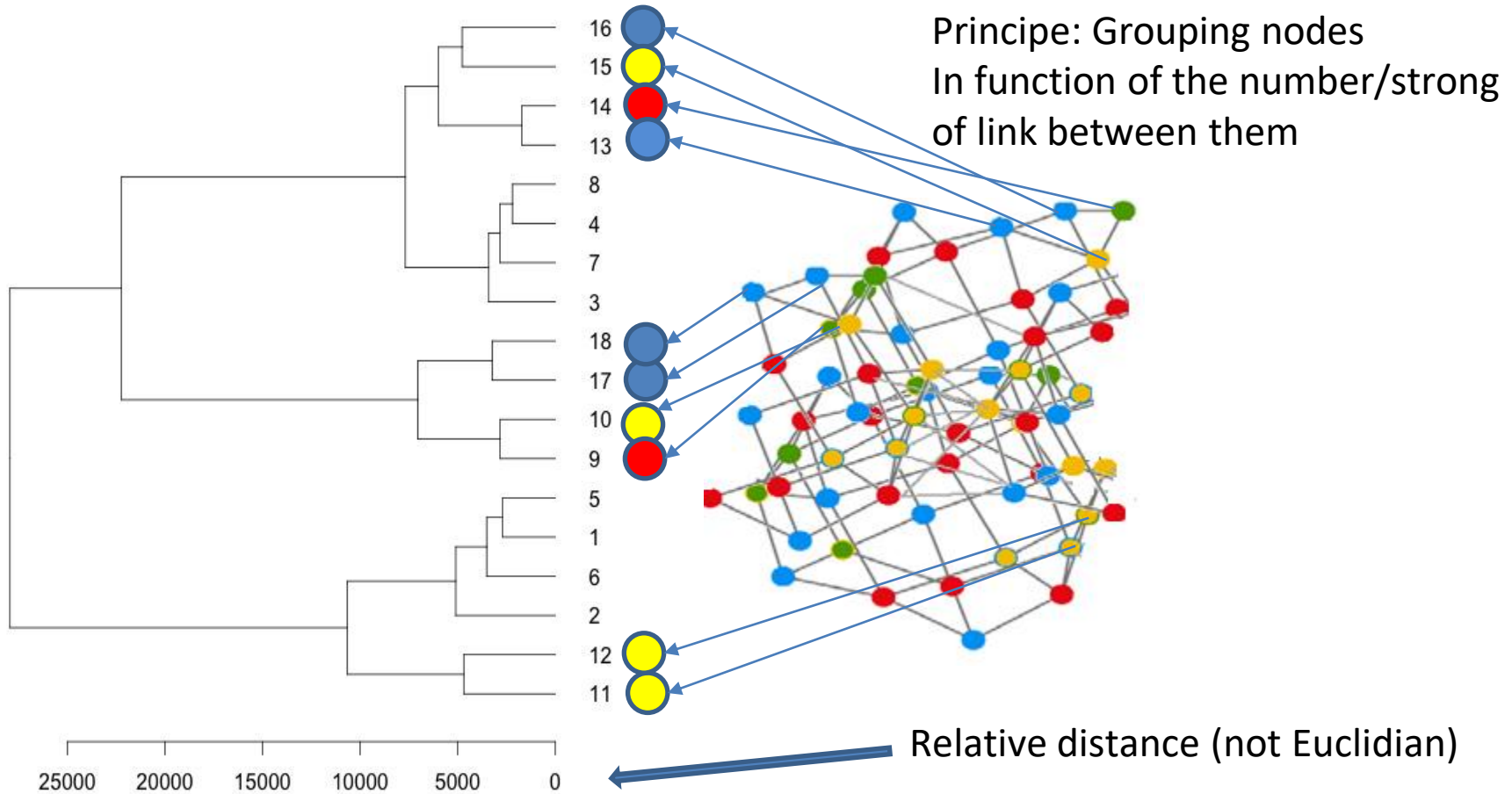


Détection des erreurs dans
les données (syntaxe,
donnée manquante)

