

# A standard for sharing data from vineyard experiments

Xavier Delpuech<sup>1</sup>, Vincent Dumas<sup>2</sup>, Jean-Yves Cahurel<sup>1</sup>, Laure Gontier<sup>1</sup>, Marion Claverie<sup>1</sup>, Arnaud Charleroy<sup>2</sup>, Viviane Bécart<sup>3</sup>, Romain Lacroix<sup>3</sup>, Eric Duchêne<sup>2</sup>, Nathalie Ollat<sup>2</sup>, Joseph Tran<sup>2</sup>, and Catherine Roussey<sup>2</sup>

<sup>1</sup>Institut Français de la Vigne et du Vin (IFV), Le Grau-du-Roi, France

<sup>2</sup>Inrae, Paris, France

<sup>3</sup>Institut Rhodanien, Orange, France

**Abstract.** To facilitate the sharing and interoperability of data collected by many different experimenters and structures, a standardized description of the data acquisition context has been set in the form of a data schema. This data schema defines the entities and the attributes to describe them. This data schema is available online, under an open license CC0 1.0 Universal, with a user's guide.

## 1 Introduction

With the acceleration of climate change, European viticulture must rapidly adapt its production methods to a new context including severe droughts, new diseases, heat waves and extreme events. Technical levers such as varietal innovation, irrigation, new agronomic techniques such as shading or canopy management are shaking up age-old production practices.

The characterization of the impact of these new practices on the vine and the wines requires a significant effort of data acquisition. In France, this effort is currently distributed among different research and development organizations to share costs and to consider the diversity of production conditions (soil, climate, plant material, technical itinerary, etc.). Data can then be produced by multiple actors (public research organizations, technical institutes, chambers of agriculture, associations, cooperatives...), in very heterogeneous forms, and therefore difficult to share and pool.

To meet these challenges, many initiatives are emerging to create shared databases (accessible through a dedicated information system), where each player is invited to make its data available for the benefit of all. For example, the OSCAR information system was developed by INRAE and IFV to monitor the deployment of downy and powdery mildews tolerant varieties in France. OCESAR was developed by the Centre du Rosé (Vidauban, France) to monitor new varieties in South-East France. SilexPorteGrefte was developed by INRAE to collect and share data on rootstock experiments. Other information systems exist with various thematic and geographic perimeters (see paragraph 2.2).

This strategy partially addresses these challenges by standardizing the data integrated across the information system, but (i) a large portion of the data remains outside

these information systems and (ii) these information systems do not share (or only partially) the same data schemas (or data model) nor the same vocabularies to populate these schemas.

For example, to search for an experiment on a variety, it will be necessary to search for the attribute called “variété” in the OCESAR system, whereas it will be called “cépage” in OSCAR or “greffon” in SilexPorteGrefte. On the other hand, different vocabularies may be used, or in other words, the values that these attributes are not the same across the information systems: for example, to name the same variety, we will find the term “CARIGNAN B” in the OCESAR system and the term “Carignan blanc” in the SilexPorteGrefte system.

Our objective was therefore to set a data schema describing the context of experimental data acquisition in the vineyard. This data schema consists of a structured list of entities, with their definition and the attributes to describe them. For each attribute, the schema will define the values that can be used (term from a list for example). Ultimately, the goal is to facilitate the interoperability of datasets produced by the vine and wine scientific and technical community, thanks to this open and shared schema.

## 2 Material and methods

### 2.1 Definitions

In this article, we will use different terms that are defined here:

- A data schema or data models describe the organisation of the data storage base of an information system.

- A data schema is composed of entities linked together by relationships. For example, a “field” is an entity in the OSCAR data schema.
- Attributes describe each entity. For example, “planting density” is a descriptive attribute for a field.
- Vocabularies are the possible values of the attributes defined in the schema. For example, the values of the “variety” attribute can be a specific list of varieties. This list of varieties is a vocabulary.

## 2.2 A workshop

First, we organized a seminar in January 2023 to bring together a technical and scientific community around the issues of interoperability and data sharing in the vine and wine sector (<https://vignevin.quarto.pub/seminaire-data-2023/>). A workshop on the terms in use within the vine and wine community was organized. This workshop allowed (i) to raise awareness of the vocabulary used, and (ii) to collect a set of terms and to group them by context (vineyard, cellar, agronomic...). A selection of the most used terms was enriched with definitions.

