e: hello@collabriculture.com
w: collabriculture.com

# First Steps Towards An Open Vineyard Data Model

## Introduction

The Australian viticulture community is at a crossroads. Advancements in digital technology are making it easier to capture and manage data, and to gain insights as it is used. Growers, agronomists, and AgTech providers are exploring opportunities to make better decisions and improve their operations through the application of new technologies. But we're in the early days of this revolution. People from all sectors of our community are reporting that they face obstacles when seeking to make the most of these technological advancements, principally around data sharing.

#Collabriculture has been having wide-ranging conversations with people working with spatial data in the Australian viticulture community, including growers, agronomists, and data & tech providers, to gain insight into how they use data, the issues they face, and what they see as key priorities. We wanted to know what they are excited about, and where they are frustrated — a key barrier for the industry stems from difficulties around data sharing, so do they see an open-access, industry-standard geospatial data model helping the community? Will building community consensus around agreed data structures and standards provide new opportunities to accelerate innovation in the AgTech/precision viticulture space?

This paper is largely based on their insights, and is intended to provoke further discussion on the data model concept.
Ultimately, an industry-standard geospatial data model for viticulture will only be as robust as the community collaboration that builds it, so please read on, and share widely in your networks.

## Why do we need a data model?

What community needs could be addressed by open-access, industry-standard geospatial data model? From our conversations with people in Australian viticulture, a few key themes emerged:

### Data portability & interoperability

A commonly expressed frustration is that it is difficult to move data between systems and platforms: a grower may be running multiple AgTech systems, &/or a GIS or database system, but then finds that these systems "don't talk to each other". There can be a number of underlying causes of this, including business drivers for AgTech and data providers, lack of available standards, and the difficulty of resolving the current impasse.

While there are instances where AgTech or GIS database systems appear to be actively designed to stymie cross-platform communication in an effort to ensure customer retention, technical problems

are a more typical barrier. Each existing system has been developed in relative isolation, with its own data model driven by vendor and application, so even when AgTech providers do want to make it easier to share data between platforms they can find it difficult in the absence of an established standard. Without an accepted industry standard, it's not always clear what data structures might work for the end user, and the existing complexities and the various user needs are too fragmented to effectively resolve.

A shared data model would provide standardisation: defining the necessary entities and relationships in a consistent way, so that AgTech providers can implement it either as a core building block of new systems or as an interchange format between existing systems. By doing this, a shared data model would facilitate the movement of data between systems and formats.

## Building consensus

Inevitably, when innovation happens and technical solutions are developed in isolation from each other, there will be differences in how people conceptualise and describe things. One application might partition vineyard blocks into "zones", another into "sub-blocks", and yet another into "patches". By collaboratively developing and examining potential data models, we can build consensus on the components that make up a vineyard operation — so that there is less room for confusion and misunderstanding when we communicate. This is another aspect of standardisation.

## Provide a resource

Building community consensus around agreed data structures and standards could accelerate innovation in the AgTech/precision viticulture space by allowing AgTech providers to build applications they felt confident would meet community needs. As it stands, each provider must invest time and resources in defining their own model from scratch, and then validating it against customer requirements. The existence of a well-defined, readily accessible, consensus-based standard model would give application developers a head start; freeing up time, energy, and investment capital to solve more specific problems.

## Mitigating risk

If we, as a community, don’t solve this problem, someone else will — and we might not like their solution. There is a risk that the industry-standard geospatial data model viticulture settles on will be paywalled, locking out some users; not fully standardised, or standardised on the wrong things; or optimised for vendors not growers.
What we have now is a unique opportunity to embed the priorities and logic of growers into the data models vendors build up from, and that ultimately benefits everyone.

## Digital literacy

As the community increases knowledge and capacity, we become better equipped to avail ourselves of opportunities to use data and technology to solve problems, and create a broader audience for

technical solutions. The potential of GIS, database, and AgTech technology could be more fully explored with increased digital literacy — generating improved outcomes for all.

