

Conception d'une base de données pluridisciplinaire : intégration de connaissances pour gérer la complexité des agro- écosystèmes

Un exemple sur les interactions Microorganismes-Sol-Plante

Patricio Soto¹, Ange Balboa¹, Didier Blavet², Dominique Desclaux¹, Jean-Jacques Drevon², Anne Zanetto¹

Résumé. Il s'agit de contribuer à développer une base sémantico-conceptuelle nécessaire pour la description de toute donnée issue des agro-écosystèmes, dans le cadre d'une approche systémique. Ceci permet les échanges transdisciplinaires et ouverts des sciences liées à l'agro-écologie et à l'environnement, dans un objectif de développement durable. La finalité est de participer à l'élaboration d'une nouvelle méthodologie pour aborder l'information agro-écosystémique, visant à rendre celle-ci à la fois rigoureuse et opérationnelle. Au cours de ce développement, dans le cadre concret d'un projet (FabaTropiMed), une terminologie rigoureuse et adéquate a été progressivement dégagée, afin de rendre possible de façon efficace l'intégration des connaissances pour gérer la complexité des informations issues des agro-écosystèmes.

Mots clés : agro-écosystème, approche systémique, base de données, intégration de connaissances, pluridisciplinarité

Introduction

Pour mieux connaître et gérer la complexité des agro écosystèmes, les projets de recherche doivent être de plus en plus pluridisciplinaires, en intégrant des concepts larges au niveau d'échelles spatiales ou temporelles très diverses. Or ces projets donnent rarement naissance à des publications interdisciplinaires ou à des analyses systémiques. Deux grands types de difficultés sont généralement observés : l'une provenant des jargons utilisés et des sens différents attribués aux termes utilisés par chaque spécialiste avec des précisions très variables, l'autre liée à l'intégration d'objets divers analysés à différentes échelles spatiales ou temporelles dans une base de données communes.

Très rarement, les chercheurs font l'exercice en début de projet de définir chaque terme qu'ils emploient. Or les échanges entre spécialistes, ayant subi des influences différentes, sont difficiles. Ce qui est rigoureux pour l'un, est flou pour l'autre, ou encore le même terme peut avoir deux significations différentes (car adoptées à partir de disciplines différentes). Il en résulte une difficulté d'échanges conduisant à masquer, sous des vocables flous, l'hétérogénéité initiale des concepts pour ne pas soulever à tout instant trop de questions.

L'objectif de cette communication est de montrer, au travers d'un projet concret réunissant des chercheurs de génétique, biogéochimie, agronomie, écophysiologie, pédologie, microbiologie qui mobilisent des concepts agro-éco-environnementaux, comment il est nécessaire de définir un langage commun pour décrire les méthodes, procédés, protocoles, notations, observations, et la manière de développer une base de données communes à toutes ces disciplines.

¹ INRA, Plateforme expérimentale DIASCOPE (0398), Domaine de Melgueil, F34130 Maugeio, France patricio.soto@supagro.inra.fr

² INRA, UMR 1222 Eco&Sols, 2 Place Viala, F34060 Montpellier, France

Matériel et Méthodes

L'intégration des connaissances produites par les diverses disciplines étudiant le réel tend vers une gestion de connaissance qui soit la plus exhaustive, pertinente et explicite possible (Bouché, 1990 ; Bouché et Soto, 1994). Ces connaissances se basent sur une information qui s'appuie, d'un point de vue conceptuel et méthodologique sur deux supports : les données organisées en ensemble et les liens (ou relations) entre ces données (Soto, 1990).

Dans ce cadre général notre démarche s'appuie sur :

a) un Schéma Conceptuel Global pour l'intégration et la gestion des données agro-éco-environnementales ; b) la construction d'une base sémantico-conceptuelle adaptée à la conception et à la modélisation d'une base de données agro-éco-environnementale ; c) des outils de développement ad-hoc.

Résultats et Discussion

Le « **Schéma Conceptuel Global** » prend en compte cinq référentiels (Soto et Bouché, 1993 ; Soto et Boglio, 1996). Il est spatio-temporel-compositionnel-protocolaire-institutionnel (voir encadré ci-dessous). Celui-ci permet l'accès direct et la recombinaison de toutes les données initiales en fonction de chaque demande formalisée d'information. La **gestion de l'information agro-éco-environnementale** est faite de telle sorte que les données acquises à une échelle spatio-temporelle donnée soient utilisables à d'autres échelles et niveaux de perception de façon rigoureuse et opérationnelle. Cependant les caractéristiques percevables et acquérables à une échelle spatio-temporelle donnée peuvent être très différentes de celles qui le sont à d'autres échelles. Par exemple, dans le cas d'échelles spatiales décroissantes, on peut enregistrer un climat régional, à l'intérieur de celui-ci une prairie, dans celle-ci un faciès de végétaux associés, pour celui-ci noter le profil d'un sol où l'on observe un horizon dont on peut décrire l'azote total, ou seulement son état ammoniacal NH_4^+ , et dans ce dernier enregistrer la fraction de l'isotope ^{15}N .

