

Le big data en écologie?
Pas encore disent certains...

... Pas si sûr!

Avec IndexMed

Relevons ce challenge!



Interopérabilité
des bases
de données
en Écologie



Origine du consortium

L'état des lieux (source : séminaire Allenvie, séminaire Indexmed1)

En écologie marine les données sont encore peu structurées et peu accessibles

Même quand elles sont structurées, elles ne sont pas accessibles:

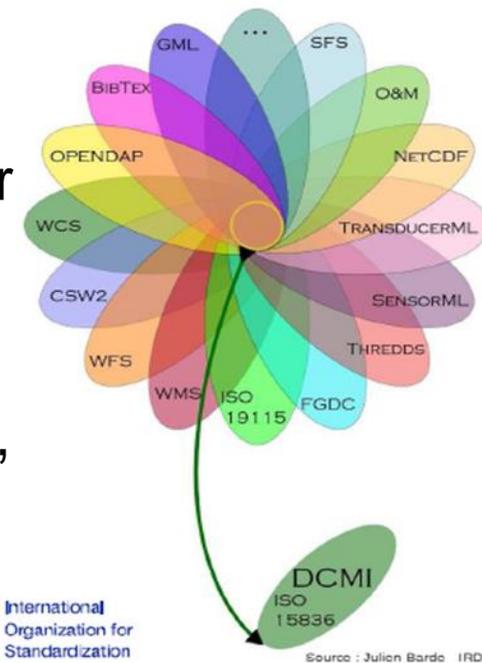
- Les normes changent beaucoup plus vite que permettent les capacités des bases de données
- Les compétences pour conserver les données et les partager sont souvent absentes des laboratoires

Et même pire, si les structures et les compétences existent

- Les données sont très peu utilisées et les bases de données, même publiques sont souvent des "trous noirs"

- **La tendance** est encore souvent à l'ultra **centralisation**

De nos jours, tout le monde veut les données de tout le monde, et ... les opérateurs changent très vite !!



Un réflexe...