A shared data model could facilitate the expansion of digital literacy in two ways. First, as a process, the creation of a shared data model would engage community members, directly developing skills. Secondly, the existence of a standardised model will assist technical people become more proficient and confident with aspects of practice outside their primary area of expertise, by offering practical and reliable guidance for implementation.

# What could a data model look like?

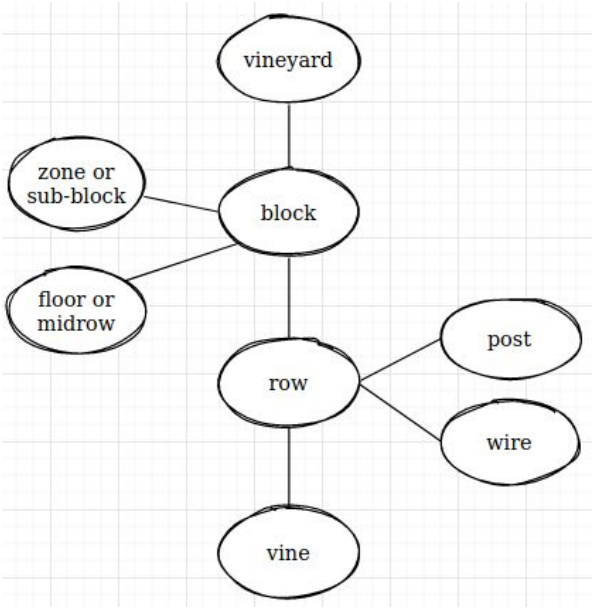
## What is a data model?

### Conceptual models

A conceptual model, or a conceptual schema, is

> "a high-level description of informational needs underlying the design of a database. It typically includes only the main concepts and the main relationships among them. Typically this is a first-cut model, with insufficient detail to build an actual database."[1]

In our case, a conceptual model would likely begin with the key physical components of a vineyard (*vineyard-block-row-vine*), and identify related entities and the relationships between them.
For example, a block may contain many rows, a row may have many posts and wires.




[1] https://en.wikipedia.org/wiki/Conceptual_schema

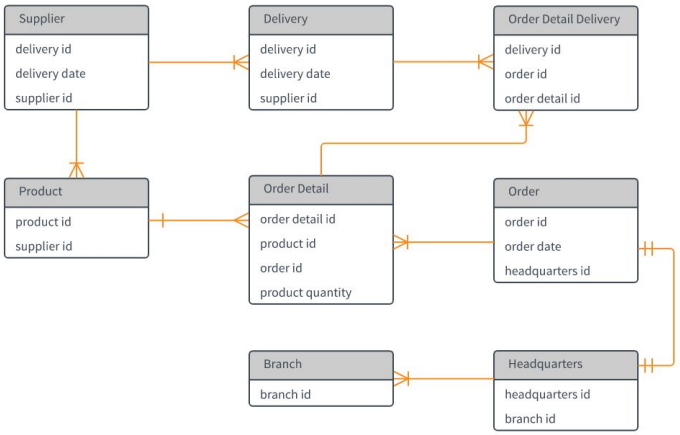
## Logical models

A logical model, or a logical schema, is

> "a data model of a specific problem domain expressed independently of a particular database management product or storage technology but in terms of data structures such as relational tables and columns, object-oriented classes, or XML tags"[2]

In a logical model, we would go beyond identifying the key entities, and start to describe them in terms of their attributes. For example, we might define the specifics of a vineyard, eg. an owner, a street address, a date of establishment, and a local government area. Each of these attributes could be further specified in terms of a column name, a data type, and a range of possible values. For example, a date of establishment could be named "*date_established*", should be a "*date*" data type, and could have the constraint that the date is in the past. A local government area could be named "*lga*", and could be required to select from a list of existing local government areas.

This kind of model could be expressed as an Entity Relationship Diagram, which has some similarity to the conceptual model diagram concept, but includes attributes, and may be more specific about the types of relationships, in database terms (eg. one-to-many).




