Here come the robots, but what do we do with the data?

Hans Loder, 2021 SA Scholar

November 2023

Nuffield Australia project number 2111

Supported by:

Wine Australia

Scholar contact details

Hans Loder

Penley Estate Pty Ltd

McLean Road, Coonawarra

08 7078 1853

hans@penley.com.au

www.penley.com.au

Social media

In submitting this report, the Scholar has agreed to Nuffield Australia publishing this material in its edited form.

NUFFIELD AUSTRALIA Contact Details

Nuffield Australia

Email: enquiries@nuffield.com.au

Address: PO Box 495, Kyogle, NSW 2474

Executive Summary

With talk of a data driven future and the function data will play in providing both efficiency and oversight to agricultural enterprises, this report hopes to deliver clarity to wine grape businesses around defining data, how its generated, how it is stored and what it can deliver.

Jason Strong of Meat and Livestock Australia said:

“Data has no value until you use it.”

Similarly, while many ask the question of “who owns the data” the first step is ensuring that a business itself properly owns and secures its data.

This report looks to explain what constitutes “data” and how it can be best managed. There are several types of data, and it can occur in different forms. How can a business be confident it is looking at the right data, let alone knowing how far along it is on the journey to achieving full data maturity?

This report concludes that every business is different, and the way data is managed will vary. Whether data needs to be hosted in a custom platform, or if the question relates to whether value is being realised from a proprietary system, each business will need to consider the needs of its own production methodology.

This report asks: “Here come the robots, but what do we do with the data?”

The likely cost of poor data management is outlined within this report, and it seems tantalizing that problems can be solved if data is done well. However, the field of data science has many facets and data management requires skill.

This report speaks of the importance of data engineering, and this will likely be a first critical step for most businesses, but so too the analytics, machine learning, artificial intelligence (AI) and automation capabilities that follow.

All rely on quality data.

Table of Contents

Executive Summary **3**
Table of Contents **4**
Table of Figures **5**
List of Tables **5**
Foreword **6**
Travel Itinerary **7**
Acknowledgments **9**
Abbreviations **10**
Objectives **11**
Chapter 1: Introduction **12**
Chapter 2: What data is being collected? **13**
2.1 The industry **13**
2.2 Collecting data **13**
2.3 Know the data being collected **14**
What is data maturity? **14**
Chapter 3: What's changed? **16**
3.1 Types of data and the broader data landscape **16**
3.2 Defining big, large, and small data? **18**
3.3 FAIR data **19**
3.4 8 Pillars of agtech and farm data **19**
Chapter 4: How is data stored? **20**
4.1 Cost of storage **21**
4.2 Rise of the data aggregators **21**
Chapter 5: Data ownership opportunities and risks **23**
5.1 The hero **23**
5.2 Investing in data **23**
5.3 The contract **24**
Opportunities; is anonymity more important than ownership? 24
5.4 Farm data principles **25**
Chapter 6: Managing data – current and future opportunities **26**
6.1 What's the opportunity? **27**
Conclusion **29**
Recommendations **30**

References **31**

Appendices **33**

Appendix 1: AxisTech & Koolin Farms – The 8 Pillars **33**

Appendix 2: Farm Data Principles – extracts Edition 1 ..... **35**

Appendix 3: FAIR Data **37**

Table of Figures

Figure 1: The AVEVA Welcome Wall: London Unified Operations Centre, March 2022 8

Figure 2: David Zadow's impressive irrigation solution, monitoring, and rainfall records 9

Figure 3: Mart Verstappen collects data throughout the season to track inputs through to results. This allows him to tell a seasonal story and demonstrate the difference plant nutrition has made to a crop 14

Figure 4: Considering the power of data with Ami Caragata at the Conexus Cultivator, Regina Canada 15

Figure 5: Vineyard naming conventions agreed by consultation with industry representatives 17

Figure 6: Common Data Model "schema" developed through the Collabriculture initiative 17

