

Vitioeno, un centre de ressources numériques pour la production de données scientifiques FAIR

Vitioeno, a web resource center for the production of FAIR research data

Arnaud Charleroy¹, Xavier Delpuech² et Pascal Neveu¹

¹INRAE

²Institut Français de la Vigne et du Vin, Le Grau-du-Roi, France

Résumé. INRAE et l'IFV ont mis en place un centre de ressources numériques ouvertes et librement accessibles, avec l'objectif d'aider et de faciliter la structuration et la qualification des données selon les principes FAIR. Ce centre de ressource pour l'aide à la gestion des données « Vitioeno » est disponible en ligne, en français et en anglais, à l'adresse <https://vitioeno.mistea.inrae.fr/resource/app/>. Il recense et donne un accès ouvert à différents types de ressources numériques existantes, en particulier des référentiels (matériel végétal et variables expérimentales), des documents téléchargeables de type protocoles, et des applications ou liens utiles.

Abstract. INRAE and the IFV have set up an open and freely accessible web resource center, with the aim of helping and facilitating the structuring and qualification of data according to FAIR principles. This “Vitioeno” resource center is available online, in French and English, at <https://vitioeno.mistea.inrae.fr/resource/app/>. It identifies and provides open access to different types of existing digital resources: a list of plant material, of experimental variables, documents such as protocols, and useful links or apps.

1 Introduction

Avec l'accélération du changement climatique, la viticulture européenne est amenée à adapter rapidement ses modes de productions à ce nouveau contexte avec son lot de sécheresses, maladies, canicules et événements extrêmes. Les leviers existants, comme l'innovation variétale, l'irrigation, les techniques agronomiques comme l'ombrage ou la gestion de la canopée, viennent bousculer des pratiques séculaires de production. L'enjeu est alors d'être en mesure de mobiliser rapidement et efficacement des données sur l'impact de ces nouvelles pratiques sur la vigne et les vins, afin de mettre en place un pilotage des vignobles véritablement « data-driven ».

Pour atteindre cet objectif dans des délais toujours plus courts, il est nécessaire de distribuer l'effort de collecte des données sur l'ensemble des partenaires de la filière : recherche, développement, professionnels... Tout l'enjeu est ensuite d'être en mesure de partager et de rendre compréhensible les données produites afin de pouvoir les analyser au bénéfice de différentes

communautés ou acteurs. Il s'agit de mettre en place de manière opérationnelle une gestion des données selon les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable). Or la gestion de données agronomiques suivant les principes FAIR est très difficile en raison de leur hétérogénéité, d'autant plus si elles proviennent de différents acteurs. Pour répondre à ces difficultés, INRAE et l'IFV ont mis en place un centre de ressources numériques ouvertes et librement accessibles, avec l'objectif d'aider et de faciliter la structuration et la qualification des données selon les principes FAIR.

L'objectif de cet article est de présenter ce centre de ressources et ses perspectives.

2 Matériel et méthodes

2.1 Un logiciel open-source

Le centre de ressources s'appuie sur un logiciel open-source développé par INRAE, OpenSILEX [3].

Le logiciel open source OpenSILEX permet de construire des systèmes d'information pour gérer les données expérimentales. OpenSILEX est une approche scientifique ouverte. Ce logiciel est basé sur une ontologie et met en œuvre les principes FAIR pour des données faciles à trouver, accessibles, interopérables et réutilisables. Son architecture permet de gérer de grands volumes de données hétérogènes complexes. OpenSILEX est dérivé en plusieurs implémentations correspondant à des communautés scientifiques spécifiques :

- PHIS (phénotypage de plantes à haut débit),
- Sixtine (unités expérimentales de plantes à l'INRAE),
- Silex-VitiOeno (vin et viticulture),
- Alfis (données sur la fermentation alcoolique),
- ENVIBIS (dépollution)
- Sun'Agri (agrivoltaïsme).

Ce logiciel a donc été choisi en raison de sa souplesse de configuration et sa conformité aux principes FAIR [4].

2.2 Identification des attentes des utilisateurs

Un atelier autour du vocabulaire expérimental s'est tenu le 19 janvier 2023 lors du séminaire organisé dans le cadre du projet VITIS DATA CROP (<https://vignevin.quarto.pub/seminaire-data-2023/>). Ce projet a pour objectif de proposer un ensemble cohérent d'outils et des méthodes pour améliorer l'interopérabilité, le partage et l'ouverture des données.

L'objectif de cet atelier était de présenter une maquette du centre de ressource et de recueillir les attentes des futurs utilisateurs pour définir les évolutions à mettre en place. L'atelier s'est déroulé sur environ 1h30, en deux sessions avec une quinzaine de personnes. Après une présentation de la maquette du centre de ressources, les participants ont été invités à réfléchir sur une série de questions :

- Quels référentiels utilisez-vous ?
- Quel cas d'utilisation imaginez-vous dans le centre de ressource ?
- Quelle serait la fonctionnalité la plus intéressante pour vous ?

Après une phase de réflexion individuelles, où les contributions de chacun ont été écrites et collectées, un temps de partage et de discussion a suivi.

