

# Standardiser les données expérimentales pour faciliter l'innovation dans le domaine des bio solutions : le projet CASDAR STAR.

Florence Amardeilh<sup>1</sup>, Arnaud Charleroy<sup>2</sup>, Baptiste Darnala<sup>1</sup>,  
Xavier Delpuech<sup>3</sup>, Catherine Roussey<sup>2</sup>, Frédéric Salvi<sup>4</sup>

<sup>1</sup> Elzeard, Bordeaux, France

<sup>2</sup> INRAE, MISTEA, Montpellier, France

<sup>3</sup> Institut Français de la Vigne et du Vin (IFV), Le Grau-du-Roi, France

<sup>4</sup> Terres Inovia, Montpellier, France

florence.amardeilh@elzeard.co, arnaud.charleroy@inrae.fr, baptiste.darnala@elzeard.co,  
Xavier.DELPUECH@vignevin.com, catherine.roussey@inrae.fr, f.salvi@terresinovia.fr

## Résumé

*Cet article présente le projet STAR sur la construction d'un standard pour partager les données d'expérimentation agronomique évaluant les biosolutions pour 3 types de cultures : oléagineux, vigne et maraîchage.*

## Mots-clés

*modèle de données, expérimentation agronomique, biosolution*

## Abstract

*This article presents the STAR project which aims to build a standard for sharing agronomic experiment data for biosolutions evaluation. STAR works on 3 types of crops : oilseeds, vines and market gardening.*

## Keywords

*data model, agronomic experiment, biosolution*

## 1 Introduction

Accélérer l'innovation et l'adaptation des systèmes de productions agricoles est de plus en plus nécessaire dans un contexte de mutations accélérées par le changement climatique et la montée des enjeux environnementaux. Ainsi, les filières sont incitées à développer des mesures préventives ou alternatives à l'utilisation de produits phytosanitaires de synthèse pour anticiper et préparer les évolutions réglementaires à venir. Parmi ces mesures, le recours au biocontrôle (c'est-à-dire les solutions de protection des cultures d'origine naturelle et présentant de faibles impacts sur l'environnement et la santé), fait l'objet d'une politique volontariste de la part des pouvoirs publics : stratégie nationale du déploiement du biocontrôle issue de la loi EGALIM de 2018, feuille de route NeoTerra en 2022 pour la région Nouvelle-Aquitaine, le Grand défi « Biocontrôle et biostimulants » qui prend la suite du consortium « Biocontrôle ». Malgré ces nombreuses initiatives, l'adoption des biosolutions sur le terrain reste freinée par leur efficacité parfois partielle et variable selon leurs conditions d'utilisation [13]; [7] [15].

Cette variabilité de l'efficacité des biosolutions au champ est encore mal comprise et il est nécessaire de consolider nos connaissances pour favoriser l'adoption en toute sécurité des biosolutions par les agriculteurs. Pour être efficace, cette consolidation des connaissances doit s'appuyer sur :

- Une description fine et détaillée des conditions de mise en œuvre des biosolutions, incluant l'itinéraire technique, la modalité d'application, le système de production, le type de sol, les conditions climatiques... Cette description du contexte est bien souvent incomplète dans les études publiées [8].
- Le partage et la mutualisation des données produites par les différents acteurs, y compris les agriculteurs eux-mêmes, pour pouvoir analyser les conditions de réussite et d'échec des solutions dans des situations variées. Toutes les données sont utiles et c'est leur combinaison qui permettra de faire un saut qualitatif dans notre compréhension des performances des biosolutions.

Le projet STAR a pour objectif de lever les freins à un partage et une mutualisation des données expérimentales (collectées par des expérimentateurs selon des protocoles définis sur des dispositifs contrôlés) et techniques (c'est-à-dire les données décrivant le ou les itinéraires techniques mis en œuvre sur l'exploitation). Pour cela, le projet vise à proposer un ou des modèles de données adaptables associés à des vocabulaires contrôlés pour obtenir des données interoperables. Le projet vise à co-construire avec les expérimentateurs des outils simples et pratiques à mettre en œuvre pour valider et produire des données standardisées. Il inclut une action de sensibilisation, formation et accompagnement des acteurs.