Figure 7: A photo as an example of unstructured data 18

Figure 8: AxisTech Managing Director Wes Lawrence demonstrates the 8 Pillars .. 19

Figure 9: Folder structures risk leaving data in silos and duplication of data. 21

Figure 10: Types of data professionals, general job requirements and weighting. Wang, A., July 2023 22

Figure 11: Chris at Turkovich Family Wines 24

List of Tables

Table 1. Travel itinerary 7

Table 2: Five levels of data maturity to infer an organisation's ability to create value from its data (SoftKraft) 15

Table 3: Examples of data categories 16

Table 4: Difference between BIG and small data (adapted from website: Geeks for Geeks online resources) 18

Foreword

Prior to studying viticulture and working as a vineyard manager, I worked for almost a decade in the resources sector as a Mine Geologist. This involved mapping everything from landscape scale features to observation of similar features under the microscope. Working in economic geology there was always a consideration given to how observations in the real world related to grades received back in assays and how well mapping aligned with this.

In transitioning to viticulture in the early 2000s, I worked in a diverse set of roles from operations through to becoming a viticulturist, finding my combined disciplines a perfect fit for this role.

With precision viticulture techniques becoming commonplace, I've been able to use both my geographic information system (GIS) and data management skills learned in the mining sector, to push the limits of detailed management in vineyards, doing so to improve fruit quality and vine performance in blocks as small as a couple of hectares, through to hundreds of hectares.

Fast forward to 2020 and I was still pursuing the ideal of leveraging data to better understand vineyards and create an avenue to 'vineyard management of the future'. To achieve this, I always felt that data management was critical. Wanting to explore this, I applied for a Nuffield Scholarship to understand how data is managed, not only in other agricultural sectors, but in industries more broadly.

What does agriculture do differently and why does there appear to be a disconnect between data managers and farm managers? Is it possible to start speaking a common language and prepare vineyards for the digital transformation that is coming?

Travel Itinerary

Table 1. Travel itinerary

Travel date	Location	Visits/contacts
3rd – 11th July, 2021	Victoria: Melbourne Chadstone Shepparton	Che Wightwick (earthSQL) Jesse Reader (Costa) Neil Carter (Microsoft) Putra Sadikin (IMDEX) Brian Riordan (Growdata)
2nd – 5th November 2021	South Australia	David Zadow (Paisley Hill Vineyards) Bradley Nott (University Adelaide, Loxton Agtech Field Day – Vitivisor) Steve Schiller Nicole Pitman (Elders Agronomist - Barmera) Susanne McLoughlin (Vine Health Australia)
19th – 25th November 2021	Western Australia: Perth	Simon Foley (AgWorld) Peter Schmidt (Microsoft) Nathan & Andrew (Specterra) Wes Lawrence & Suzanne Loveland (Axistech) Andrew Paterson (AVEVA) Grant Gilmour (Beanstalk Agtech) Darren Gibbon (WA DPI) & Alex Jenkins (Curtin) Joanne Wisdom (UWA) Ravi Nichani (maptaskr) Vineyard Automation Field Day
15th – 20th March 2022	United Kingdom:	Andrew McAllister & Simon Sekula (AVEVA) Ian Beecher-Jones (JoJo's Vineyard) Tony Millanowski & Cameron Roucher (Rathfinny)
August 2022	Global Focus Program	Data a common theme at visits with many insights
30th August - 4th Sept. 2022	United States: California New York State	Karissa Isaacs (Superior Farms) Stephanie Boulton (LODI Winegrape Commission) Brent Sams (E&J Gallo) Roger Boulton & Mason Earles (UC Davis) Chris Turkovich (Turkovich Family Wines) Terry Bates (Cornell University)

