



**HAL**  
open science

## L'hétérogénéité des données au sein du projet Cosac: Enjeux et solutions

Jean Villerd, Francois Brun, Ophélie Jolys, Pascale Métais, Nathalie G.  
Munier-Jolain, Olivia Pointurier, Fanny Vuillemin

### ► To cite this version:

Jean Villerd, Francois Brun, Ophélie Jolys, Pascale Métais, Nathalie G. Munier-Jolain, et al.. L'hétérogénéité des données au sein du projet Cosac: Enjeux et solutions. Séminaire de restitution à mi-parcours du projet de recherche ANR COSAC, Jan 2017, Paris, France. 85 p., 2017, COSAC un projet pour la conception de Stratégies durables de gestion des Adventices dans un contexte de Changement (climat, pratiques agricoles, biodiversité). hal-01604465

**HAL Id: hal-01604465**

**<https://hal.science/hal-01604465>**

Submitted on 2 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



## Conception de Stratégies durables de gestion des Adventices dans un contexte de Changement (Climat, pratiques agricoles, biodiversité)



**CoSAC** (ANR-14-CE18-0007)

**Séminaire de restitution à mi-parcours  
du projet de recherche ANR COSAC**

***Mardi 31 janvier 2017 de 10 h à 17h et  
mercredi 1 février de 9h à 12h30***

**INRA - 147 rue de l'université à Paris**

# TABLE DES MATIÈRES

TABLE DES MATIÈRES	1
INTRODUCTION	3
CoSAC, un projet pour la Conception de Stratégies durables de gestion des Adventices dans un contexte de Changement (climat, pratiques agricoles, biodiversité)	4
Réchauffement observé et régionalisation des projections climatiques pour la Bourgogne Franche-Comté	7
ÉTUDE DE PROCESSUS BIOPHYSIQUES	10
PRÉSENTATIONS ORALES	10
Capacité des adventices à germer en surface dans diverses conditions	11
Modélisation de l'architecture racinaire dans le projet COSAC : pourquoi et comment ?	17
L'architecture racinaire des adventices en relation avec la disponibilité en azote du sol	19
POSTERS	22
Mise au point d'une méthode de caractérisation morphologique de cultures et d'adventices en vue de modéliser leur aptitude de réponse à la lumière	23
Les transferts d'azote dans le sol à partir d'engrais granulaires dans un contexte d'étude de l'intérêt des apports localisés pour favoriser la régulation des adventices	25
Mise au point d'un dispositif expérimental pour caractériser le tropisme racinaire	27
MODÈLES ET OUTILS	29
PRÉSENTATIONS ORALES	29
Comment représenter la dispersion des adventices dans les modèles : comparaison par analyse de sensibilité	30
Prototypes d'outil d'aide à la décision à partir de FLORSYS pour la gestion intégrée des adventices	32
L'hétérogénéité des données au sein du projet Cosac Enjeux et solutions	34
Modélisation de la résistance aux herbicides dans FlorSys : introduction d'un module permettant de simuler l'évolution conjointe de la résistance liée à la cible et non liée à la cible	36
Évaluation de la valeur pollinisatrice de la flore adventice de différents systèmes de culture	38
POSTERS	41
Évaluation du modèle de dynamique de flore multispécifique FLORSYS. Analyse d'incertitude, domaine de validité, erreur de prédiction et règles d'utilisation	42
Modélisation des effets des systèmes de culture sur la dynamique de l'orobanche rameuse ( <i>Phelipanche ramosa</i> (L.) en interaction avec les adventices	44
Méta-modélisation de l'interception de la lumière dans un couvert cultures: adventices	46
Mieux prédire les conditions hydriques dans les premiers centimètres du sol pour déterminer la germination des adventices	48
Impact de la flore adventice sur la rétention des pesticides : développement d'un nouvel indicateur pour l'évaluation des systèmes de culture	50

