

Réseau Bases de Données

Proposition de Groupe de Travail « Interopérabilité des données »

Eric Garnier^{1,2}, Marie-Claude Quidoz¹ & Baptiste Laporte²

1



CENTRE D'ÉCOLOGIE
FONCTIONNELLE
& ÉVOLUTIVE

UMR 5175, Montpellier

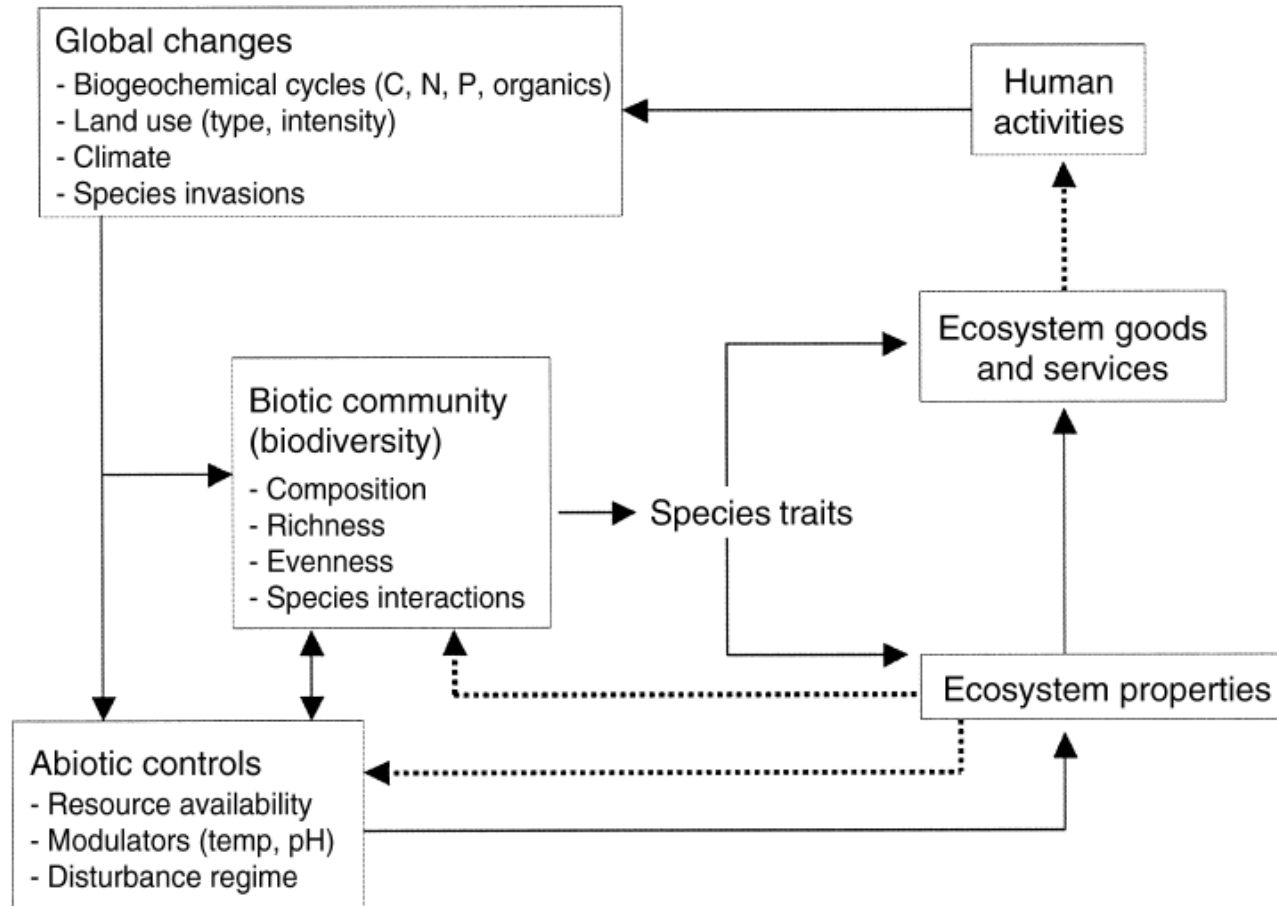


2



Un centre créé et développé par la FRB

Champ d'application: les systèmes écologiques (InEE)



Contrôles sur le fonctionnement des systèmes écologiques

- Déterminants de la structure et du fonctionnement des systèmes écologiques sont multiples et se rapportent à différents domaines scientifiques:
 - Sciences de la nature: facteurs du milieu (climat, sciences du sol), différents types de composantes biotiques (organismes, communautés, écosystèmes)
 - Sciences humaines et sociales: déterminants des variables de forçage, organisation sociale, etc.