## 2 Travaux antérieurs

Pour standardiser les données d'expérimentations agronomiques plusieurs initiatives ont été mises en oeuvre depuis plus de 10 ans pour produire des modèles de données parta-

gés. Nous compléterons cet état des lieux avec les modèles de données sur le suivi des exploitations agricoles.

## 2.1 Les modèles de données d'expérimentations agronomiques

ICASA [19] est un des premiers standards disponibles pour partager les données d'expérimentation agronomique. Il s'agit d'un format de données tabulaires qui décrit l'ensemble des paramètres et des types de mesures nécessaires au suivi des expérimentations agronomiques et à leurs analyses statistiques. Ce standard n'a pas été construit par des informaticiens et donc le modèle de données sous jacent n'est pas explicite.

Le standard ISA [17, 16] est un modèle de données open source générique pour partager les données d'expérimentations agronomiques. Il décrit une expérimentation agronomique à l'aide des entités « Investigation » (le contexte du projet), « Study » (une expérimentation) et « Assay » (une mesure analytique). La structure hiérarchique de ce modèle permet la représentation d'expérimentation utilisant une technologie ou une combinaison de technologies. Ainsi, il est possible de détailler la mesure analytique en décrivant les caractéristiques de l'échantillon, la technologie et les types de mesure.

La construction du standard de « Minimal Information About Plant Phenotyping Experiment » (MIAPPE) a été lancée en 2014, avec l'aide d'experts d'infrastructures et d'instituts européens comme Elixir, Emphasis, INRAE, WUR, iBet, IPK, EBI et IPG PAS [11]. Il répond au besoin de publication et de réutilisation des données grâce à une liste de contrôle qui formalise et documente les métadonnées minimales d'une expérimentation agronomique. MIAPPE reprend les entités du standard ISA, sauf que l'entité « Assay » est supprimée au profit de l'entité « Observation Unit », la portion de parcelle agricole qui est la plus petite unité d'observation d'un essai agronomique. De plus, MIAPPE est plus riche que ISA il contient les entités « Factor », « Factor Value », « Growth Facility », etc... La version 1.1 a donné naissance à une ontologie Plant Phenotype Experiment Ontology (PPEO) [10].

INRAE, avec l'UMR MISTEA à Montpellier, s'est intéressé depuis 2002 à la problématique d'intégration des données agronomiques complexes et hétérogènes, et a développé une suite logicielle innovante open-source OpenSILEX. OpenSILEX permet de développer des systèmes d'information composé de plusieurs couches logicielles, basé sur les principes du Web sémantique et du Big Data [9]. La suite OpenSILEX s'appuie sur une ontologie d'application privée intitulée « Experimental Scientific Objects Ontology » (OESO). OESO a pour objectif de contextualiser les ensembles de mesures stockées dans des fichiers. Elle a été créée en 2012 en s'inspirant des éléments de l'ontologie des expérimentations scientifiques EXPO [18]. EXPO est une ontologie générique datant de 2006 qui utilise SUMO. Elle n'est pas orientée sur la structuration des données mais plus sur la documentation d'une expérimentation. OpenSILEX est déjà utilisé de façon opérationnelle pour la gestion de données dans différents projets.

Concernant certaines cultures, le projet VITIS DATA CROP, piloté par l'IFV de 2021 à 2023, a initié des travaux de standardisation sur un cas d'usage en innovation variétale. Au cours de ce projet, un modèle de données pour la description des expérimentations agronomiques en vigne a été proposé [4]. Ce modèle de données est accompagné par un guide d'explication accessible en ligne et propose un fichier de saisie Excel. Ce projet a aussi permis de créer un centre de ressources [2] dédié à la filière viticulture-œnologie qui contient des documents (protocoles expérimentaux), des référentiels (variables, matériel végétal), des applications et des liens utiles pour la communauté [4].