ÉVALUER LA PERFORMANCE DES PRATIQUES CULTURALES AU CHAMP	52
PRÉSENTATIONS ORALES	52
Étude de l'effet du travail du sol et des couverts sur les adventices dans des contextes de production varies	53
Analyse des mécanismes impliqués dans la détection automatisée des adventices	56
APPLICATION À LA GESTION DURABLE	58
PRÉSENTATIONS ORALES	58
Comment les conseillers et les agriculteurs raisonnent la gestion des adventices?	59
Les adventices réduisent le rendement et la réduction d'usage d'herbicides aggrave cette perte si elle n'est pas compensée par des mesures préventives et curatives alternatives	62
Proposition d'une méthodologie de conception multi-objective de systèmes de culture à partir de simulations. Application à la gestion des adventices	64
Freins et leviers à la réduction de l'usage d'herbicides en grande culture	67
Guide interactif ECOHERBI : Accompagner les agriculteurs dans les changements de pratiques	70
POSTERS	71
Risque de résistance au glyphosate dans les systèmes de culture à base de maïs et effet sur la production agricole et la biodiversité	72
Mots-clés : modèle, adventice, résistance au glyphosate, système de culture, trait, agroécologie	72
À quelle échelle concilier production agricole et biodiversité : au niveau du paysage ou dans chaque parcelle? Étude appliquée à la gestion des adventices	74
Impact sur le long terme d'une pulvérisation localisée sur la flore adventice. Une étude de Simulation	77
Développement d'un outil d'aide à la décision pour la gestion intégrée des adventices : implication des futurs utilisateurs	79
Projet Inter-instituts Syppre : production de connaissances sur des stratégies de gestion durable des adventices en grande culture	81
Mots-clés :	81
LISTE DES AUTEURS	84

# **INTRODUCTION**

# L'hétérogénéité des données au sein du projet Cosac

## Enjeux et solutions

J. Villerd <sup>1</sup>, F. Brun <sup>2</sup>, O. Jolys <sup>3</sup>, P. Métais <sup>4</sup>, N.M. Munier-Jolain <sup>3</sup>, O. Pointurier <sup>3</sup>, F. Vuillemin <sup>5</sup>

<sup>1</sup> LAE, INRA Univ. Lorraine, F-54518 Vandœuvre-lès-Nancy, France

<sup>2</sup> ACTA, INRA, UMR 1248 AGIR Castanet Tolosan cedex, France

<sup>3</sup> Agroécologie, AgroSup Dijon, INRA, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

<sup>4</sup> ARVALIS Institut du végétal, Biopôle Clermont Limagne, 63360 Saint Beauzire

<sup>5</sup> Terres Inovia, Avenue Lucien Brétignières, 78 850 Thiverval-Grignon

Correspondance : [jean.villerd@univ-lorraine.fr](mailto:jean.villerd@univ-lorraine.fr)

### Résumé

Les données considérées dans le cadre du projet CoSAC sont fournies par les différents partenaires utilisant des protocoles différents, leurs propres référentiels et formatages de données. L'intégration de ces données dans une base de données commune est une étape indispensable pour mener à bien les différentes tâches d'analyse et d'exploitation de ces données capitalisées. L'intégration de données hétérogènes n'est pas un problème trivial et sa résolution réside dans des choix techniques mais aussi dans un consensus d'experts autour des référentiels retenus. Cet exposé offrira un panorama des problématiques liées à l'intégration de données hétérogènes et de leurs solutions. Nous préciserons celle retenue dans le cadre du projet CoSAC.

**Mots-clés** : gestion des données, référentiels, intégration de données, qualité des données, données hétérogènes

### Contexte, enjeux

Les données ont toujours été au cœur de la recherche. La nouveauté est que, depuis une quinzaine d'année, la croissance de la production de données est telle que l'enjeu de l'organisation et de l'analyse des données disponibles est devenu aussi prégnant que la production des données elles-mêmes. Afin de répondre à cet enjeu du *Big Data* et de ne pas être emporté par le « déluge de données », de nouvelles méthodes et bonnes pratiques ont vu le jour. Elles sont finalement en passe de s'imposer comme standard y compris pour les jeux de données de volumes habituels mais très hétérogènes, tels que ceux manipulés dans le cadre du projet CoSAC. En effet elles permettent de résoudre plusieurs problèmes inhérents à la manipulation de données hétérogènes dans le cadre des projets multi-partenaires et issus de différentes expérimentations, non planifiées initialement comme faisant partie d'un même réseau d'essai.

Un premier problème concerne la réutilisation de données anciennes, par exemple dans le cadre d'une méta-analyse, c'est-à-dire une synthèse de connaissance quantitative, ou bien en vue d'être comparées à des résultats de simulations comme c'est le cas dans le cadre du projet CoSAC. Pour pouvoir être réutilisées, elles doivent être accompagnées de données sur les données elles-mêmes – les métadonnées – fournissant les informations relatives au contexte dans lequel elles ont été produites. Par exemple, les métadonnées accompagnant des données floristiques incluent la date, la localisation géographique du relevé, la méthode de prélèvement, des informations sur l'environnement, etc. Les métadonnées assurent ainsi la traçabilité et, donc, la qualité des données. En ce sens elles sont un élément clé de la reproductibilité en recherche.