3 Résultats et Discussion

3.1 Attentes des utilisateurs

Certains participants n'utilisent pas vraiment (ou n'en n'ont pas conscience) de référentiels, mais d'autres citent des référentiels de matériel végétal ou des ontologies existantes. Parfois, ce sont des normes qui sont citées, comme l'AFNOR pour l'analyse sensorielle.

Concernant les cas d'usage, les utilisateurs ont souligné l'importance d'avoir un outil qui regroupe des référentiels existants actuellement dispersés, que ce soit sur le matériel végétal, sur les sols, les échelles phénologiques...mais aussi sur les maladies par exemple (Fig. 1).

Le partage de protocoles fait aussi partie des fonctionnalités attendues.

La présence d'œnologues et de spécialistes de l'analyse sensorielle a permis aussi de souligner l'importance d'avoir une communauté élargie et transversale pour faire le lien entre l'amont et l'aval.

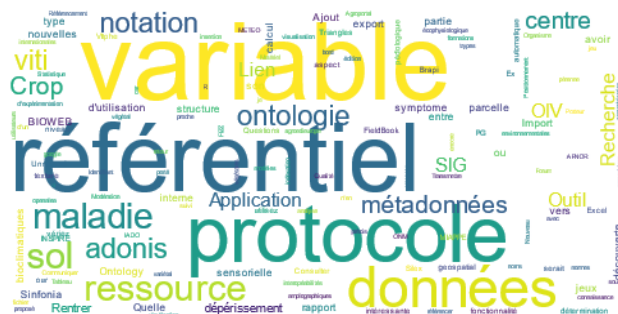


Figure 1. Nuage de mots issu de l'atelier du 19 janvier 2023.

La demande des utilisateurs est aussi forte en fonctionnalités permettant de traiter et analyser les données, par exemple les fonctions suivantes ont été citées :

- Visualiser et éditer des rapports,
- Calculs d'indicateurs bioclimatiques
- Outils d'harmonisation (liste de variables en téléchargement, transformation et alignement de référentiels...)

L'interopérabilité avec Adonis, un outil de collecte de données expérimentales développé par INRAE a été citée plusieurs fois.

3.2 Sélection et mise à disposition des ressources numériques

Le centre de ressources a été enrichi à partir de ressources numériques existantes. Ces ressources sont de différentes natures :

- Des référentiels sous forme de liste comme pour le matériel végétal ou les variables ;
- Des documents téléchargeables : protocoles, guide, jeux de données ;
- Des applications ;
- Une liste d'ontologies ;
- Des liens vers des ressources tierces.

Une première version a été mise en ligne à l'adresse <https://vitioeno.mistea.inrae.fr/resource/app/>.

Chaque ressource est identifiée par un URI (Uniform Resource Identifier), une chaîne de caractère qui a pour but de l'identifier sur un réseau (ici le Web) de manière permanente et non ambiguë. Le centre de ressource s'appuie sur des technologies du web sémantique, qui facilitent l'ouverture et l'interopérabilité des données.

Le référentiel matériel végétal est issu de la 'Base de données du Réseau Français des Conservatoires de Vigne [6]. Cette base de données regroupe les informations du CRB Vigne de Vassal-Montpellier et du Réseau Français des Conservatoires de Vigne. Elle répertorie les informations décrivant les espèces, variétés, clones et accessions utilisés par le RFCV qui regroupe 36

partenaires régionaux impliqués dans la sauvegarde des ressources génétiques et la sélection. Plus de 180 conservatoires sont actuellement implantés dans toutes les régions viticoles du pays. Le centre regroupe des identifiants ainsi que des informations basiques dédiés aux utilisateurs et applications qui souhaite utiliser un référentiel de matériel végétal.

La liste des variables (Fig. 2) est issue de la crop ontologie Vitis 356 [1].

Les protocoles proposés sont sous forme de document au format pdf (Fig. 3), dont par exemple des extraits du guide de l'OIV [2].

Une liste d'applications est proposée, avec généralement des liens externes, mais aussi certaines applications qui sont exécutables directement sur le centre de ressources, comme l'application WaLIS qui permet de simuler le bilan hydrique du sol à partir des données météorologiques.

Une liste d'ontologies qui présentent les vocabulaires et les différents concepts qui sont explicités et qui peuvent servir de référence pour annoter les données provenant des différentes étapes qui mène la vigne à devenir du vin.

Un ensemble de site web sur la gestion de données vient enrichir ce centre de ressource.

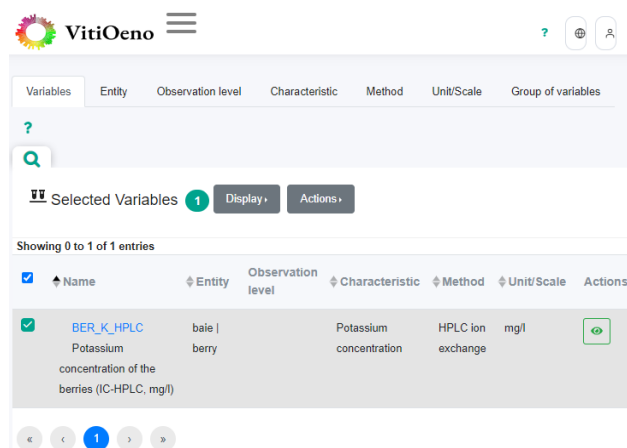


Figure 2. Capture d'écran du formulaire variable du Centre de ressources VitiOeno.