5th - 15th Sept. 2022	Canada	Anthony Shaw (Brock University) Shawn Moen (9 Mile Legacy Brewery) John & Barb (Black Fox Farm & Distillery) Jordan McFarlen & Ami Caragata (Conexus Cultivator Regina)
January 13th – 20th January 2023	Adelaide, Brisbane & New Zealand	Kerrie Mengerson, Bryce Polley, Will Browne & Chris Lehnert (QUT) Ian Woodhead (Lincoln Agritech) Mark Eder & Armin Werner (Manager – Waiata Vineyard Waipara) Kate Lattey & David Allen (Pernod-Ricard) Joris Besamusca (Integrape) Ben McLauchlan (Nuffield Alumni Blenheim NZ)
	COVID Online (Zoom)	Peter Melville (Nutrien Ag Solutions) Chris Mendes (The Yield) Ben Harris (TWE) Hamish Munro (Pairtree)

Figure 1: The AVEVA Welcome Wall: London Unified Operations Centre, March 2022

Acknowledgments

Being a Nuffield Scholar has provided me the opportunity to take a deep dive into the topic of data management in viticulture. Thank you to Nuffield Australia for the opportunity provided through the scholarship, the alumni network, once-in-a-lifetime travel experience and the license to question.

Thank you also to project investor Wine Australia who believed in my topic and provided me with ongoing support and advice. Special thanks to Paul Smith, Jo Hargreaves, Alex Sas and Dave Gerner who have helped me through an extended Nuffield experience.

With COVID being a significant factor in extending the timeline of this project, I have attended countless presentations and conferences, arranged in – person and Zoom meetings, met formally over a coffee, or just noted chance conversations on the side. Thank you to all those who have contributed their thoughts, ideas, experience, and connections to my understanding of data management and its application in viticulture. I reflect on many of these and how pearls of wisdom were gleaned from the full spectrum of experiences, from sitting in a farm kitchen and seeing handwritten data on the back of kitchen cupboards (albeit with ethernet cables coming in from the irrigation shed! (Figure 2), to world leading examples of digital twins and enterprise visualisation in the world's megacities.

Thank you also to my local and international 2021 scholar group, who have become lifelong friends with the encouragement, support, and perspective they have provided.

To my wife and extended family, who kept life rolling for our young family in my absence. COVID provided plenty of curve balls, the biggest of all being that we were unable to travel as a family and much of my travel was completed solo.

Figure 2: David Zadow's impressive irrigation solution, monitoring, and rainfall records

Abbreviations

AI	–	Artificial Intelligence
AIM	–	Asset Information Management
API	–	Application Programming Interface
AWS	–	Amazon Web Service
BI	–	Business Intelligence
CDO	–	Chief Data Officer
CIO	–	Chief Information Officer
CSV	–	Comma Separated Values
CTO	–	Chief Technology Officer
ELT	–	Extract, Load, Transform
ESG	–	Environmental, Social & Governance
ETL	–	Extract, Transform, Load
FAIR	–	Findable, Accessible, Interoperable, Reusable
GIS	-	Geographic Information System
IP	–	Intellectual Property
IoT	–	Internet of Things
JSON	–	JavaScript Object Notation
ML	–	Machine Learning
OI	–	Operational Intelligence
PDF	–	Portable Document Format (a digital “hard copy” that is difficult to edit)
PI	–	Production Intelligence
SARS-CoV-2		Coronavirus disease 2019 (COVID-19)
SQL	–	Structured Query Language
XML	–	eXtensible Markup Language

Objectives

- Objective 1: Provide insight and understanding to viticulturists and vineyard managers around data, its storage, terminology, and opportunities.
- Objective 2: Raise awareness as to where data is being captured in a viticultural business.
- Objective 3: Test how data will change the way vineyards are viewed.
- Objective 4: Answer the question of, “What do we do with the data?” as it applies to data storage and application.

Chapter 1: Introduction

The wine sector employs 163,790 full and part-time employees nationally (Luke Mancini, Nuffield Project No 1614, p10). According to the Bureau of Meteorology, the mean January temperature in Coonawarra, South Australia, in 2023 was 20.2 degrees. According to the author, there were 179 vines identified for layering in a block of Cabernet at Penley Estate last year.

All these are examples of data and as a wine sector, much is collected!