My PREEEECIOUSS DATABASE

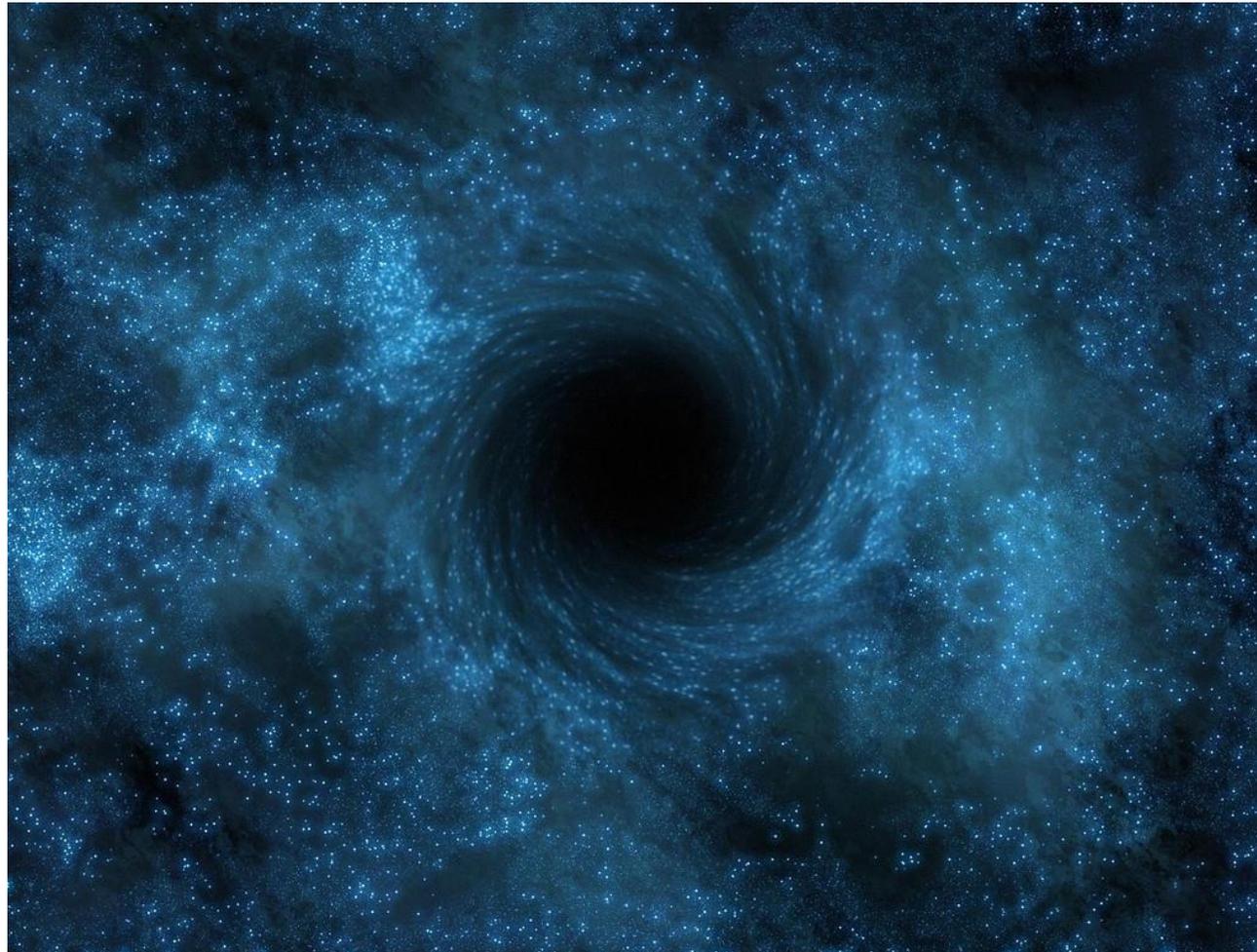


Tout le monde veut les données de tout le monde

**Besoins de mutualisation, partage
Pour les outils de conservation et d'utilisation des données**

Un état des lieux

La plupart des systèmes d'information publics n'informent personne



Origine du consortium

IndexMed est un consortium
EN CROISSANCE

créé par l'axe Gestion de la biodiversité et des espaces naturels de l'IMBE
(Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale).

Objectif principal :

Développer la culture des bases de données et leur utilisation efficace dans le milieu de la recherche en écologie et biodiversité.

Définition du consortium

Objectif principal :

développer la culture des bases de données

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,
Des données traçables et bien qualifiées,

Périmètre concerné : la biodiversité méditerranéenne terrestre et marine

Partenaires (à compléter)

- Un premier groupe formé pour réponse à appel à projet MASTODONS

Objectifs

Développement de points nodaux d'indexation massive et de qualification des données sur l'environnement littoral et marin méditerranéen

Construction de graphes à partir des indexations de données

Mise en évidence des relations « discrètes » entre valeurs de facteurs écologiques, économiques et sociaux

Partenaires (à compléter)

- Un premier groupe formé pour réponse à appel à projet mastodon
- Un séminaire qui a élargi ce groupe et permis de définir les objectifs du consortium

Présents inscrits et participants au séminaire

Séminaire "interopérabilité des bases de données en écologie"

5 juin 2014, Marseille



Inscrits : 70 [IMBE : 29]

dont présents : 55 [IMBE : 23]

+ non inscrits : 8

Total présents : 63



Partenaires (à compléter)

- Un premier groupe formé pour réponse à appel à projet mastodon
- Un séminaire qui a élargi ce groupe et permis de définir les objectifs du consortium
- Un site internet avec un appel à manifestation d'intérêt par formulaire

Seulement quelques minutes pour acter votre participation au consortium IndexMed

À l'origine, IndexMed est une réponse à un appel d'offre du CNRS (MASTODONS). L'intérêt montré par les participants et celui, encore plus grand qui s'est manifesté depuis, pour un groupement qui permettrait un rassemblement pluridisciplinaire et une organisation pour faire émerger et cerner les questions concernant les données de masse, rendent nécessaire la création d'un consortium.

Afin de renforcer ce qui sera le consortium IndexMed et de préparer des projets s'appuyant sur tous les volontaires et toutes les compétences disponibles, nous proposons ce questionnaire aux personnes ayant assisté au séminaire ou ayant manifesté leur intérêt (vous pouvez éventuellement le faire suivre aux personnes susceptibles d'être intéressées).

L'objectif est de compléter le relevé de discussion du séminaire et de préparer un projet global, pluridisciplinaire, s'appuyant sur des personnes issues des différents instituts de recherche, agences et associations environnementales, et d'augmenter le potentiel de futures collaborations.

Merci de répondre à ces questions de la manière la plus complète et argumentée possible, afin que vos éléments de réponse puissent utilement compléter les tâches/tasks des projets en préparation, et nous permettent de mieux cerner les besoins de financement. Une vision européenne est aussi à l'ordre du jour (H2020).

Ce questionnaire est aussi destiné à l'inventaire des forces qui constituent IndexMed (personnes, unités de recherche, instituts, agences, associations). Les fiches réponses seront utiles à l'établissement de la liste officielle des participants.

Il peut être rempli en ligne (ci-dessous), ou bien en utilisant ce fichier à renvoyer à cette adresse jean-pierre.feral@imbe.fr, romain.david@imbe.fr

Ce formulaire comporte deux parties, (1) vos commentaires sur le séminaire Interopérabilité des bases de données en Ecologie [les 3 premières questions] et (2) votre adhésion aux développements futurs [questions suivantes].

Bilan du Séminaire du 5 juin 2014 et futur du consortium IndexMed -----

D'après les conclusions du séminaire (<http://www.indexmed.eu/Bilan-et-conclusions-du-seminaire.html>), êtes-vous intéressés par les thèmes de recherche proposés par IndexMed ? (1 à 5) [Obligatoire] :

- 1 - pas du tout intéressé(e) 4 - intéressé(e)
 2 - un peu intéressé(e) 5- très intéressé(e)
 3 - intérêt moyen

ATTENTION: UN SEUL LIEN UNIQUE

Présents inscrits et participants au séminaire ...