En matière de base sémantico-conceptuelle, si les changements d'échelles se font avec une relative aisance lorsque l'on peut référer les données à des échelles métriques (au sens général du terme) d'espaces et de temps, il n'en va pas de même lorsque les référentiels spatio-temporels ne sont pas métriques. Nous sommes alors amenés à parler, par exemple, de lieu-dit, station, secteur, région naturelle, ... pour indiquer certains changements d'échelles spatiales.

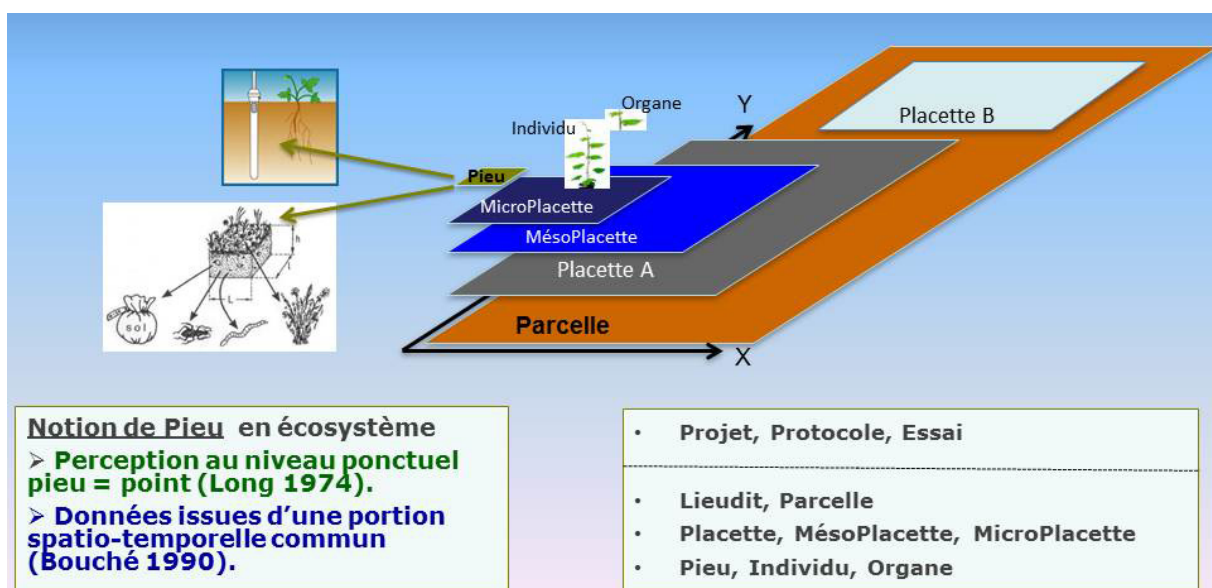


Schéma inspiré sur les projets bases de données : Adonis (Gavaland, 2011), Ecordre-Lombriciens (Soto, 1989) et Ecordre-Viticulture (Soto et Boglio 1996, Soto, 1997)

En permettant une gestion ordonnée de l'information exhaustive, non bornée par un objectif sectoriel, les Bases de Données élaborées devraient rendre possible des échanges optimisés. Elles seront accessibles par des modules d'interrogation et de saisie préexistants ou à développer qui, fédérés et couplés à des réseaux, permettront une mobilisation quasi-instantanée des données objectives et formalisées, quelle que soit leur origine disciplinaire ou géographique.

Intégration de connaissances : les cinq référentiels de base du Schéma Conceptuel Global

Espace (où ?)	→	Repères spatiaux : GPS, niveau, profondeur,...
Temps (quand ?)	→	Repères temporels : instant, date, durée,...
Composition (quoi ?)	→	Variable(s) analytique(s) décrivant l'objet étudié
Protocole (comment ?)	→	Description du protocole d'analyse : unités, cotes,...
Obtenteur (qui ?)	→	Personne(s), institution(s),...