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

3) Sélection de concepts sur les systèmes d'information

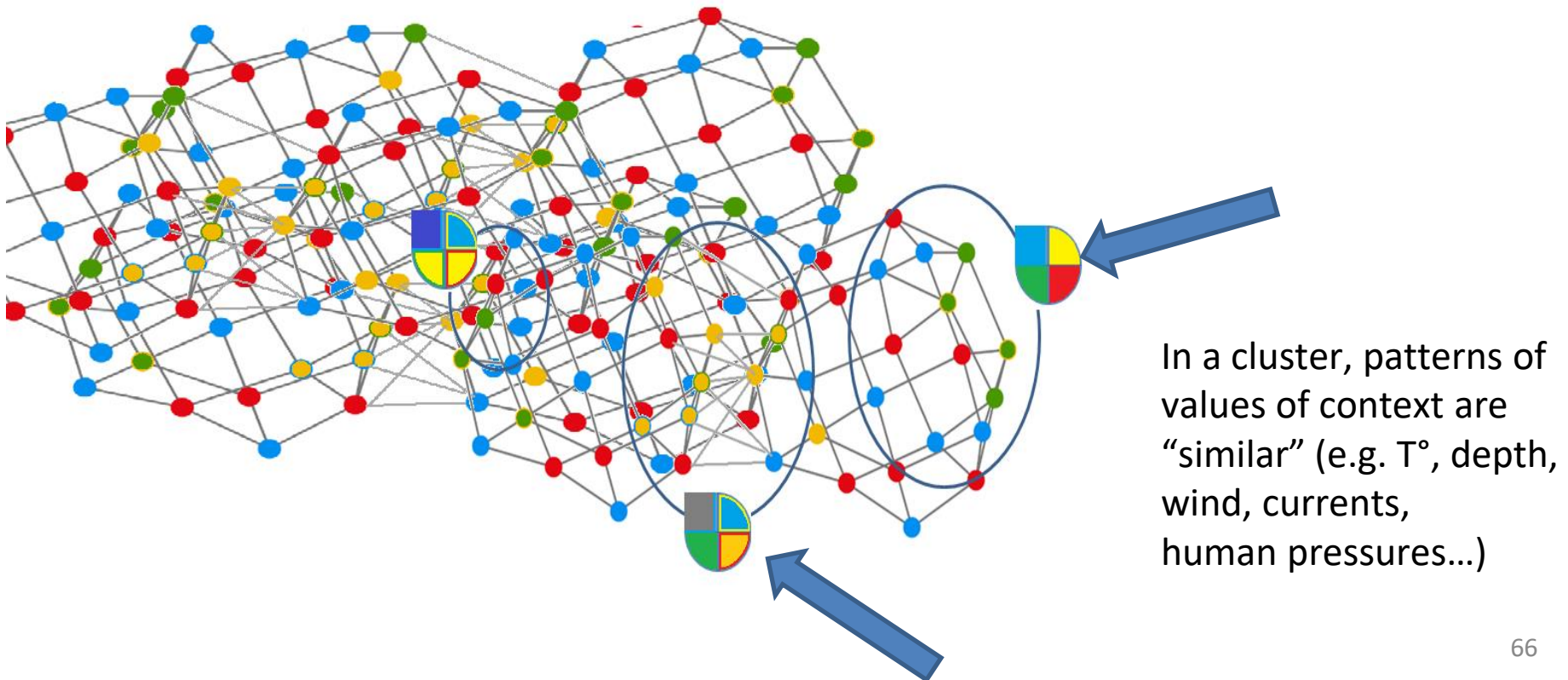
Comment faire parler ces données hétérogènes et forcément imparfaites ?



3) Discussion : 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



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)

Questions ?