## 2.2 Les modèles de données agricoles

Des standards pour les données de traçabilité des exploitations agricoles existent aussi, en enregistrant les données de l'itinéraire technique, c'est-à-dire la combinaison logique et ordonnée des interventions qui permettent de contrôler la culture à différentes échelles (parcelle, sous-parcelle, plante, etc) et l'environnement associé pour obtenir une récolte. L'itinéraire permet ainsi de relier un ensemble de données telles que le mode de production, les données d'implantation, les semences et plants utilisés, les différentes tâches qui jalonnent la production (fertilisation, irrigation, protection, traitement, désherbage, etc.), les matériels utilisés, le temps de travail, les données de récoltes et de rendement, etc.

En 2003, Le projet « Gestion des Informations de l'Exploitation Agricole » (GIEA) conduit par l'Assemblée Permanente des Chambres d'Agriculture (APCA) réunit l'ensemble des acteurs de la sphère agricole désireux d'échanger des données agricoles répondant à la fois aux besoins de management technique et environnemental des exploitations et aux demandes des partenaires économiques et institutionnels (filiales, traçabilité, PAC...)[12]. Ce projet a défini des dictionnaires métiers favorisant l'interopérabilité des applications informatiques ou statistiques d'évaluation, de comparaison, de prévision de risque. Il a donné lieu à un site Web et une documentation UML des données nécessaires à la gestion des exploitations.

Ces données d'exploitation, et en particulier les données d'itinéraires techniques, font l'objet d'une proposition de standard au sein de l'association AgroEDI Europe avec le format DAPLOS pour Data Plot Sheet. DAPLOS permet ainsi l'interopérabilité entre les outils de traçabilité à la ferme, comme MesParcelles, SMAG, et le système d'information AgroSyst, qui gère les données du réseau DEPHY du Plan Ecophyto. DAPLOS devrait toutefois à moyen terme être remplacée par le futur standard ISO Smart Farming [6], proposée par l'ISO (Organisation internationale de normalisation). Une limite identifiée est que ces standards sont privés et leur accès est payant, ce qui est un frein à leur diffusion et leur réutilisation par une communauté élargie.

En open data, Elzeard a développé l'ontologie « Crop Planning and Process Production Ontology » (C3PO) en partenariat avec INRAE, EducAgri et d'autres acteurs du monde agricole pour standardiser la représentation des itinéraires

techniques, quelle que soit la filière végétale concernée [3]. Ce modèle de données propose une représentation des itinéraires techniques. Il a aussi l'avantage de relier les itinéraires techniques à un ensemble de facteurs extérieurs à la culture comme le contexte pédoclimatique, la région, le parcellaire, les bioagresseurs, l'équipement matériel pour offrir une collecte et une analyse des données multifactorielles.

### 2.3 Les vocabulaires contrôlés

Des vocabulaires contrôlés et partagés sont nécessaires pour développer l'interopérabilité des données en fournissant l'ensemble des valeurs possibles à utiliser pour renseigner les modèles de données. Ces vocabulaires contrôlés peuvent prendre la forme d'un modèle SKOS ou d'un module d'une ontologie plus importante.

En vigne, la Vitis ontology [5] propose une liste de variables phénotypiques permettant de caractériser les traits du végétal suivant 4 composantes : la caractéristique, l'entité observée, la méthode de mesure et l'unité. Dans le cadre du précédent projet CASDAR VITIS DATA CROP, cette ontologie a été intégrée dans un centre de ressources ouvertes et partagées pour la filière [2]. Néanmoins, en l'état actuel cette ontologie ne permet pas d'identifier sans ambiguïté des variables proches. Par exemple, une « hauteur de plante » est plus générique qu'une « hauteur de plante à floraison » : ces deux variables partagent la même méthode, la même unité mais l'une est plus précise car l'entité est dans un stade phénologique particulier : la floraison. Nous devons ainsi faire évoluer la description des variables pour préciser leurs liens entre elles. Pour l'observation des plantes, les stades phénologiques sont fondamentaux. Les échelles phénologiques et leurs stades, déjà formalisés sous la forme d'une extension d'un modèle SKOS publiée sur le Web [14], pourront être remobilisés dans le cadre du projet STAR.