## 2.3 Comparative analysis of data structuring in existing systems

In addition, we analysed some existing information systems in France to identify the entities and vocabularies used. We selected existing information systems, associating several partners with the objective of sharing data in the field of French viticultural research:

- SilexPorteGrefte: information system that aims to gather existing experimental data (in nurseries and vineyards) on rootstocks used in viticulture (only the vineyard part was analysed).
- Oscar: an information system associated with a collaborative network of fields in production of new varieties tolerant to mildew and powdery mildew, planted by wine growers and monitored by technicians. The OSCAR database is hosted on IFV's Epicure platform.
- OCESAR: information system centralizing information from the regional observation network on new grapevine varieties in the south of France.
- Sinfonia: IFV information system for the management of vine and wine experimental data.
- VitisExplorer: information system whose objective is to pool experimental data acquired on genotypes from recent varietal creation.

This review was completed by similar approaches conducted during networking projects, such as the Recap&Dep project (a PNDV project), as well as by the analysis of the experimental data collection tool Adonis (INRAE).

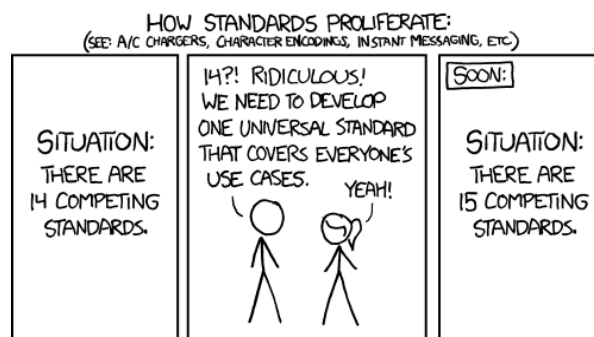
## 2.4 Building on existing standards

Once this review of the existing system was completed, we compared it to existing standards for sharing experimental data in agronomy. Indeed, the challenge is

not to create a new standard (Fig. 1), but to understand why the community not or hardly uses the existing ones, and to suggest ways to facilitate their adoption.

We have identified and analysed the following standards:

- ICASA [1]: ICASA is a standard available in tabular form to describe the set of parameters and types of measurements needed to monitor agricultural experiments.
- MIAPPE [2]: MIAPPE (Minimal Information About Plant Phenotyping Experiment) is an open and shared standard to harmonize data from plant phenotyping experiments. This standard includes general entities describing an experiment (project, ...), the experimental design, and the environmental conditions. The MIAPPE standard is itself largely inspired by and interoperable with the generic ISA (Investigation-Study-Assay) model.
- ISA [3]: ISA is a generic open source data model, built on the metadata categories “Investigation” (the context of the project), “Study” (an experiment) and “Assay” (an analytical measurement). The extensible and hierarchical structure of this model allows the representation of studies using a technology or combination of technologies, focusing on the description of its experimental metadata (i.e., sample characteristics, technology and measurement types, sample-data relationships).



**Figure 1.** Avoid the proliferation of standards ! Source: <https://xkcd.com/927>

## 3 Results and Discussion

### 3.1 Commons entities and terms

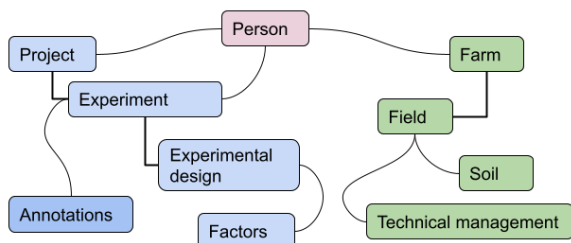
The workshop at the seminar showed consistency in the vocabulary used by the experimenters, but also a need to share a definition for about 30% of the terms used. Participants also identified synonyms for some terms. For example, the French term “cep” (vinestock) has the synonyms “plante”, “plant”, “shouche”, “pied” and “individu”.

On the other hand, the entities used to describe the context of data acquisition were identified in the existing information systems. These entities are listed and defined in Table 1.

**Table 1.** Entities describing the experimental context.

| Entity               | Definition   |
|----------------------|--|
| Person               | Human person   |
| Project              | A research project is the general framework for research activities, including one or more experiments, depending on the project objectives. It is generally the object of a contract or a grant agreement between the laboratory performing the research activities and funding agencies providing financial support. |
| Experiment           | A planned and organized inquiry designed to test a hypothesis, answer a question, or discover new facts.   |
| Experimental design  | The set of rules and procedures by which the treatments are assigned to experimental units.  |
| Experimental factor  | The object of a study is to ascertain the impact of one or more factors on the biological material. Thus, a factor is, by definition, a condition that varies between observation units, which may be biotic (pest, disease interaction) or abiotic (treatment and cultural practice) in nature.                       |
| Annotation           | Annotations are free comments, which provide additional synthetic information to facilitate the understanding and interpretation of the data.  |
| Farm                 | A farm is the structure or legal entity whose objective is to organize production on the plot. It is characterized by a unique management and its own means of production.   |
| Field                | A field is a plot of land with the same plant population to which the same technical management is applied.  |
| Soil                 | Soil is an environmental material, which is primarily composed of minerals, varying proportions of sand, silt, and clay, organic material such as humus, gases, liquids, and a broad range of resident micro and macro organisms.  |
| Technical management | The technical management is the logical and ordered combination of techniques implemented on a plot of land to obtain a production.  |

Entities are linked together by relationships (Fig. 2).