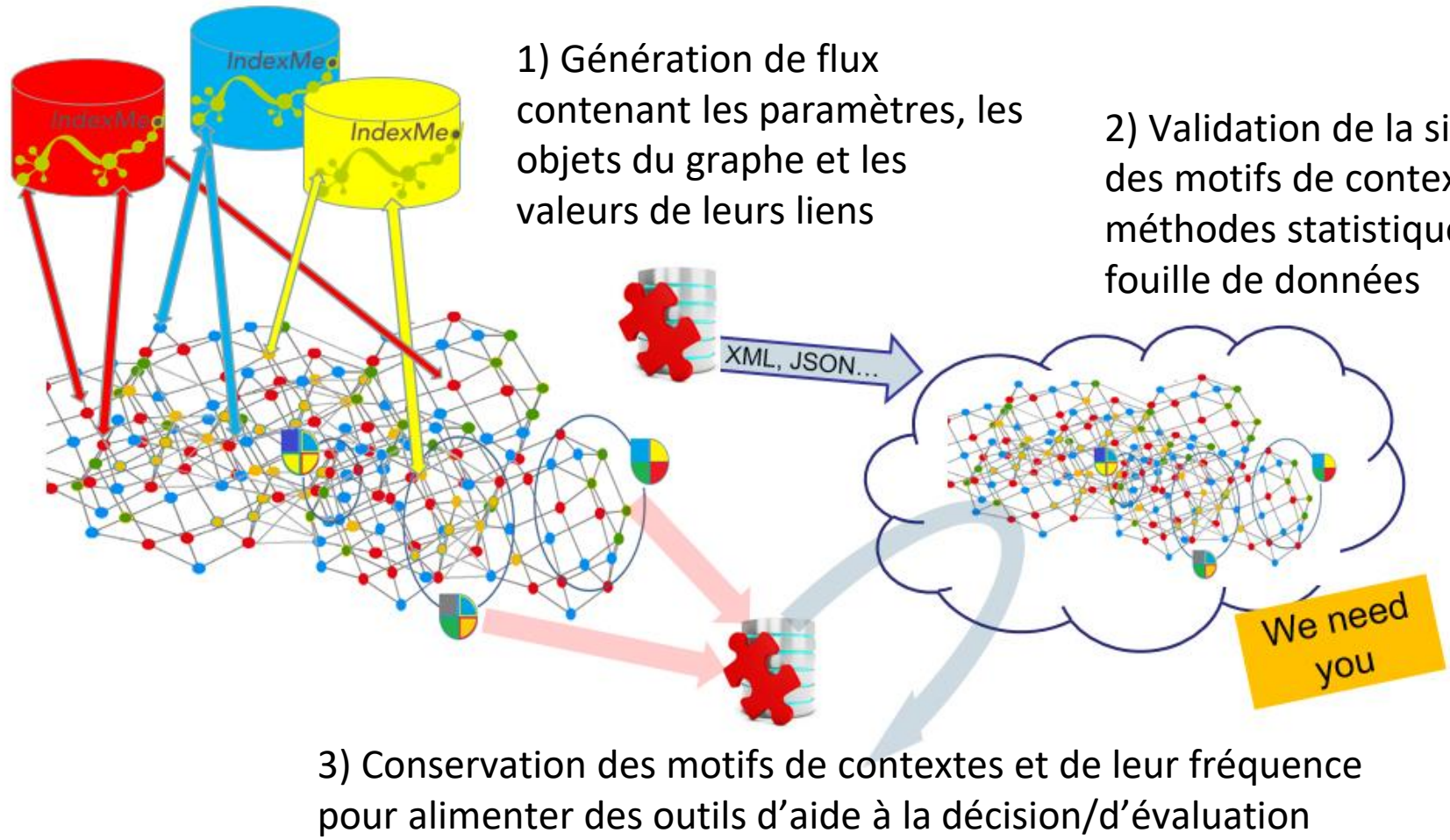


Cycle pictures from : http://cargocycling.org/category/riding_type/family-cycling