Bilan et conclusions du Séminaire du 5 juin 2014

Relevé d'idées et conclusions, issus des présentations, du débat et des conversations en marge du séminaire

Points à améliorer pour un prochain séminaire

Définition et séminaire d'action proposés pour IndexMed

Adhérer au consortium IndexMed

Agenda

2014						
<< Novembre						
Aujourd'hui						
Lu	Ma	Me	Je	Ve	Sa	Di
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
Aucun événement à venir les 6 prochains mois						

= plus de 100 manifestations d'intérêt

Partenaires (à compléter)

Consortium



IndexMed

Interopérabilité
des bases
de données
en Écologie

Plan du site | Rejoignez nous | Actions

Accueil du site > Actions

Rechercher dans le site >>

- Qui sommes nous ?
- Actions >
- Les projets
- Comité de pilotage

- Contact
- Connexion

RSS VALID

Actions

Actions

- ▶ Réponse à l'appel d'offre CNRS MASTODONS
- ▶ Organisation du premier **symposium sur l'interopérabilité des bases de données en écologie**.
- ▶ **PREDON**  **Workshop on Scientific Data Preservation** (MASTODONS)

Atelier sur la Préservation des Données Scientifiques

5-6 Novembre 2014, APC, Paris

Methods, practices and projects for data preservation : community projects, work on data preservation standards, exchanges with libraries and information sciences, policies, legal aspects of data preservation.

Romain David : Préservation de données dans le contexte IndexMed

Navigation

Sous-rubriques

Réponse à l'appel d'offre MASTODONS 2013
Premier séminaire "Interopérabilité des bases de données en écologie"

Agenda

2014						
<< Novembre						
Aujourd'hui						
Lu	Ma	Me	Je	Ve	Sa	Di
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
Aucun événement à venir les 6 prochains mois						

Partenaires (à compléter)



Consortium
IndexMed
Interopérabilité
des bases
de données
en Écologie

Plan du site | Rejoignez nous | Actions

Accueil du site > Actions

Rechercher dans le site

Navigation

Sous-rubriques

Réponse à l'appel d'offre MASTODONS 2013
Premier séminaire "Interopérabilité des bases de données en écologie"

Agenda

2014

<< Novembre

Aujourd'hui

Lu	Ma	Me	Je	Ve	Sa	Di
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

Aucun événement à venir les 6 prochains mois.

Actions

- ▶ Réponse à l'appel d'offre CNRS MASTODONS
- ▶ Organisation du premier **symposium sur l'interopérabilité des bases de données en écologie**.
- ▶ **PREDON @ Workshop on Scientific Data Preservation** (MASTODONS)
Atelier sur la Préservation des Données Scientifiques
5-6 Novembre 2014, APC, Paris
Methods, practices and projects for data preservation : community projects, work on data preservation standards, exchanges with libraries and information sciences, policies, legal aspects of data preservation
Romain David : [Préservation de données dans le contexte IndexMed](#)

Actuellement :

Renforcement du consortium

Constitution du comité de pilotage

Définition des principaux work packages

Qui d'autre?

Quelles autres idées?

Méthode proposée

S'appuyer sur les compétences extérieures au domaine de l'écologie:
Astronomie, Physique, Mathématique, Informatique, Fouille de données...

Proposer des réponses à appel à projet de recherche transdisciplinaires:

En se reposant sur des principes partagés
(et des prototypes à améliorer/adapter)

Des bases de données ouvertes, interopérables,

Des données traçables et bien qualifiées,

Un périmètre bien identifié : la biodiversité méditerranéenne terrestre et marine

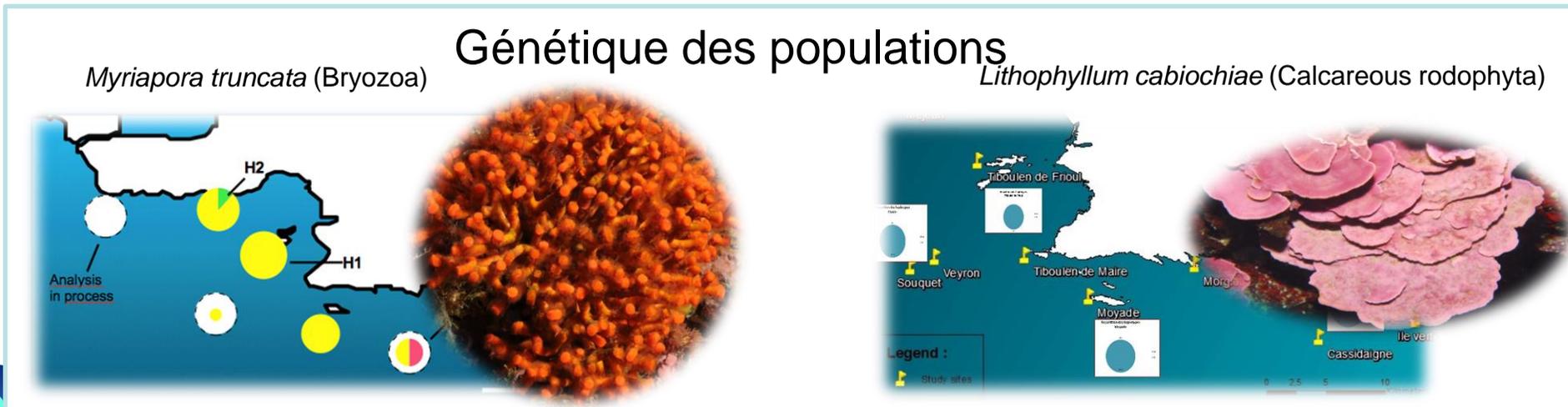
ET SURTOUT UNE COMMUNAUTE VOLONTAIRE (environ 100 personnes ont manifesté leur intérêt pour la démarche) et ouverte aux différents instituts de recherche!