In a logical model, we would go beyond identifying the key entities, and start to describe them in terms of their attributes,  building from the conceptual model. In practice, this can be accomplished using any one of multiple languages and frameworks, each suitable for different application development environments.  SQL is a very common, standards-based language favoured for use in database systems[3]. A simplified example of how a model could be defined in SQL:

[2] https://en.wikipedia.org/wiki/Logical_schema

[3] It should be noted that while SQL is a common standard, most database systems use their own "flavour" of SQL and implementing each may require some minor modifications to the code.

```sql
CREATE TABLE public.vineyard (
    id uuid NOT NULL,
    "name" varchar NOT NULL,
    "owner" varchar NULL,
    street_address varchar NULL,
    date_established date NULL,
    CONSTRAINT vineyard_pk PRIMARY KEY (id)
);
```

This usage of SQL is known as Data Definition Language, or DDL, and is specifically used for defining a database structure in terms of things like tables, relationships, and indexes. A full model would consist of multiple tables and relationships.

A benefit of expressing the model in code, is that if it's in the same framework as the target technology it can be quickly and easily implemented. Modifications can also be made relatively simply, and tracked using familiar development patterns. It could be useful to provide DDL-like code in additional formats, such as JSON, or for specific ORM frameworks like Flask or Ruby on Rails.

## Who will use it, and what for?

The data model will potentially be of use to anyone working with vineyard data.

- Those involved in implementing and managing systems — such as a GIS, a database, or a field app — will be able to use the specification to design their own data models, to ensure compatibility with others following the same standard.
- Growers could reference the standard when requesting data from their providers.
- Data providers, such as government agencies, could use the standard to design their open data feeds.
- Application developers could use the standard to guide their own projects, by implementing it directly, or by using it to guide how data is shared with other applications.

If there is an agreed standard, with broad uptake by the viticulture community, it will make it easier to design data systems, and remove an impediment to sharing data between platforms.

## What next?

> "How do we get on that road... the biggest challenge we're facing is that everybody's waiting on everybody else to make those first few steps, and nobody's really getting anywhere because no one is taking the first steps."

#Collabriculture exists so that the Australian viticulture community can start moving together towards the opportunities that advancements in digital technologies are creating.
To do that, we're going to need a road: digital infrastructure like accurate digital maps and interoperable spatial data. That is not something that can be built by outsiders, not if it's going to work for us on our terms. It will need community involvement from the ground up, to make a model that makes sense for the work we do now and the work we want to be doing in the future.

The Australian viticulture industry has a strong sense of community, and we're joined in this project by technical experts and enthusiasts from industries as diverse as mining, defence, and city planning. They're as keen to learn from us as we are from them.

None of us, alone, have all or even most of the pieces needed to make this work.
But together, when we share our knowledge and our experience, we're going to be on the road to something of significant practical value.

## About us (and you)

#Collabriculture is an ongoing project initially made possible by funding from a Primary Industries and Regions South Australia (PIRSA) grant. Beginning with a series of six workshops, we also have an online site and other tools to capture and share learnings from the workshops, and to provide online space for discussions and resource exchange.
#Collabriculture is the venue and the facilitators: you are the community and the experts.
Nobody knows your industry, your work, and your needs like you do.
We need doers and builders and growers. People who ask questions, people who see connections, curious people. Stubborn people who won't budge on what matters, and people who know when to compromise. Traditionalists, and early adopters. So that we can find the most complete answers together, and build a digital infrastructure for viticulture's future that will serve us all.

Some key questions #Collabriculture will be working through as the project unfolds:

- Does the viticulture community actually need a data model?
- Would you use a data model in your own work? What are the obstacles?
- This approach is centred on identifying things that can be mapped in space, ie. geospatial objects, and the relationships between them. Should geospatial information be at the centre? If not, what are alternative approaches?
- If we make the model too complex, it may be too difficult to implement; if we make it too simple, it may be too general to actually serve a tangible purpose. What is the sweet spot - how will you know if it's either too simple or too complex for your needs?
- How important is it that the community retains ownership of the data model? How can we achieve this?
- How can we ensure the ongoing participation of the community in maintaining the data model, as viticulture and technology evolve? How do we keep the model from becoming obsolete?
- Thinking beyond conceptual and logical models, what other resources would make this a useful asset? How far do we take this?