Remerciements au jury, à mes
encadrants et à tous les partenaires,
collègues et étudiants qui ont
participé à ces travaux

3) Discussion : recherche opérationnelle et perspectives

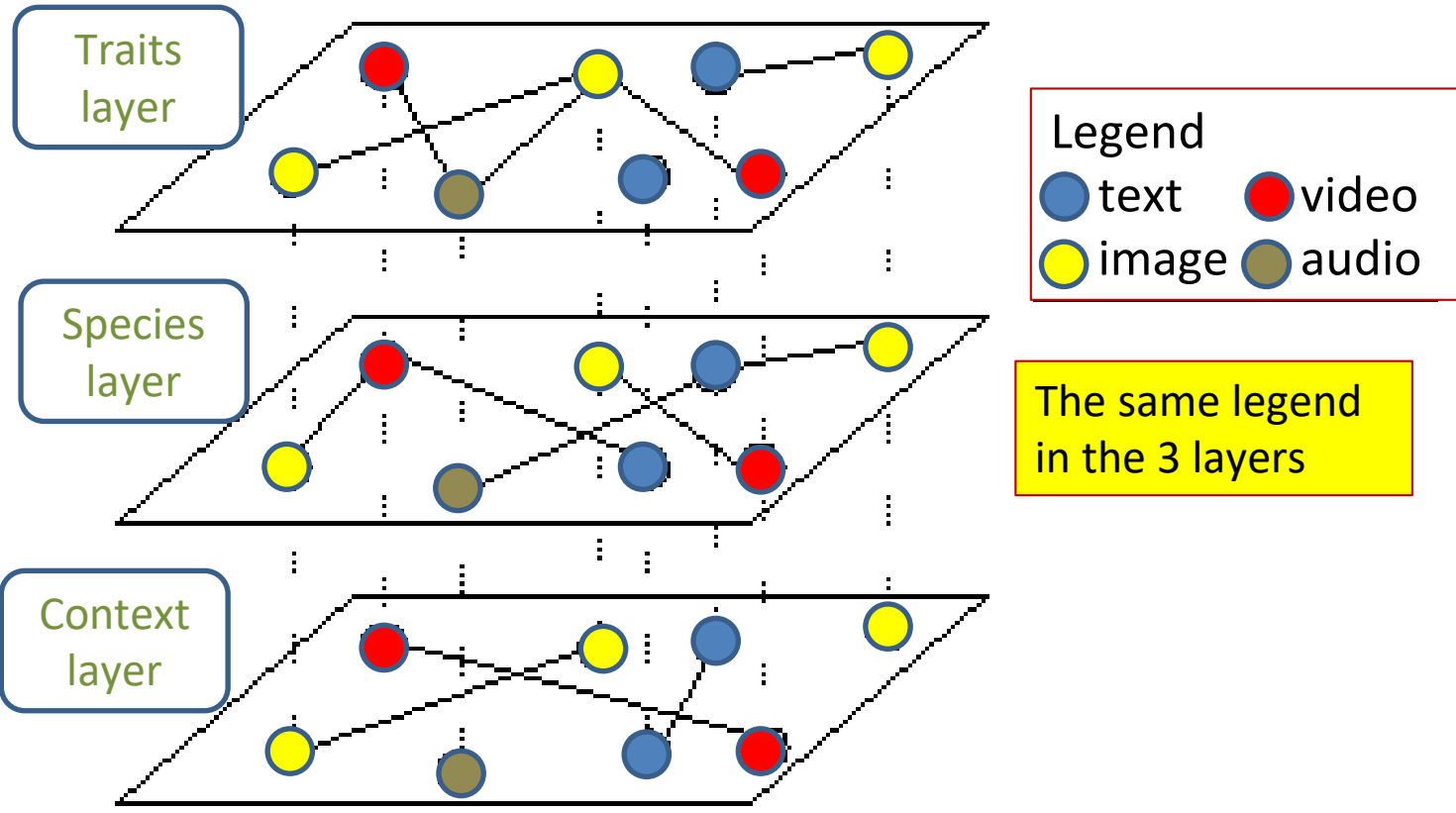
Quels systèmes prévisionnels d'aide à la décision ?