Les données concernées

Des données aux formats hétérogènes... (exemple sur les habitat coralligènes)



Observations d'espèces via photo, vidéo ou prospection de terrain



Les données concernées

Des données aux formats hétérogènes... (exemple sur un type d'habitat : les coralligènes)

Données écologiques

Méditerranée :
Hot spot de biodiversité

Données socio-économiques

Plongée
Pêche artisanale
Pêche

En cours de déploiement :
Protocole de suivi opérationnel
à long terme, à large échelle, à moindre coût

Les données concernées

Des facteurs supplémentaires à prendre en compte

A large échelle : Effet
des compétences des
opérateurs



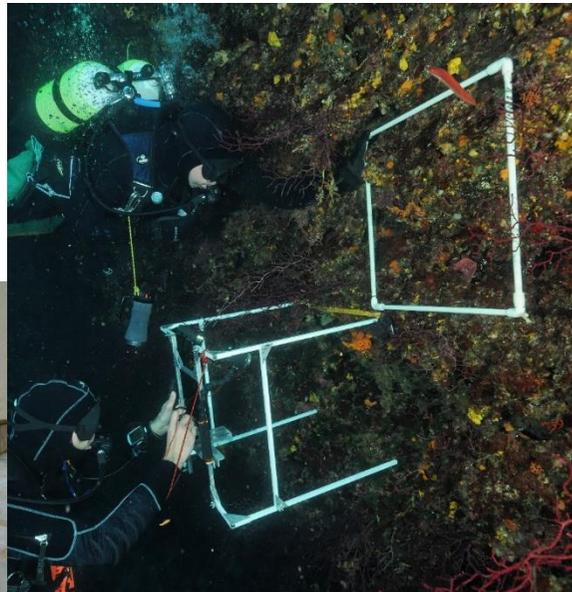
Les données concernées

Des facteurs supplémentaires à prendre en compte

A large échelle : Effet
des compétences des
opérateurs



Impact du type de
protocole sur les variables



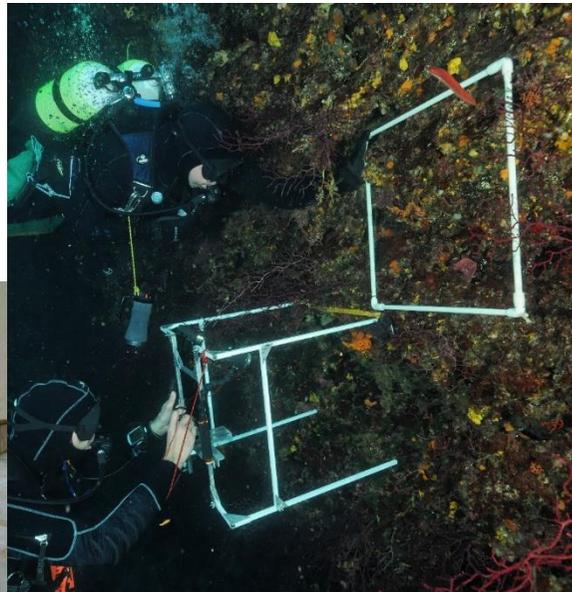
Les données concernées

Des facteurs supplémentaires à prendre en compte

A large échelle : Effet des compétences des opérateurs



Impact du type de protocole sur les variables



Effet de la qualité du matériel



Exemple de variables utilisées

Premier niveau

Dénombrement d'espèces reconnaissables *in situ*

Surface de recouvrement des espèces

% de surface occupée par espèce/taxon

Présence/Absence

Description morphométrique
(périmètre, diamètre, orientation)

Interpretation plus avancée

Abondance relative des espèces /
taxons

Fractionnement de la couverture des
espèces

Espèces voisines

**Pour une utilisation plus large (rapide et aisée)
les méthodes d'acquisition
de beaucoup de ces variables
doivent être reconsidérées**