Position du problème

- **Objectif:** trouver des moyens efficaces pour découvrir, accéder, intégrer, conserver et analyser d'importants volumes de données pour l'étude des systèmes écologiques
- **Problèmes:**
 - seule une petite fraction des données collectées en écologie est facile à découvrir et accessible: moins de 1% des données écologiques est accessible après publication et résultats associés
 - pas d'accès direct aux données: interprétations des données à travers des présentations et des publications
 - [limite la réutilisation possible dans d'autres contextes et/ou pour d'autres questions]

Etat des lieux et objectifs

(a) Data for research



Distributed repositories

- Protocols for contributing data, managed locally

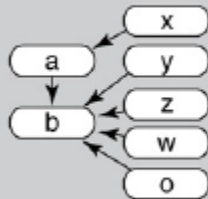
Open access

- Available to other projects and groups, enable reuse

Data diversity

- Support for wide range of data types and disciplines

(b) Standards for describing and interpreting data



Metadata

- Data structure, methods, access rights, authors, etc.

Core Ontology

- Community endorsed
- Observation and measurement

Specialized Ontologies

- Project-specific concepts
- Novel areas and viewpoints

(c) Tools for scientists and data managers



Data publishing

- Semantic annotation
- Automated (when possible)

Data discovery

- Ontology-enabled
- Summaries/ranked results

Scientific workflows

- Access and integrate data
- Locate, create, run analyses

Nécessité de l'interopérabilité

- Combinaison et intégration de différentes recherches afin d'améliorer la généralité et l'applicabilité des résultats de la recherche scientifique:
 - mise en commun de données issues d'observations/expérimentations conçues initialement de façon indépendante et/ou dans des contextes disciplinaires différents
- Produire de nouvelles connaissances à partir de ces données
- Interface entre écologie et sciences de l'information: développement de *l'écoinformatique*

Obstacles à l'interopérabilité

□ Techniques:

- Dispersion des données: petits jeux de données et sujets abordés très différents
- Hétérogénéité syntaxique et sémantique des données
- Provenance des données: traçabilité, assurance qualité de collecte à la publication

□ Socio-culturelles:

- Peu d'incitation au partage
- Peu de protection contre les prédateurs
- Manque de reconnaissance pour la collecte des données
- Déficit chronique de financements et de temps dévolus à la mise en forme et au stockage des données

Les conditions de l'interopérabilité

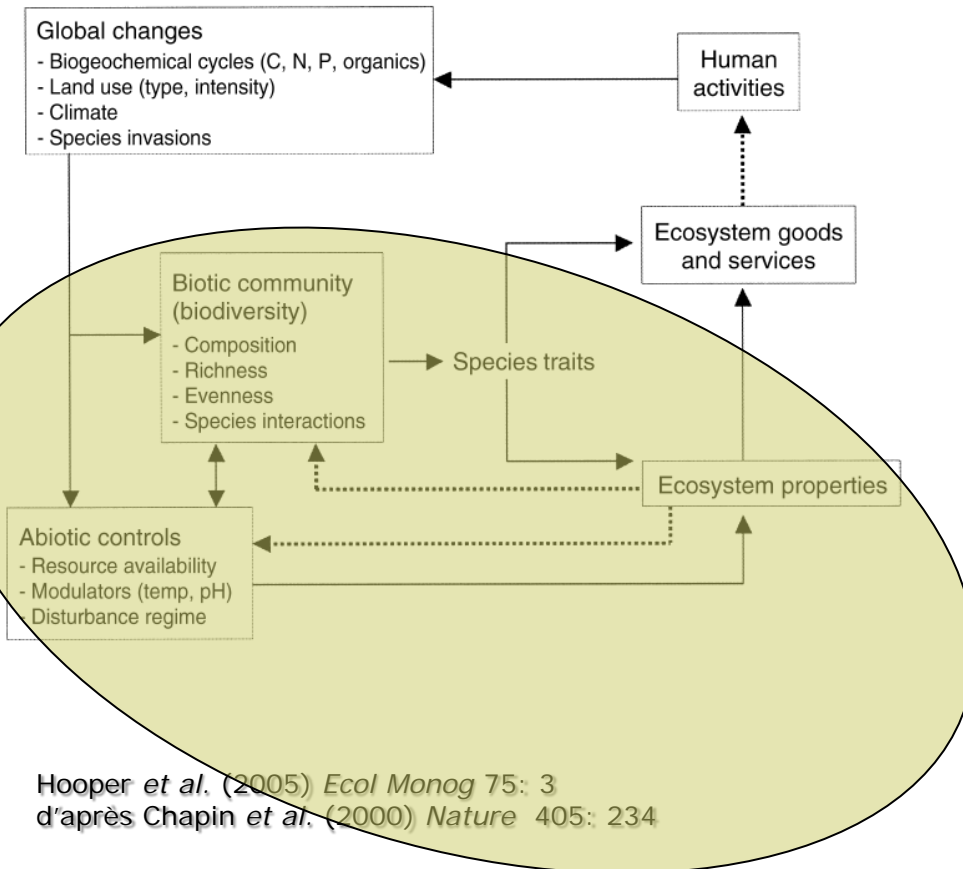
□ Volet technique:

- Fédérer les jeux de données indépendants (*e.g.* DataOne)
- Adoption de standards communs: métadonnées, vocabulaires contrôlés et ontologies
- Adoption d'outils communs utilisables par tous (*e.g.* interfaces, frameworks communs, API...)