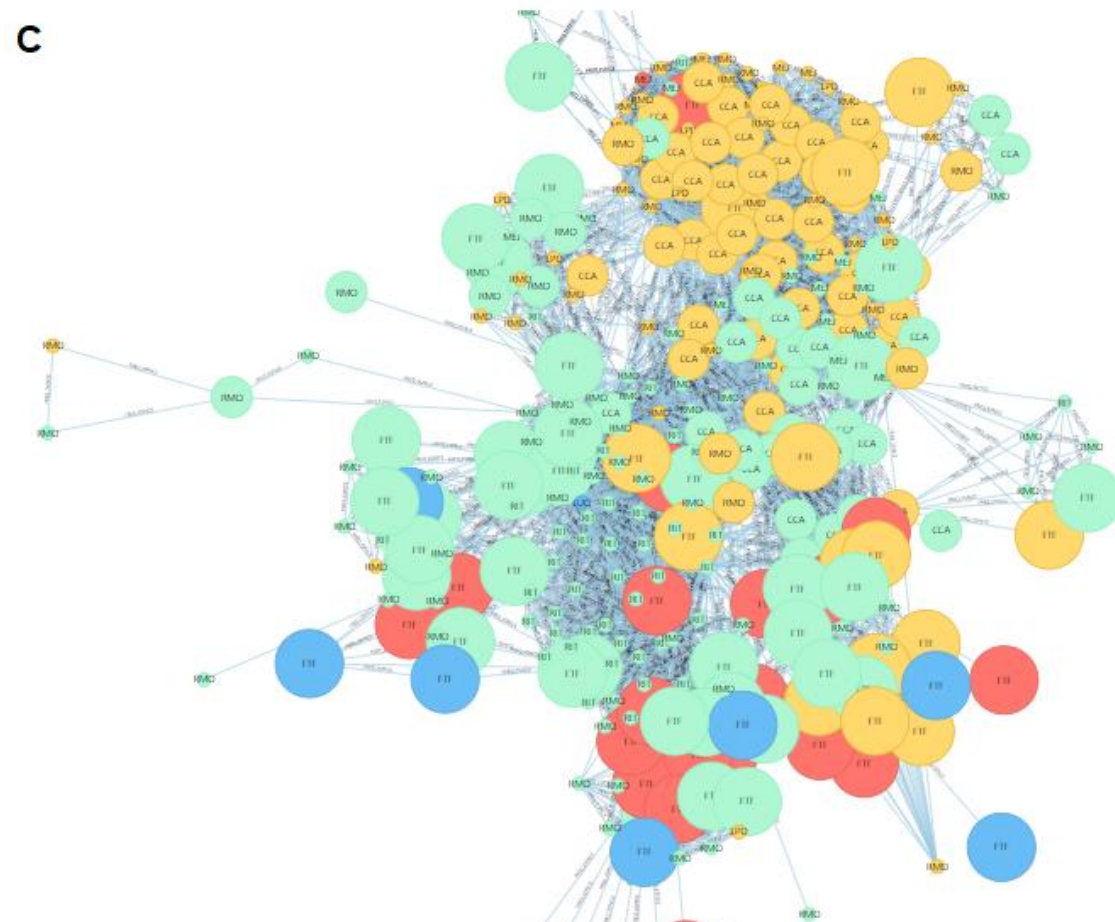
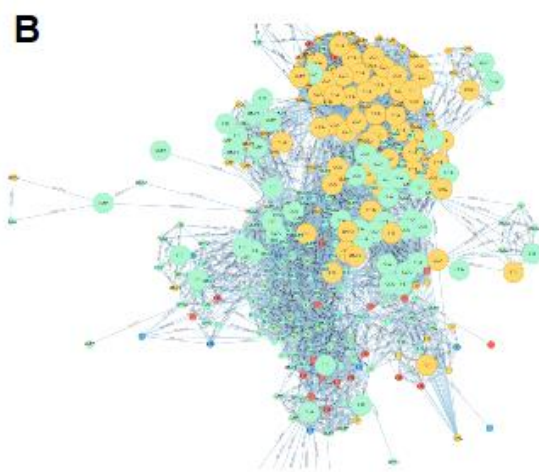
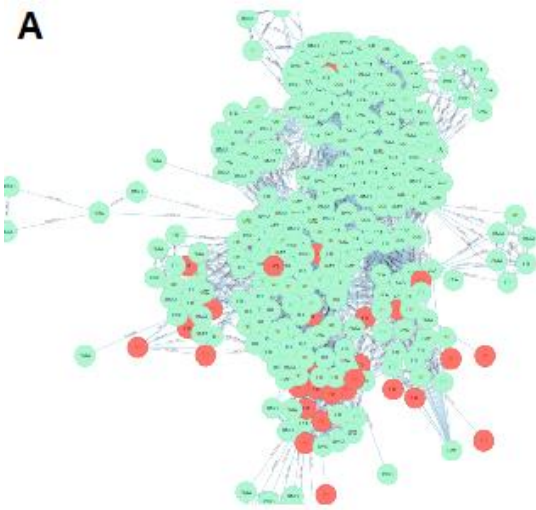