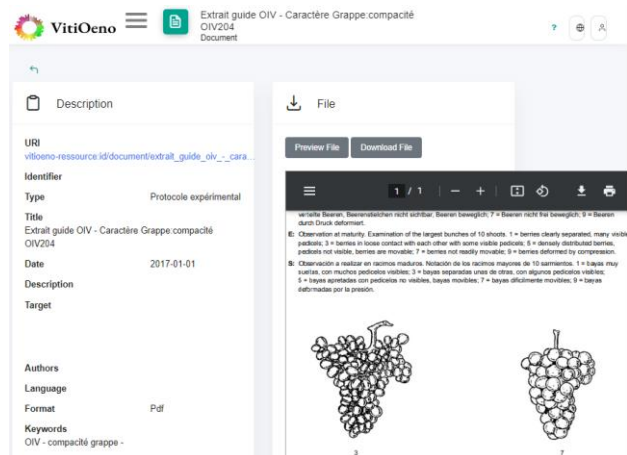


Figure 3. Capture d'écran d'un document protocole disponible sur le centre de ressource.

3.3 Une interface de programmation (API)

Une API, permet à un ordinateur de demander une information à un autre ordinateur, par internet. Le centre de ressource propose une API pour requêter toutes les ressources numériques présentées. Il implémente aussi BrAPI [5] à la version 1.2. C'est une spécification d'API normalisée pour la communication de données sur la sélection végétale. BrAPI permet de communiquer facilement des données. BrAPI est une spécification d'API de service Web RESTful standardisée pour la communication de données sur la sélection végétale.

Cette API permet l'interopérabilité machine-machine du centre de ressource.

4 Perspectives

Notre objectif est d'enrichir ce centre de ressources, et de pouvoir proposer une palette complète de ressources et d'outils pour faciliter le partage et l'interopérabilité des données, à la fois pour l'expérimentation au vignoble et en cave. L'enjeu est maintenant de construire et d'animer une communauté d'expérimentateurs pour rendre ce centre de ressource attractif et continuer à l'enrichir. A terme, ce centre de ressources pourrait aussi intéresser d'autres acteurs de la filière (fournisseurs de services ou structures de production).

5 Conclusions

Un centre de ressources a été construit pour partager des ressources numériques ouvertes afin de faciliter le partage et l'interopérabilité des données (référentiels, protocoles, applications...).

Ce centre est accessible à la fois via une interface web pour les humaines et pour les machines *via* une API.

Ce travail a été réalisé dans le cadre du projet VITIS DATA CROP, avec le soutien du ministère de l'Agriculture et de l'Alimentation et la contribution financière du compte d'affectation spéciale développement agricole et rural (CASDAR).

Références

1. E. Duchêne, C. 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, 2019)
2. OIV, Liste des Descripteurs OIV pour les Variétés et Espèces de Vitis (Organisation Internationale de la Vigne et du Vin, Paris, France, 2008)
3. P. Neveu, A. Tireau, N. Hilgert, V. Negre, J. Mineau-Cesari, N. Brichet, R. Chapuis, I. Sanchez, C. Pommier, B. Charnomordic, F. Tardieu, L. Cabrera-Bosquet, New Phytol **221**, 588 (2019)
4. M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. Da Silva Santos, P. E. Bourne, J.

- Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, 't Hoen, Peter A. C., R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *Scientific Data* **3**, 160018 (2016)
5. P. Selby, R. Abbeloos, J.E. Backlund, M. Basterrechea Salido, G. Bauchet, O.E. Benites-Alfaro, C. Birkett, V.C. Calaminos, P. Carceller, G. Cornut, B. Vasques Costa, J. D. Edwards, R. Finkers, S. Yanxin Gao, M. Ghaffar, P. Glaser, V. Guignon, P. Hok, A. Kilian, P. König, J. E. B. Lagare, M. Lange, M.-A. Laporte, P. Larmande, D. S. LeBauer, D.A. Lyon, D.S. Marshall, D. Matthews, I. Milne, N. Mistry, N. Morales, L. A. Mueller, P. Neveu, E. Papoutsoglou, B. Pearce, I. Perez-Masias, C. Pommier, R.H. Ramírez-González, A. Rathore, A. M. Raquel, S. Raubach, T. Rife, K. Robbins, M. Rouard, C. Sarma, U. Scholz, G. Sempéré, P.D. Shaw, R. Simon, N. Soldevilla, G. Stephen, Q. Sun, C. Tovar, G. Uszynski, M. Verouden, *Bioinformatics (Oxford, England)* **35**, 4147 (2019)
6. Réseau Français des Conservatoires de Vigne, https://bioweb.supagro.inra.fr/collections_vigne/Home.php, accessed May 17, 2023