Toute information issue des agro-écosystèmes se situe à l'intérieur d'un hyper-espace, où les informations spatiales et temporelles font partie d'un ensemble plus vaste de variables élémentaires. Elles sont aussi des éléments de la composition de la même portion d'espace-temps prise en compte.

La composition est l'ensemble des **caractéristiques** que l'on peut attribuer à un objet d'étude de niveau d'observation quelconque, quelle que soit l'analyse. Les informations issues de la composition peuvent être recomposées en ensembles moins analytiques, ou au contraire précisées par des analyses plus fines. La composition peut enfin s'exprimer selon une infinité de niveaux d'observation analytiques allant du global au plus particulier selon des modalités techniques définissables.

La somme des informations possibles constitue donc un hyper-espace complexe et souvent hiérarchisé. Une portion d'espace à un instant donné peut comporter, par exemple, une branche d'arbre dont on analyse la morphologie foliaire dans laquelle on choisit par sélection hiérarchique dans une série juvénile, un pétiole, pour en extraire les protéines, puis on sépare l'acide aminé cystéine dont on isole le carbone pour en caractériser l'isotope ¹⁴C...

Remarquons que cette démarche constitue une cascade de **choix**... On n'a pas étudié les autres branches, ni les autres isotopes du carbone, ni les autres éléments, ni les autres acides aminés... Il y a à la fois une hiérarchie analytique et une sélection entre caractéristiques étudiées ou non étudiées sur la composition d'un objet d'étude.

La distinction claire de ces trois composantes majeures (l'espace, le temps et la composition), s'impose principalement dans le référentiel des données, c'est-à-dire, dans le repère spatio-temporel-compositionnel de chaque donnée acquise dans l'**écosphère** et ses liens aux données ayant la même position dans cet espace-temps.

Deux autres composants liés à la **motivation** de l'acquisition des données : le protocole (enchaîné selon la procédure d'analyse) et l'obteneur (qui a observé ? pour qui ? quels partenaires institutionnels ?...), permettent de référer la responsabilité de l'acquisition de chaque donnée, leur propriété initiale et leurs raisons d'acquisition (problématique, objectif, réalisation, projet, essai, contrat, thèse, ...).

Chaque donnée est ainsi dans un "contenant" analytique (la caractéristique), dans un contenant physique (portion d'espace-temps **indissocié**) inclus dans un contenant général (l'écosphère), et est liée à sa justification sociale initiale (la motivation).

Remerciements

Ce travail est soutenu par le grand projet fédérateur FABATROPIMED (« Services écologiques des légumineuses pour les cycles bio-géochimiques de l'azote et du phosphore et la séquestration du carbone dans les systèmes de cultures céréalières en Afrique et dans le Bassin méditerranéen ») d'Agropolis Fondation, sous la référence ID 1001-009 et le programme de bourses d'études AVERROES fournie par l'UE pour le séjour de Rim Tihinen Maougal à Montpellier.

Références bibliographiques

Bouché M B (1990) *Ecologie opérationnelle assistée par ordinateur*. Ed. Masson, Paris.

Bouché MB, Soto P (1994) The Transdisciplinary Information System ECORDRE : an open manager of the eco-environmental knowledge. In : Anonyme "Data and Knowledge in a changing world". Abstract 14th Int. CODATA Conf., Chambéry, 18-22/09/94, 94-95.

Gavaland A (2011) Point d'avancement du projet ADONIS, Réunion IGEC, Présentation PowerPoint, 21 novembre 2011.

Long G (1974) *Diagnostic phytoécologique et aménagement du territoire. I- Principes généraux et méthodes*. Ed. Masson et Cie, Paris, 1-252.

Soto P (1989) Base de données ECORDRE. Sous-schéma lombriciens. Organisation conceptuelle et modélisation des données lombriciens. Doc. int. ZES, 1-35 + Biblio.

Soto P (1990) Ecologie et environnement : méthodologie technique de normalisation des données dans la conception de la base de données d'ECORDRE. Doc. ZES n°7, 1-306.

Soto P (1997) Géoviticulture : la base de données CEPAGES. Vitiscience. 150^e anniversaire de l'Agro. Montpellier, 21 novembre 1997, pages 17-22.

Soto P, Bouché MB (1993) ECORDRE : Bases de données relationnelles pour toutes données écologiques y compris les activités humaines. In : Anonyme « Les systèmes d'information environnementale ». Ed. ICALPE, Le Bourget du Lac (France), 69-93.

Soto P, Boglio P (1996) La base de données ECORDRE-CEPAGES. Rapport Final, décembre 1996. Réseau Européen de Recherche en Viticulture. U.E. Projet AIR 1728, 1-235.