Au-delà des métadonnées de premier niveau tels que la date et le lieu, les informations plus complexes nécessitent de se rattacher à des référentiels afin de pouvoir être comprises par les futurs réutilisateurs. Ainsi la métadonnée fournissant le type de sol sur lequel le relevé floristique a été effectué a tout avantage à utiliser un référentiel connu pour décrire le type de sol. Ces standards prennent la forme d'ontologies, i.e. de formalisations des concepts d'une discipline (ici la pédologie). L'ontologie identifie de façon unique chaque concept de la discipline (par exemple chaque type de sol) au moyen d'une URI (*Uniform Resource Identifier*). Utiliser une URI pour décrire le type de sol du relevé assure que chaque réutilisateur examinant les métadonnées saura retrouver l'information du type de sol, quelle que soit sa langue, et pourra en parcourant

l'ontologie situer le type de sol considéré parmi les autres concepts de la discipline, traduire le type de sol dans sa langue ou trouver son équivalent dans un autre référentiel. A terme, l'ensemble des données ainsi sémantisées et stockées dans des entrepôts de données constituent le « web des données » (*web of data*). L'intérêt de ces données liées (*linked data*) via les ontologies est qu'elles sont interrogeables par les technologies de web sémantique. Il serait ainsi possible d'interroger l'ensemble des données étiquetées par l'URI d'un certain type de sols ou d'effectuer des requêtes plus complexes.

Ces aspects dépassent largement le cadre de la gestion des données du projet CoSAC. Néanmoins, ces bonnes pratiques – étiquetage par des métadonnées, sémantisation par des ontologies, archivage dans des entrepôts – sont bénéfiques y compris pour des données utilisées localement dans le cadre d'un projet car elles assurent leur traçabilité et leur réutilisabilité future.

## Les données dans CoSAC

Les données mobilisées dans le projet sont issues d'une compilation de différentes expérimentations passées chez différents partenaires (INRA, Arvalis, Terres Inovia,...) pour répondre à des questions très variées. Ainsi, ces données sont caractérisées par différents niveaux d'hétérogénéité : (i) hétérogénéité syntaxique : différents formats de stockage (tableur, base de données relationnelle) ; (ii) hétérogénéité structurelle : différents schéma de données, i.e. différents découpages, différents noms de colonnes ; (iii) hétérogénéité sémantique : absence de référentiel ou de vocabulaire partagés. Un groupe de travail regroupant les différents partenaires fournisseurs de données a été constitué. Nous exposerons les différentes solutions identifiées et étudiées. Le choix de la solution retenue a dû tenir compte à la fois de contraintes techniques, de contraintes temporelles et des moyens humains disponibles pour intégrer l'ensemble de ces données. Ceci explique que le choix ne se soit pas porté vers la solution la plus complète au vu des enjeux énuméré précédemment (mais aussi la plus coûteuse à mettre en œuvre). La solution retenue prend la forme d'un fichier structuré (format csv) utilisant un format d'identifiant unique et des référentiels permettant d'assurer l'interopérabilité des données. Toutefois la prise en considération de ces enjeux a permis d'identifier les différents impératifs correspondant aux différentes réutilisations des données : méta-analyse, entrées de modèles, etc.

## L'exemple des données stockées dans Agrosyst

Nous illustrerons l'étape effective d'intégration au travers des données Dephy stockées dans le SI Agrosyst et montrerons comment l'utilisation d'une base de données assure la pérennité de l'interopérabilité entre Agrosyst et le modèle FLORSYS.

## Analyse des données

Les méthodes de synthèse de connaissance quantitative comme l'analyse des réseaux d'essais ou la méta-analyse, basée sur la valorisation de la littérature doivent nous permettre d'extraire des connaissances de ces données hétérogènes. Par ailleurs, la modélisation profitera de ces données variées pour des améliorations dans les formalismes, mais aussi pour l'évaluation de la qualité de prédiction de ces outils.

## Conclusion

Nous concluons par les leçons tirées par cette expérience, notamment par la nécessité de prendre en considération la question de la gestion des données très en amont. Ainsi, de nombreux appels à projets recommandent aujourd'hui l'établissement d'un « data management plan » dans les premiers mois du projet. Collectivement, on peut également se poser la question du minimum de métadonnées nécessaires pour une mise en commun et leur utilisation dans les travaux de modélisation.

## Références bibliographiques

Conseil scientifique de l'Inra, 2012. Rapport du groupe de travail sur la gestion et le partage des données