Outils de prospection envisagés

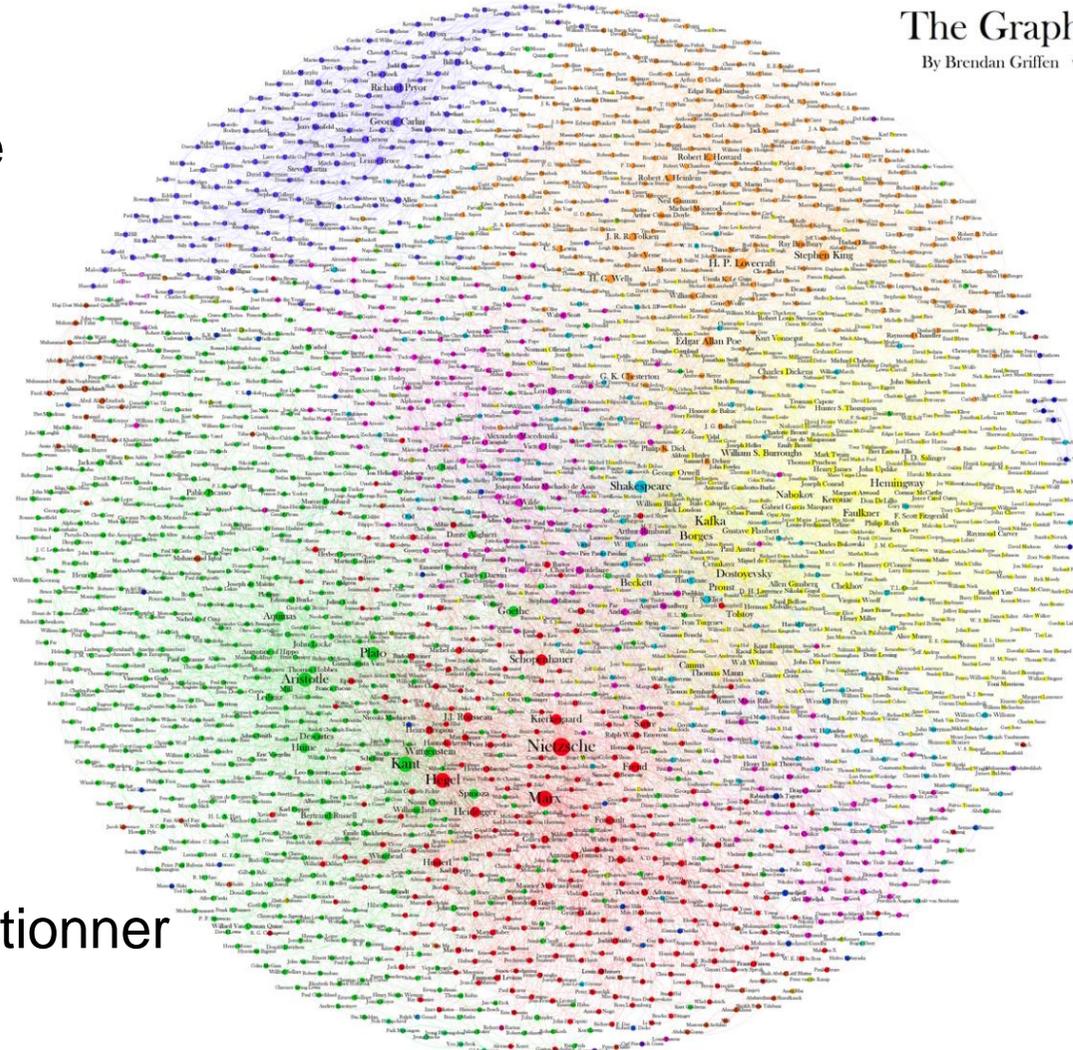
Un exemple prenant en compte
les idées en littérature

Identification par la densité
des liens, les valeurs des
catégories
qui sont proches

En ayant la possibilité de sélectionner
le type et les valeurs des liens

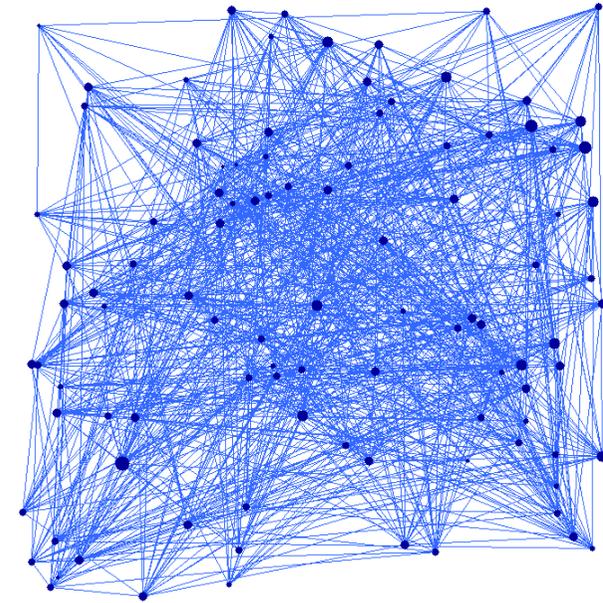
The Graph Of Ideas

By Brendan Grifflin www.grifflgraphs.com



Questions à prospecter (exemples)

- Facteurs pertinents pour qualifier un bon état écologique
- Valeurs de facteurs ou d'associations de facteurs pertinents pour mesurer une perturbation
- Dynamique et précurseurs (associations de valeurs) favorisant l'apparition de perturbations du milieu (perte de rendement, de fourniture de services éco-systémiques...

