

Seminaire résidentiel INRAE Semantic Linked Data, 11-14 Octobre 2021 à Sète

*Le comité d'organisation lance un **appel à communication** aux ingénieur.e.s et chercheur.e.s des CATIs INRAE qui souhaitent présenter leurs travaux de conception et réalisation de systèmes d'information et d'applications utilisant les méthodes et les technologies du Web Sémantique. objectif faire un snapshot de ce qui se passe à inrae et voir si on ne peut pas faire du data linking de maniere generalisée*

Publication et annotation des Bulletins de Santé du Végétal

Stephan Bernard, Robert Bossy, Anna Chepaykina, Catherine Roussey

Nous présentons le projet de construction de navigateur augmenté pour les bulletins officiels d'alertes agricoles français, appelés Bulletins de Santé du Végétal (BSV).

Le projet ANR "Des Données aux Connaissances en Agronomie et Biodiversité" (D2KAB) (www.d2kab.org) illustre comment la science des données sémantiques contribue au développement d'applications agricoles innovantes. L'objectif de D2KAB est de créer un cadre pour transformer les données d'agronomie et de biodiversité en connaissances décrites sémantiquement, interopérables, exploitables et ouvertes.

Pour construire un tel cadre, nous nous appuyons sur des ressources sémantiques pour décrire nos données (par exemple, des thésaurus ou des référentiels métiers) et les publier en tant que données ouvertes liées. Nous utilisons le portail web AgroPortal (<http://agroportal.lirmm.fr>) pour trouver, publier et partager des ressources sémantiques.

L'un des scénarios agricoles de D2KAB consiste à construire un navigateur web augmenté pour les bulletins officiels d'alertes agricoles français, appelés Bulletins de Santé du Végétal (BSV). Ces bulletins sont actuellement disponibles au format pdf sur les sites web des DRAAF. Une première étape a été de collecter les bulletins et de transformer leur contenu textuel en fichier HTML. Ensuite, chaque bulletin a été annoté sémantiquement en identifiant dans le texte les mentions d'éléments d'une ressource sémantique. Les annotations sont produites par plusieurs chaînes de traitement automatique de la langue utilisant les outils développés par l'équipe Bibliome. Des annotations de différents types ont été construites pour enrichir le corpus et sont publiées sur le web de données liées.

L'objectif de cette présentation est de montrer comment les référentiels métiers intéressants pour l'annotation des BSV ont été transformés et mis à disposition sur Agroportal. Deux types de référentiels métiers ont été publiés dans le projet D2KAB: le thésaurus des cultures intitulé FrenchCropUsage et des échelles phénologiques (de la vigne, des céréales et de quelques légumes). Le langage RDF choisi pour structurer ces référentiels est Simple Knowledge Organisation System (SKOS). Ce langage permet d'aligner les éléments des ressources sémantiques entre eux.

Nous présenterons aussi les différentes étapes de la publication du corpus des BSV sur le web de données liées et comment les annotations ont été structurées en utilisant un réseau d'ontologies basé sur Web Annotation Data Model, PROV et Semantic Sensor Network.