Concernant la description des variables mesurant les impacts des stress biotiques observées dans les expérimentations de biosolutions nous utiliserons les ressources associées à l'ontologie des bioagresseurs BAGO [1] pour décrire quels sont les phénomènes observés. L'enjeu est de mettre à niveau cette liste de variables, de la partager et de l'aligner avec les thésaurus existants.

## 3 Problématique

Avec l'avènement du numérique et d'Internet, la production de données s'est démultipliée : chaque acteur, de l'agriculteur au chercheur en passant par le fournisseur de services, produit toujours plus de données. Le partage et l'exploitation de ces données sont un enjeu stratégique pour des filières agricoles, en particulier pour améliorer les connaissances existantes sur l'efficacité des biosolutions et favoriser son adoption par les agriculteurs. Si des standards existent pour organiser et décrire les données expérimentales, ces standards ont un taux d'adoption faible dans la communauté technique (et scientifique) française. Il est probable que la plupart des expérimentateurs ne connaissent pas leur existence, et même quand ils les connaissent, le coût d'adoption (temps d'apprentissage et

de mise en œuvre) est jugé comme trop important au regard des bénéfices perçus. Par exemple, un des frein identifié est l'absence de traduction en langue française des vocabulaires proposées. Il est aussi nécessaire d'adapter le modèle de données à la thématique traitée pour prendre en compte les pratiques des expérimentateurs, en différenciant les éléments indispensables à la compréhension des données, et ceux qui sont optionnels. Enfin, modèle de données et vocabulaires doivent être accessibles et facilement utilisables par les expérimentateurs eux-mêmes : l'enjeu est alors d'intégrer cette approche de standardisation des données dans des interfaces utilisateurs ergonomiques, répondant aux besoins des utilisateurs-expérimentateurs.

## 4 Approche envisagée

Le projet s'articule en plusieurs actions complémentaires :

1. L'action 1 vise à co-construire avec les experts métiers un modèle de données d'expérimentation associé à un modèle de description des variables observées. Les modèles proposent une structuration et organisation des données auxquels on associe des vocabulaires contrôlés (sous forme de modèle SKOS étendu) pour proposer les valeurs possibles des éléments du modèle. Ces modèles sont adaptés aux cas d'usages identifiés. Fort du travail déjà réalisé sur la construction de ressources sémantiques de la vigne dans de précédents projets (D2KAB, VITIS DATA CROP, ...), un des objectifs du projet STAR est de tester la robustesse des ontologies et des ressources associés sur plusieurs systèmes de cultures, en cultures pérennes (vigne) et annuelles (maraîchage, protéagineux, oléagineux).
2. L'action 2 a pour objectif de co-construire les outils associés, permettant d'implémenter les livrables de l'action 1 dans des interfaces simples et répondant aux besoins des utilisateurs, afin de faciliter et de rendre attractif la démarche de standardisation pour les expérimentateurs : aide à la recherche de la bonne variable, autocomplétion grâce aux vocabulaires contrôlés (listes de valeurs possibles), fonctions de validation et d'exploration des données...
3. L'action 3 permettra d'animer la démarche, en proposant des formations et des ateliers pour la mise en pratique des outils construits en action 2. Ainsi, les productions des actions 1 et 2 seront testées sur le terrain afin de faciliter leur adoption et de mettre en place un pipeline allant de la collecte des données jusqu'aux analyses.