Data is collected in a myriad of forms, from online forms, invoices and accounting software, spray diaries, apps on mobile devices, aerial imagery, IoT devices, emails, weather stations and the traditional, but broadly used, clipboard, pen, and paper.

Do many realise however, that each of these can be described as a type of data? That businesses are on a journey of data maturity and that the conversation around data is much broader than who owns it.

In the 20+ years working in the wine sector, viticulture specifically, I have pondered about data and the increasing role it plays in managing vineyards and businesses effectively. Data opportunities have been missed, with good IoT, agtech and precision viticulture techniques not being fully utilised as interrogation did not go beyond a "pretty picture."

Dave Lawie of Perth-based IMDEX points out that *"humans are good at recognizing trends and patterns, but a lot of what may be important is in the outliers."*

Thoughts therefore were always along the lines of whether representatively collected data, potentially extracted from existing formats, and interrogated correctly, could provide an alternate way of managing vineyards by bringing "outliers" directly to the attention of vineyard managers. Could data also provide a form of scale and efficiency to ensure that particularly small to medium size wine businesses could remain competitive in a world where scale seems necessary for long term viability?

Too often data is left in silos, used for a specific purpose or calculation, and not easily reviewed relative to other measures and/or temporally. Similarly, data entry is often duplicated, or existing (industry) data sets not accessed by producers. Duplication occurs with spray diaries, along with manual re-entry of data is synonymous with errors, and no one knows the "right" data to be looking at. This can introduce doubt into the conclusions made from data, particularly when it's realised there are differing versions of the "raw data."

The United States Geological Service (USGS) currently provides an estimate of the cost of poor data management:

"Poor data quality, redundant data, and lost data can cost companies 15 percent to 25 percent of their operating budget."

With tightening margins, increased compliance and environmental, social and governance (ESG) expectations, the next decade could be critical to the future of the sector. There are clear benefits to doing data well.

Chapter 2: What data is being collected?

2.1 The industry

Practitioners involved in the production of grapes are excellent at collecting data. The sector has benefitted from years of data collected and reported, given the importance placed on site-specific factors of climate, soils, aspect, canopy management and interrelationship of vine balance and quality of production reflective of terroir.

For good decisions to be made from data, it needs to be collected representatively. While this is not the focus of this report, the way data is collected is changing. From an individual going into the vineyard to take observations and counts, viticulture is rapidly moving towards data sourced from IoT devices, sensors and imagery incorporating computer vision, machine learning and edge computing. As per the title of this report, here come the robots but what do we do with the data?

2.2 Collecting data

Viticultural production through its nature results in significant data collection throughout the season, but much of this fieldwork will no longer be done by individuals on the ground, but by sensors and IoT devices deployed either in:

- i. static locations providing a continuous feed of data;
- ii. Tractor/robot mounted collected data on the go or specifically related to an operation; and
- iii. remote sensing.

Mike Porter in his online interview states that, "*this changes the skills required, with individuals needing to be able to relate or question data correctly and with a technical understanding" (Porter, M. Around the world in 80 projects)*

This will be increasingly critical, with the expertise required to question rather than just collect data, knowing what spurious values or unrealistic trends look like.

Mart Verstappen a Marlborough-based agronomist in New Zealand who specialises in reconciling soil and foliar nutrition sampling (Figure 3), provided the explanation that, *"it requires a shift in mindset to look at the data critically and address the changes that need to be made."*

Mart explained that if looking at petiole results that indicate a nutrient deficiency, it remains prudent that professionals understand the data they are looking at to know if a result is:

- Expected or probable.
- Aligns with recorded observations.
- Fit a trend.

To overcome concerns around "rubbish data" and ensure integrity, it is important not to take all results at face value. As such, validation is an early step in taking ownership of data.