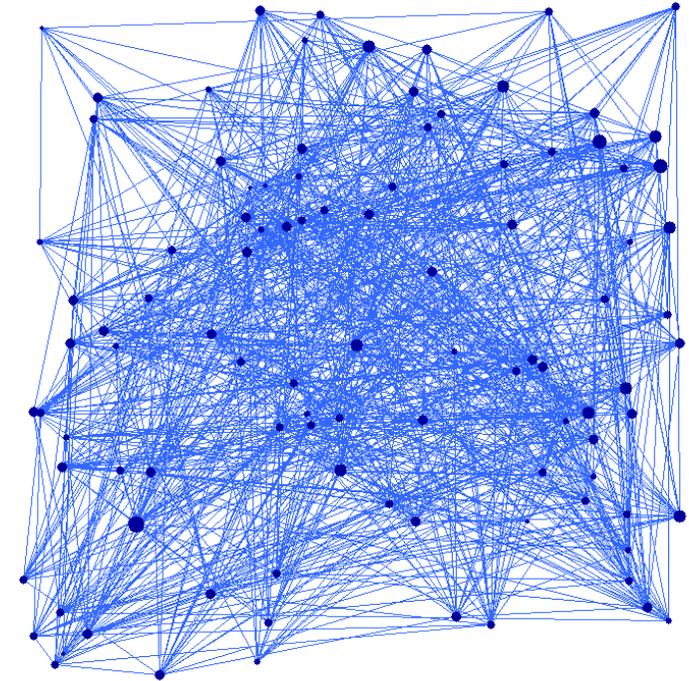


Questions à prospecter (exemples)

- Facteurs pertinents pour qualifier un bon état écologique
- Valeurs de facteurs ou d'associations de facteurs pertinents pour mesurer une perturbation
- Dynamique et précurseurs (associations de valeurs) favorisant l'apparition de perturbations du milieu (perte de rendement, de fourniture de services écologiques...)

Moyens (en très bref...)

- > Identifier grâce aux densités de lien, Les valeurs de catégories qui sont proches (ou éloignées)
- > avoir la possibilité de sélectionner le type et les valeurs des liens ainsi que le type d'objets composant ces graphes
- > Faire des analyses de proximité, de dispersion des valeurs et tester leur signification



Verrous

Les enjeux scientifiques sont liés à la gestion de données rendues complexes par

- leur volume, et la dynamique de leur mise à jour, de celle des référentiels et des normes qui permettent de les administrer
- leur hétérogénéité et leur complexité intrinsèques, notamment liées au croisement de données biodiversité et métriques contextuelles,
- l'hétérogénéité des utilisateurs, des réseaux d'acteurs producteurs et leurs motivations à maintenir et alimenter leurs systèmes d'information.