## 5 Conclusion

Le projet STAR a pour objectif de faciliter le partage et la mutualisation des données produites par différents acteurs (expérimentateurs, agriculteurs, fournisseurs...) pour analyser les conditions de réussite et d'échec des biosolutions dans des situations variées. La stratégie proposée est de rendre interopérables et de partager les données pro-

duites pour éviter la création de silos de données propriétaires, en s'appuyant sur les technologies du Web sémantique. Ces technologies seront intégrées dans des interfaces utilisateurs ergonomiques pour limiter le coût d'adoption pour les utilisateurs finaux. Il s'agit de rendre autonomes les expérimentateurs dans la gestion et l'exploration de leurs données, en utilisant des outils répondant à leurs besoins et en s'appuyant sur des ressources (formation, ressources en lignes).

## Références

- [1] F Amardeilh, S Bernard, R Bossy, M Courtin, M Hirschy, P Larignon, Catherine Roussey, and N Sauvion. Une ontologie pour modéliser les bioagresseurs des plantes. In Haïfa Zargayouna, editor, *IC2024 : 35es Journées francophones d'Ingénierie des Connaissances*, number 2024 in Plate-Forme Intelligence Artificielle (PFIA), pages 8–13, La Rochelle, France, July 2024. AFIA-Association Française pour l'Intelligence Artificielle, AFIA-Association Française pour l'Intelligence Artificielle.
- [2] Arnaud Charleroy, Xavier Delpuech, and Pascal Neveu. Vitioeno, un centre de ressources numériques pour la production de données scientifiques FAIR. In *44th World Congress of Vine and Wine*, volume 68, page 01030, Cadix, France, May 2023. Organisation internationale de la vigne et du vin, EDP Sciences.
- [3] Baptiste Darnala, Florence Amardeilh, Catherine Roussey, Konstantin Todorov, and Clement Jonquet. C3PO : a crop planning and production process ontology and knowledge graph. *Frontiers in Artificial Intelligence*, 6 :1187090, October 2023.
- [4] Xavier Delpuech, Vincent Dumas, Jean-Yves Cahurel, Laure Gontier, Marion Claverie, Arnaud Charleroy, Viviane Bécart, Romain Lacroix, Eric Duchêne, Nathalie Ollat, Joseph Tran, and Catherine Roussey. A standard for sharing data from vineyard experiments. *BIO Web of Conferences*, 68(2023) :01031, November 2022.
- [5] Eric Duchêne and Cyril Pommier. The Vitis Ontology : sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) for consistent and complete data description through biologist friendly ontologies. In *The First Annual Meeting of COST Action CA17111 INTEGRAPPE 2019*, Chania, Greece, 26-28 March 2019, Greece, March 2019.
- [6] STRATEGIC ADVISORY GROUP REPORT ON SMART FARMING. Iso/ tmb/ sag sf : Final report with recommendations. Technical report, ISO, 2023.
- [7] Ashebir T Godebo, Naomi Marie J Wee, Christopher K Yost, Fran L Walley, and James J Germida. A meta-analysis to determine the state of biological control of aphanomyces root rot. *Frontiers in Molecular Biosciences*, 8 :777042, 2022.
- [8] Keiji Jindo, Travis L Goron, Paloma Pizarro-Tobías, Miguel Ángel Sánchez-Monedero, Yuki Audette, Ayodeji O Deolu-Ajayi, Adrie van der Werf, Misghina Goitom Teklu, Moshe Shenker, Cláudia Pombo Sudré, et al. Application of biostimulant products and biological control agents in sustainable viticulture : A review. *Frontiers in Plant Science*, 13 :932311, 2022.