Figure 3: Mart Verstappen collects data throughout the season to track inputs through to results. This allows him to tell a seasonal story and demonstrate the difference plant nutrition has made to a crop

2.3 Know the data being collected

When asked about data, the cross-section of producers interviewed either answered: “we don’t do data well” or “we could be doing better.” There were many examples of data being recorded but not used beyond a single purpose, if at all. A systematic review is required to understand the data being collected and where value can be generated by breaking it out of silos. Importantly though, the data is a) being collected, b) used (if only for a single purpose) and c) presumably stored… somewhere.

Examples that occur daily, weekly, and monthly includes observations of phenology, pest and disease monitoring, weather data, irrigation readings, bud numbers at pruning and bunches prior to harvest, harvest tonnes and historical yield. It goes on.

To address this, advice received has been along the lines of undertaking a “data audit” (Jesse Reader pers. comms.) or due to the possible negative connotations’ audits imply, creating a “data map” of the vineyard business (Suzanne Loveland pers. comms). This will help visualise all the data being captured in business processes, both financial, production-related, and technical.

Considering how data is being used and what role it plays in a business’ decision-making processes is the next step. Understanding this allows a business to assess its “data maturity” and to better understand use of data.

What is data maturity?

Data maturity is a measurement that demonstrates the level at which a company makes the most out of their data. To achieve a high level of data maturity, data must be firmly embedded throughout the business and fully integrated into all decision-making and activities.

The box above provides a simple explanation with Table 2 adapted from “SoftKraft” consultancy’s summary of four models detailing levels of data maturity (link in references).

Understanding these aspects can then assist in understanding a business, the data it collects and the data it requires to improve.

Neil Carter from Microsoft suggested having an “*outcome in mind”* (Carter, pers.comm., 2021), to know what data is already being captured and what may be required to advance. This can guide aspects such as which agtech or IoT device to invest in, how data can be stored in a database and then shared or presented and which comparative data to capture to understand performance. As stated by Ami Caragata from the Conexus Cultivator in Regina Canada:

“Data in its raw form is useless. How do you extract, process, and absorb data?”

Figure 4: Considering the power of data with Ami Caragata at the Conexus Cultivator, Regina Canada

Table 2: Five levels of data maturity to infer an organisation’s ability to create value from its data (SoftKraft)

Level 1: Unaware	No data management or integration, organization has no awareness of data management as a discipline. "We don't know that we don't know.”
Level 2: Initial	Basic data management and integration, organizations may use simple tools that they already have in the organization - spreadsheets, documents, databases etc to identify, collect, organize, and maintain data
Level 3: Fragmented	Advanced data management and integration. An organization recognizes the importance of data management in improving efficiency and effectiveness. However, data remains siloed to business units and/or software. Each department makes decisions about their own data (management) systems and responsibilities without considering the broader organizational needs making data management fragmented
Level 4: Enterprise	Fully integrated data management and integration. there is a recognition of the importance of data management at the highest levels of the organization. Likely, there is a Chief Data Officer or an equivalent role and an enterprise-level Data Management team
Level 5: Strategic	Data driven innovation with no decision made unless supported by data

Chapter 3: What's changed?

What of this focus on data anyway? What has changed that data is becoming more critical and what do growers need to do better? Decisions based on data provide opportunities for efficiency, productivity, and transparency. In a 2021 meeting with Neil Carter (Microsoft), he explained that the opportunities from data have increased due to factors of:

- Cheap, low – power (IoT) sensors becoming readily available.
- Improved speed and efficiency of field devices.
- Elasticity in "compute" and storage that have improved through the advent of virtual machines. This means that both the volume of storage and speed of compute can be dialled up and down as required and the necessity for heavy capital outlays to purchase physical servers are a thing of the past.
- Edge computing which is pre-processing at source and thereby reducing the overall volume of data being transferred and stored.

Improvements have also meant that it is now possible to leverage AI to extrapolate probable values for unknowns, based on measured variables. Quality inputs are required, but the number of samples can be reduced.

Beyond these technical aspects of data storage, within viticultural production and the broader wine sector, reasons that necessitate good data management include such things as:

- Legislated requirements of traceability and food safety.
- Greenwashing and requirements around ESG i.e., proving your credentials.
- Continual improvement and benchmarking.