Défis techniques à relever

Système d'information

- **Distribué**
- Inter-opérable
- Donnée tracée

Work in link with productors
+ taking into account existing system



= distributed information system
Reference data by the productors

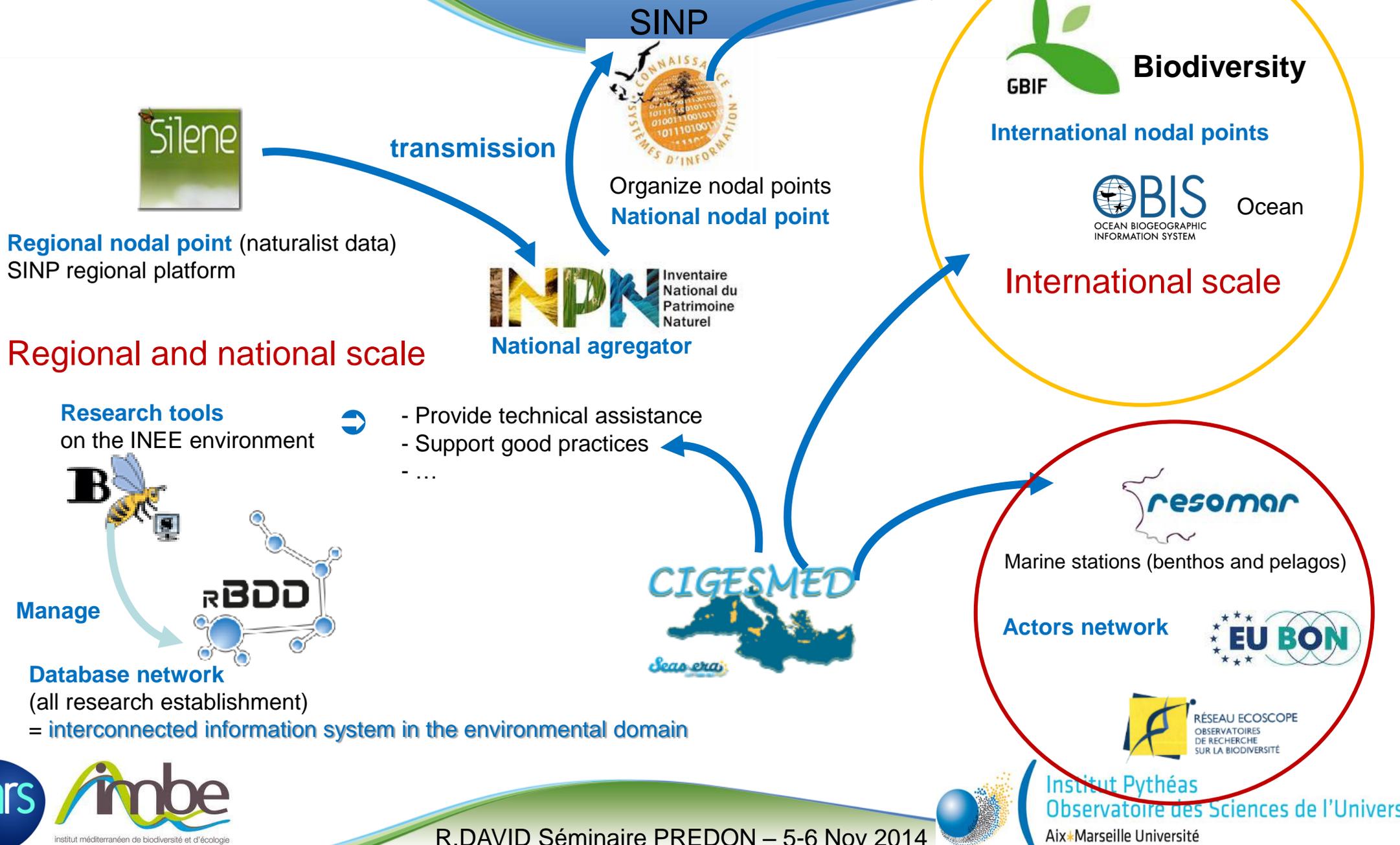
Défis techniques à relever

Systeme d'information

- Distribué
- **Inter-opérable**
- Donnée tracée



Inter-operability challenge



Défis techniques à relever

Système d'information

- Distribué
- Inter-opérable
- **Donnée tracée**

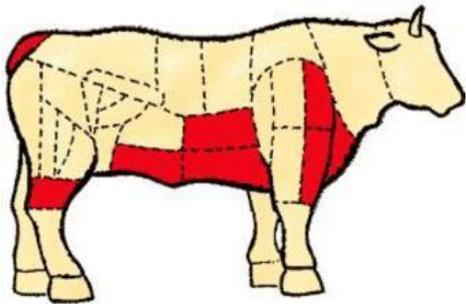
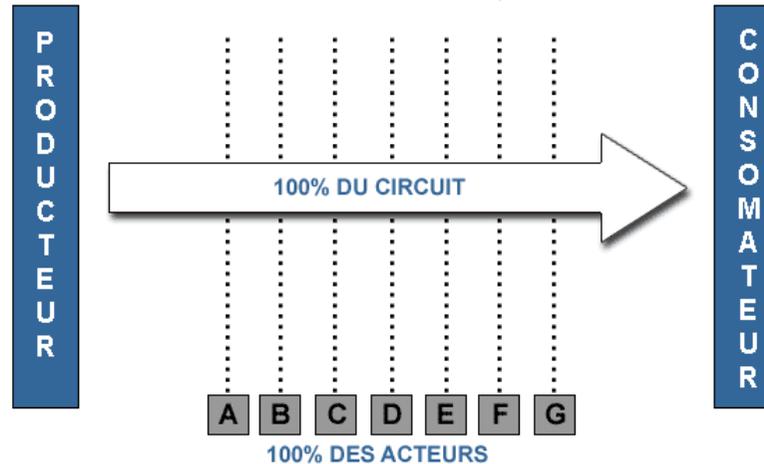


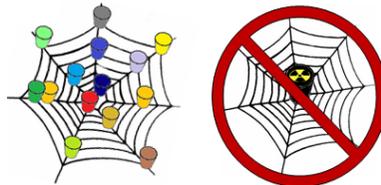
Schéma illustrant le 100% secteur, 100% circuit



Autrement dit :

Facteurs de qualité

Work in link with producers
+ taking into account existing system



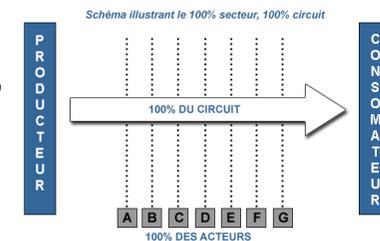
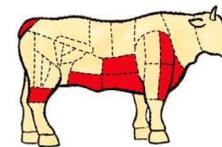
= **distributed information system**
Reference data by the producers

Système d'information

- Distribué
- Interopérable
- Donnée tracée

Objectif poursuivi :

Qualifier à large échelle
les objets de la biodiversité
D'une manière commune
(vocabulaire contrôlé, thésaurus et ontologie...)



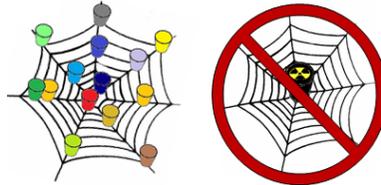
Autrement dit :

Facteurs de qualité

Système d'information

- Distribué
- Interopérable
- Donnée tracée

Work in link with producers
+ taking into account existing system

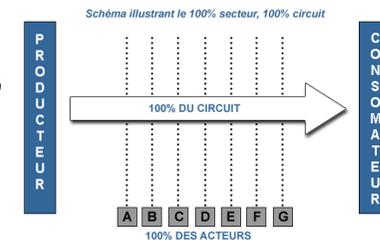
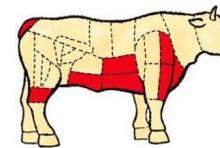


= **distributed information system**
Reference data by the producers

Les objets (sites, photos, espèces)
seront les nœuds du graphe,
les qualifications seront les liens

Objectif poursuivi :

Qualifier à large échelle
les objets de la biodiversité
d'une manière commune
(vocabulaire contrôlé, thésaurus et ontologie...)