- [9] Pascal Neveu, Anne Tireau, Nadine Hilgert, Vincent Negre, Jonathan Mineau-Cesari, Nicolas Bricchet, Romain Chapuis, Isabelle Sanchez, Cyril Pommier, Brigitte Charnomordic, Francois Tardieu, and Llorenç Cabrera Bosquet. Dealing with multi-source and multi-scale information in plant phenomics : the ontology-driven Phenotyping Hybrid Information System. *New Phytologist*, 221(1) :588–601, 2019.
- [10] Evangelia A Papoutsoglou, Daniel Faria, Daniel Arend, Elizabeth Arnaud, Ioannis N Athanasiadis, Inês Chaves, Frederik Coppens, Guillaume Cornut, Bruno V Costa, Hanna Ćwiek-Kupczyńska, et al. Enabling reusability of plant phenomic datasets with miappe 1.1. *New Phytologist*, 227(1) :260–273, 2020.
- [11] Cyril Pommier, Guillaume Cornut, Thomas Letellier, Célia Michotey, Pascal Neveu, Manuel Ruiz, Pierre Larmande, Paul J. Kersey, Hanna Cwiek Kupczynska, Pawel Krajewski, Frederik Coppens, Richard Finkers, Marie-Angélique Laporte, Daniel Faria, Célia M. Miguel, Inês Chavez, Anne-Françoise Adam Blondon, and Bruno Costa. Data standards for plant phenotyping : MIAPPE and its implementations. In *26. Plant and Animal Genome Conference (PAG XXVI)*, page 24 slides, San Diego, Californie, United States, January 2018.
- [12] P. Poyet and Thomas Brun. GIEA : gestion des informations de l'exploitation agricole - Vers des concepts et un langage communs pour les partenaires économiques et institutionnels de la profession agricole. *Ingénieries eau-agriculture-territoires*, pages 167–175, 2003. [Departement\_IRSTEA]GEAPA [TR1\_IRSTEA]41 - PRECITECH / COPAIN.
- [13] Thomas Presseccq. *Développement d'outil d'aide à la décision pour favoriser l'usage du biocontrôle microbien*. Theses, Université d'Avignon, December 2023.
- [14] Catherine Roussey, Xavier Delpuech, Florence Amardeilh, Stephan Bernard, and Clement Jonquet. Semantic Description of Plant Phenological Development Stages, starting with Grapevine. In Emmanouel Garoufallou and María-Antonia Ovalle-Perandones, editors, *Communications in Computer and Information Science (CCIS)*, volume 1355 of *Metadata and Semantic Research. MTSR 2020*, pages 257–268, Madrid, Spain, December 2020. Springer International Publishing. The final authenticated version is available online at [https://doi.org/10.1007/978-3-030-71903-6\\_25](https://doi.org/10.1007/978-3-030-71903-6_25).
- [15] M Sandroni, Erland Liljeroth, T Mulugeta, and Erik Alexandersson. Plant resistance inducers (pris) : pers-

pectives for future disease management in the field. *CABI Reviews*, pages 1–10, 2020.

- [16] Susanna-Assunta Sansone, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, et al. Towards interoperable bioscience data. *Nature Genetics*, 44 :121–126, 2012.
- [17] Susanna-Assunta Sansone, Philippe Rocca-Serra, Alejandra Gonzalez-Beltran, David Johnson, and ISA Community. *ISA Model and Serialization Specifications 1.0*, October 2016.
- [18] Larisa N Soldatova and Ross D King. An ontology of scientific experiments. *Journal of the royal society interface*, 3(11) :795–803, 2006.
- [19] Jeffrey W White, LA Hunt, Kenneth J Boote, James W Jones, Jawoo Koo, Soonho Kim, Cheryl H Porter, Paul W Wilkens, and Gerrit Hoogenboom. Integrated description of agricultural field experiments and production : The icasa version 2.0 data standards. *Computers and electronics in agriculture*, 96 :1–12, 2013.

## Remerciements

Les auteurs tiennent à remercier le projet CASDAR STAR « Standardiser les données expérimentales et techniques pour faciliter leur réutilisation et accélérer l’innovation et le développement agricole : application aux travaux sur les bio-solutions » financé par FranceAgrimer.