Yet David Armour of the Bregato Institute in Blenheim warns that while data will help optimise the use of assets, the sheer quantum of data could become overwhelming.

3.1 Types of data and the broader data landscape

Following the understanding of how data is collected and how it is being used, it is timely to introduce the concept of how data is broadly classified, where it may occur and in which format. Primarily, data can be categorised into one of three broad types:

- Structured;
- Semi – structured; and
- Unstructured.

Examples of each of these are detailed in Table 3, noting that most data collected by vineyard scouting could be categorized as structured data, while the other two categories are more of an enigma.

Table 3: Examples of data categories

Data Type	Examples
Structured	Bunch counts, pest & disease scouting, irrigation records
Semi – structured	Email, PDF, XML, JSON
Unstructured	Videos, audio files, social media posts, email

Structured data is most easily loaded into a database with minimal additional formatting required. Other forms of data will require processing, special import, and additional validation before they can be referenced for reliable insights.

In terms of structure, there is also consistency of data. This can relate to aspects of nomenclature and format, ensuring that comparative data sets are delivered in the same way. Wine Australia has invested to future-proof the industry with support of the www.collabriculture.com initiative, making digital assets such as grapevine varietal naming conventions easily available to programmers, app developers and data engineers. These are stored in the CollabricultureORG “GitHub” repository, which is open source and available for all to access (https://github.com/CollabricultureOrg/). Figure 5 is the diagrammatic of vineyard naming conventions and Figure 6 the “Common Data Model” database schema developed by the initiative.

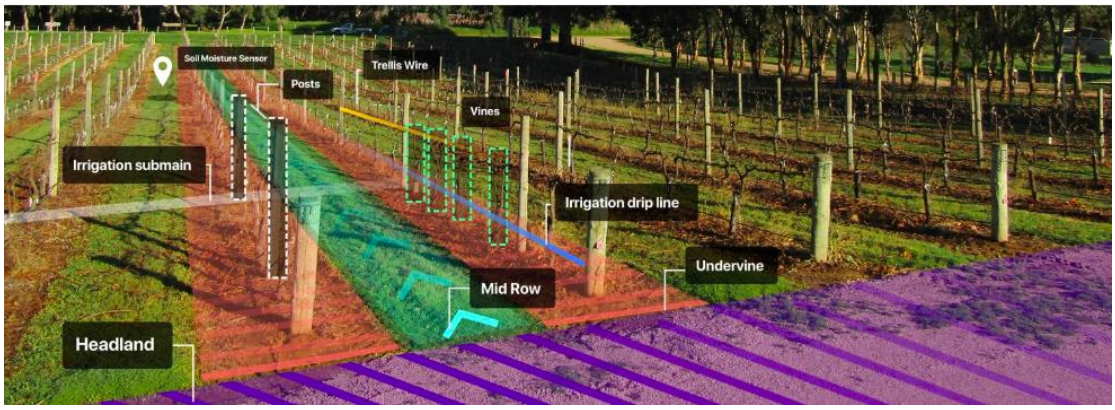


Figure 5: Vineyard naming conventions agreed by consultation with industry representatives