**Figure 2.** Relations between the entities of the standard. In green the specific entities of the agronomic experimentation domain.

It should be noted that these entities belong to the generic domain of agricultural experimentation. It is the selection of attributes and the associated vocabularies that are

specific to the context of vineyard experimentation.

### 3.2 Description of entities

Depending on the information system, the entities have more or less attributes and sometimes could even be absent (Table 2). For example, OSCAR and OCESAR do not use project, experiment and experimental factor entities. Indeed, these information systems are designed to collect data on networks of grapevine fields in production, on which no experimentation is being carried out.

On the other hand, information systems dedicated to the management of experimentation data such as SilexPorteGreffe, Sinfonia or VitisExplorer include these entities and their description. Other entities, such as the vineyard or the field, are found in all information systems (Table 2).

**Table 2.** Number of attributes per entity according to the information system.

| Entity               | Silex PorteGreffe | OSCAR | OCESSAR | Sinfonia | Vitis Explorer |
|----------------------|-------------------|-------|---------|----------|----------------|
| Person               | 4                 | 3     | 4       | 3        | 3              |
| Project              | 14                | 0     | 0       | 13       | 13             |
| Experiment           | 11                | 0     | 0       | 12       | 12             |
| Experimental factor  | 5                 | 0     | 0       | 4        | 4              |
| Farm                 | 6                 | 22    | 9       | 4        | 4              |
| Field                | 24                | 22    | 17      | 15       | 4              |
| Technical management | 0                 | 17    | 0       | 15       | 0              |
| Sool                 | 2                 | 0     | 2       | 12       | 0              |
| Annotation           | 0                 | 0     | 0       | 3        | 3              |

Generic entities, such as the project, experimentation, experimental design are part of the elements generally described in existing standards. For their definition and their descriptive attributes, we based ourselves on their implementation in the OpenSilex software [4], which is used in SilexPorteGreffe, VitisExplorer and Sinfonia.

The work carried out thus focused on the selection and definition of the attributes useful for describing the context of an experimentation in the vineyard. This schema was completed by vocabularies integrated as lists within the schema. For example, the “production system” can take a value from the following list: “conventional”, “organic farming” or “biodynamic”. For the variety or the rootstock, the vocabulary available online in the Vitioeno web resource center was recommended: <https://vitiogeno.mistea.inrae.fr/resource/app/germplasm>. All the proposed descriptive attributes are available online, in an open and shared repository (see Sect. 3.4).

### 3.3 Mapping with existing standards

Entities as the project, the description of the experiment and the factors studied already exist in previous standards. *A priori*, the alignment of entities and vocabularies to fill in the attributes of the entities used in the vine and wine community should not pose any problem.

A search on the data warehouse <https://entrepot.recherche.data.gouv.fr/> with the query "(vitis OR vigne OR grapevine)" identifies 51 datasets. None of these datasets is based on an existing standard. This result seems to support the analysis of the non-adoption of existing standards by the French vine and wine scientific community. Alignment with an existing standard is therefore not a prerequisite for maintaining interoperability with existing datasets, but nevertheless it remains a relevant exercise to validate and compare the proposed data schema.

In particular, the ICASA standard [1] is used by the agronomic and modeling communities, and proposes attributes for agronomic management practices, treatments, environmental conditions, and crop measurements. ICASA is applicable to any field experiment or agricultural production situation. Because of this generic scope, the age of this standard (10 years) and its integration into data sharing tools we chose to map first our data schema with the ICASA standard data dictionary, available online: <https://agmip.github.io/ICASA.html>.

An equivalence was found for 38% of the attributes.

### 3.4 Diffusion and access

A GitHub directory has been set up, under open license CC0 1.0 Universal. It contains a xlsx file lists the metadata fields, with their description and expected format (see <https://github.com/vignevin/vitiscadacrop>). This basic file is then declined in different forms:

- a user's guide, which explains the entities and their fields to collect, available online (see [https://vignevin.github.io/standard\\_guide/](https://vignevin.github.io/standard_guide/))
- a xlsx template file to facilitate its appropriation by the experimenters.