Moyens pour lever ces verrous

L'efficacité des services répondant aux contraintes contradictoires (modèles des bases de données, optimisation des flux grâce à des services d'administration et des outils de paramétrages, gestion de droits d'utilisation en fonction des statuts) peut être gérée par une **indexation globale** appuyée par des **méthodes de qualifications** définies et validées par les acteurs sous forme de communautés thématiques.

Un soin particulier sera apporté aux aspects suivants :

- objectifs de chaque qualification, rapport coûts / avantages,
- inter-calibration de protocoles, d'observateurs, typologie adaptée,
- qualités des tatouages de jeux de données / systèmes de traçabilité (qualification / marquage) et rapports coûts / avantages.

Premiers work packages
Premières tâches proposées

Acquisition automatisée
Systèmes de reconnaissance automatique
Outils d'aide à la détermination
Data management
Indexation de la donnée
Gestion de la qualité et réutilisation de la donnée
Qualification et ontologie
Fouille de données
Conservation des données (Nouveau : les variables essentielles de biodiversité)
Représentations des données
...

Prochains rendez vous

Une réunion des membres du comité de pilotage

**Un work shop pour la mise en commun
des drafts des work packages**

Prochains rendez vous

Une réunion des membres du comité de pilotage

Seulement quelques minutes pour acter votre participation au consortium IndexMed

A l'origine, IndexMed est une réponse à un appel d'offre du CNRS (MASTODONS). L'intérêt montré par les participants et celui, encore plus grand qui s'est manifesté depuis, pour un groupement qui permettrait un rassemblement pluridisciplinaire et une organisation pour faire émerger et cerner les questions concernant les données de masse, rendent nécessaire la création d'un consortium.

Afin de renforcer ce qui sera le consortium IndexMed et de préparer des projets s'appuyant sur tous les volontaires et toutes les compétences disponibles, nous proposons ce questionnaire aux personnes ayant assisté au séminaire ou ayant manifesté leur intérêt (vous pouvez éventuellement le faire suivre aux personnes susceptibles d'être intéressées).

L'objectif est de compléter le relevé de discussion du séminaire et de préparer un projet global, pluridisciplinaire, s'appuyant sur des personnes issues des différents instituts de recherche, agences et associations environnementales, et d'augmenter le potentiel de futures collaborations.

Merci de répondre à ces questions de la manière la plus complète et argumentée possible, afin que vos éléments de réponse puissent utilement compléter les tâches/tasks des projets en préparation, et nous permettent de mieux cerner les besoins de financement. Une vision européenne est aussi à l'ordre du jour (H2020).

Ce questionnaire est aussi destiné à l'inventaire des forces qui constituent IndexMed (personnes, unités de recherche, instituts, agences, associations). Les fiches réponses seront utiles à l'établissement de la liste officielle des participants.

Il peut être rempli en ligne (ci-dessous), ou bien en utilisant ce fichier à renvoyer à cette adresse jean-pierre.feral@imbe.fr, romain.david@imbe.fr

Ce formulaire comporte deux parties, (1) vos commentaires sur le séminaire Interopérabilité des bases de données en Ecologie [les 3 premières questions] et (2) votre adhésion aux développements futurs [questions suivantes].

Bilan du Séminaire du 5 juin 2014 et futur du consortium IndexMed -----

D'après les conclusions du séminaire (<http://www.indexmed.eu/Bilan-et-conclusions-du-Seminaire.html>), êtes-vous intéressés par les thèmes de recherche proposés par IndexMed ? (1 à 5) (Obligatoire) :

- 1 - pas du tout intéressé(e)
- 2 - un peu intéressé(e)
- 3 - intérêt moyen
- 4 - intéressé(e)
- 5 - très intéressé(e)

AGENDA DE LA SEMAINE

Présents inscrits et participants au séminaire ...

Bilan et conclusions du Séminaire du 5 juin 2014

Relevé d'idées et conclusions, issus des présentations, du début et des conversations en marge du séminaire

Points à améliorer pour un prochain séminaire

Définition et périmètre d'action proposés pour IndexMed

■ Adhérer au consortium IndexMed

Agenda

2014

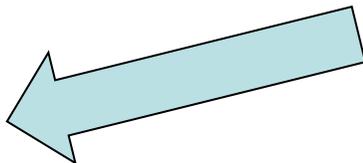
<< Novembre

Aujourd'hui						
Lu	Ma	Me	Je	Ve	Sa	Di
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

Aucun événement à venir les 8 prochains mois

Un work shop pour la mise en commun des drafts des work packages

L'appel à manifestation d'intérêt est toujours ouvert!



- **Question ou commentaires?**