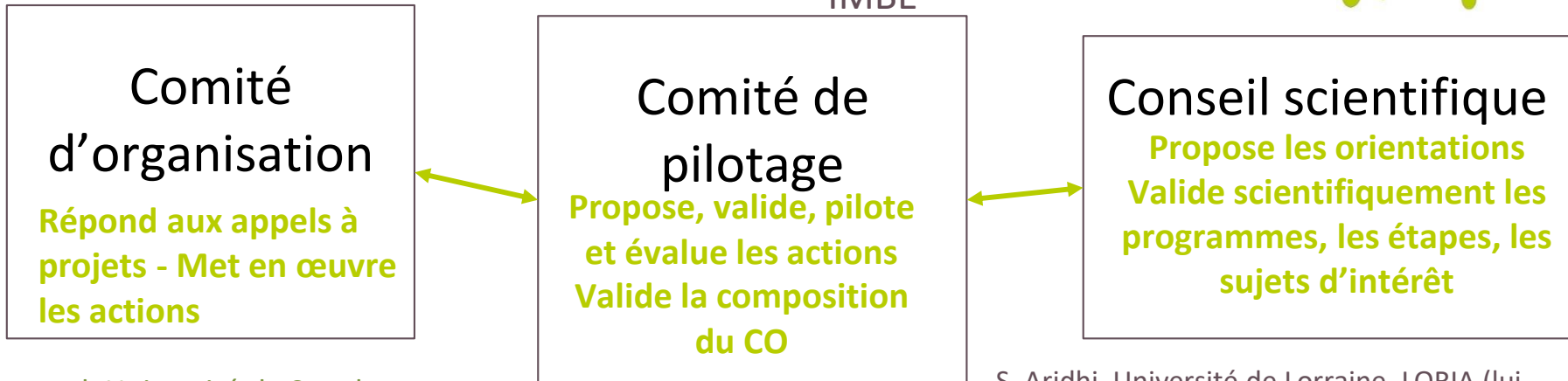
3) Discussion : la fouille de données

Quels systèmes prévisionnels d'aide à la décision ?

Multi Layer Graphs







Comité d'organisation

Répond aux appels à projets - Met en œuvre les actions

Comité de pilotage

Propose, valide, pilote et évalue les actions
Valide la composition du CO

Conseil scientifique

Propose les orientations
Valide scientifiquement les programmes, les étapes, les sujets d'intérêt

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

{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 écologuesd'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 (iEMSS) 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> . <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>