## 4 Discussion

This first work allowed us to share and define vocabularies, and to define schemas for describing the context of experimental data acquisition in the vineyard.

If the entities used by the vine and wine community are well reflected in the existing standards, one of the first obstacles to their use is the absence of a vocabulary translated into the native language of the experimenters. On the other hand, some very specific terms of the wine context must complete the vocabulary. The extension of this vocabulary to the domain of oenological experimentation and sensory analysis will make it possible to cover all vine and wine experimental data. Beyond the description of the data acquisition context, the next step will be to work on a way to associate raw

data, *i.e.* data collected during measurements performed on the experimental objects.

Finally, it is likely that the cost of learning and implementing existing standards is too high for experimenters in relation to the perceived added value. The challenge is to facilitate the use of this standard by proposing simple tools to (i) standardize the data and (ii) exploit them. This standard could be a necessary building block for the implementation of a decentralized, interoperable and searchable data architecture according to FAIR principles [5]. Demonstrating the benefit of using the standard will be all the easier and more obvious if it is used by a large part of the experimenters: the co-construction of the tools associated with the support of the future users through training and online tutorials is necessary. In this way, the support of the vine and wine Vitioeno resource center <https://vitoeno.mistea.inrae.fr/resource/app/> will allow the standard and the associated tools to be disseminated.

As a corollary, this success will require strong and continuous coordination of the community. This standard must also be supported by a trusted structure whose legitimacy is obvious to the community. In France, we think that the collective approach initiated in the VITIS DATA CROP project can be the governance framework of this standard. If some metadata is very specific to the French context, for example, the production area, this work would also benefit from going beyond national borders. The International Organisation of Vine and Wine could fully play its role in this internationalization process.

## 5 Conclusions

A data schema has been proposed to describe the vineyard experiments.

The associated vocabulary has been defined in French.

This data schema has been analysed with respect to existing standards and a first mapping with ICASA has been achieved.

This work must be continued and strengthened to ensure its appropriation by the experimenters.

This work was carried out within the framework of the VITIS DATA CROP project, which aims to propose a coherent set of tools and methods to improve interoperability, sharing and openness of data in the vine and wine sector, with the support of the French Ministry of Agriculture and Food, with the financial contribution of the special allocation account for agricultural and rural development (CASDAR).

## References

1. J.W. White, L.A. Hunt, K.J. Boote, J.W. Jones, J. Koo, S. Kim, C.H. Porter, P.W. Wilkens, G. Hoogenboom, *Computers and Electronics in Agriculture* **96**, 1 (2013)
2. E.A. Papoutsoglou, D. Faria, D. Arend, E. Arnaud, I. N. Athanasiadis, I. Chaves, F. Coppens, G. Cornut, B.V. Costa, H. Ćwiek-Kupczyńska, B. Droesbeke, R. Finkers, K. Gruden, A. Junker, G.J. King, P.

- Krajewski, M. Lange, M.-A. Laporte, C. Michotey, M. Oppermann, R. Ostler, H. Poorter, R. Ramı Rez-Gonzalez, Ž. Ramšak, J.C. Reif, P. Rocca-Serra, S.-A. Sansone, U. Scholz, F. Tardieu, C. Uauy, B. Usadel, R.G.F. Visser, S. Weise, P.J. Kersey, C.M. Miguel, A.-F. Adam-Blondon, C. Pommier, *New Phytologist* **227**, 260 (2020)
3. S. Susanna-Assunta, R.-S. Philippe, F. Dawn, M. Eamonn, T. Chris, H. Oliver, F. Hong, N. Steffen, T. Weida, A.-Z. Linda, B. Kimberly, B. Tim, B. Lydie, B. Gully, C. Brad, C. Tim, C. Lee-Ann, C. Jay, D. Sudeshna, d. D. Antoine, d. M. Paula, D. Ian, E. Scott, T.E. Chris, J.F. Mark, G. Pascale, G. Jack, G. Carole, L.G. Julian, J. Daniel, K. Jos, H. Lee, H. Kenneth, H. Henning, J.H.S. Shannan, L. Alain, L. Shaoguang, M. Stephen, M. Annette, M. Emily, R. Dorothy, R. Magali, E.S. Caroline, A.S. Catherine, S. Christoph, T. Anne, W.-J. Bryn, W. Katherine, X. Ioannis, H. Winston, *Nature Genetics* **44** (2012)
  4. 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)
  5. 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)