□ Volet socioculturel:

- Améliorer la protection de la propriété intellectuelle
- Rendre les jeux de données publiables (*e.g.* « ecological archives »)
- Mise au point d'un modèle de gestion des données pérennes => financements à la hauteur des enjeux

Proposition d'un groupe de travail

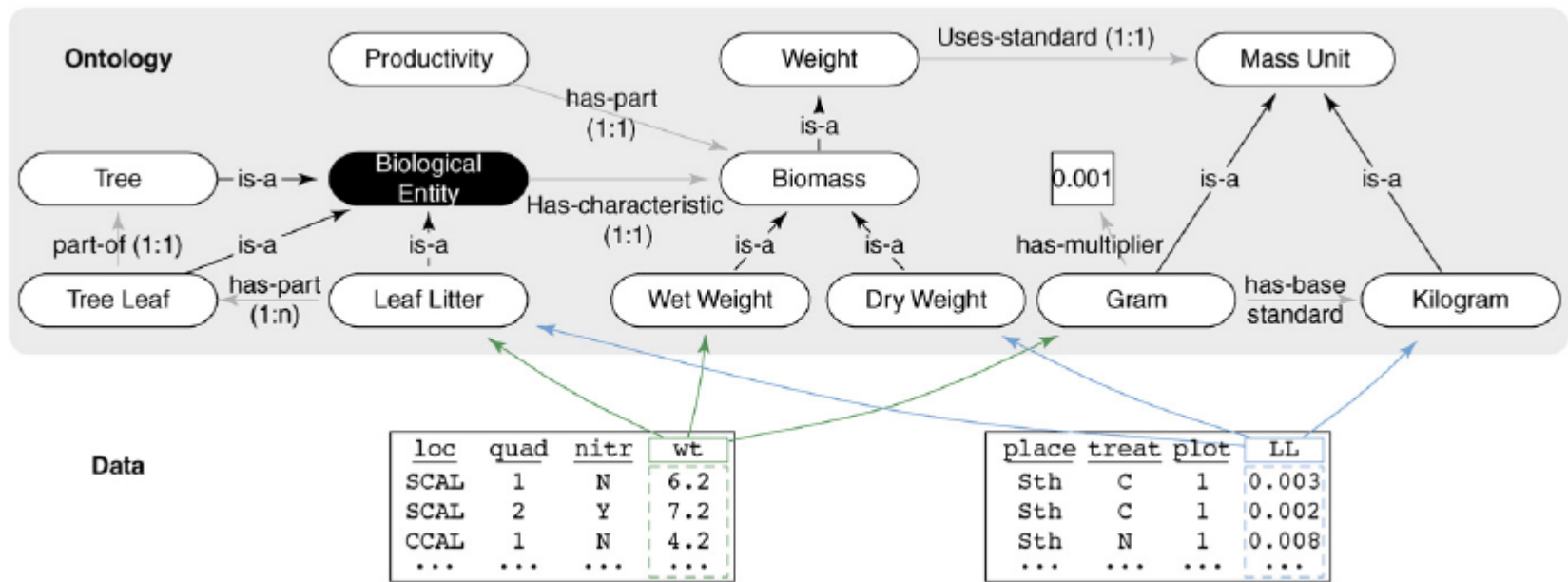


- **Objectif:** établir un prototype d'une méthodologie visant à rendre les données écologiques interoperables
- **Périmètre:** trois composantes des systèmes écologiques (prototype du prototype)
- **Composition:** groupe mixte de chercheurs en écologie et d'informaticiens

Organisation en trois tâches

- ❑ **Tâche 1:** articulation de jeux de données existants formatés selon des standards définis pour les trois types de composantes: *e.g.* « Xeml Environment Ontology » pour les facteurs biotiques, « Darwin Core » pour la taxinomie et « PLATON » pour les traits (plantes)
- ❑ **Tâche 2:** établissement de vocabulaires contrôlés, thesaurii et ontologies de domaine pour les différents champs disciplinaires
- ❑ **Tâche 3:** structuration et mise à disposition de bases de données existantes pour permettre le mapping des données avec les ontologies

Le Graal: mapper données et ontologies



TRENDS in Ecology & Evolution

**MERCI POUR VOTRE
ATTENTION**