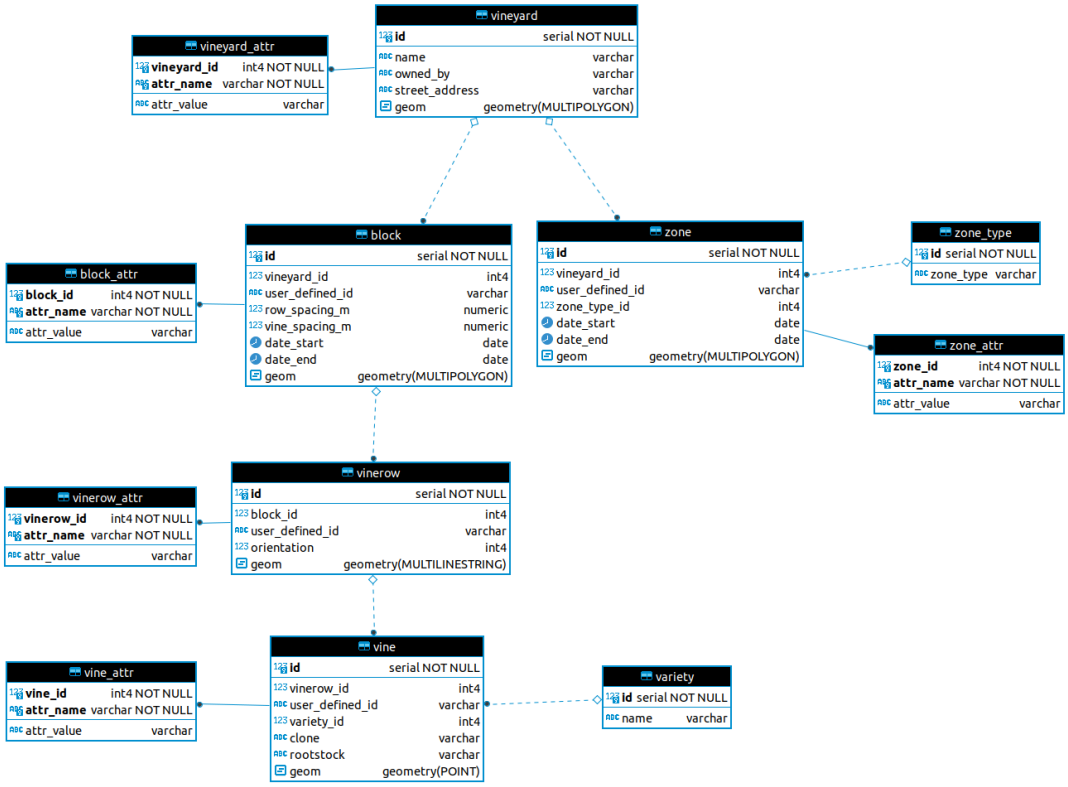


Figure 6: Common Data Model “schema” developed through the Collabriculture initiative

Semi-structured data is not in a tabular structure but due to embedded tags and other markers, data can be mined from within files or structures. JSON files are such an example, which are easy to transfer, open in GIS and with basic code data can be extracted for upload into a structured database. PDFs on the other hand are the least easy to work with and in their design provide limited opportunity for data transfer.

Figure 7: A photo as an example of unstructured data

Unstructured data is not arranged according to a preset data model or schema and is prolific as it could be anything. It may have an internal structure but is not defined, can differ between files and be of any format. An example could include the image in Figure 7 taken during the Nuffield travels. Within the picture there is data about my location and I that could be identified using image-recognition software. The image also contains associated "metadata" detailing location, date, and time of the picture, along with other details.

3.2 Defining big, large, and small data?

As practitioners it can be too easy to think that a large data set constitutes BIG data. However, most viticultural data and particularly structured data, is considered either small to large, but not big. It may seem trivial whether data is large or whether one is truly dealing with "big data," but understanding the difference can be helpful when considering data storage and analysis. Table 4 provides further explanation.

Table 4: Difference between BIG and small data (adapted from website: Geeks for Geeks online resources)

Data Type	Examples
Small	Data is high quality and reliable, datasets can influence/impact current decisions, data can be accumulated in an Excel file, simple enough to be workable by an individual, analyzed using traditional statistics, tens, or hundreds of Gigabytes, generally structured
Large	Large accumulations of small data, traditional statistics still apply, locally hosted and/or centralized database, often mistakenly thought of as big data on account of volume
Big	Data quality and reliability can vary, data volume is immense, size and complexity make conventional data processing techniques inadequate, advanced analytical techniques such as machine learning are required, more than Terabytes, large "chunks" of structured and/or unstructured data

3.3 FAIR data

The principle of Findable, Accessible, Interoperable, Reusable (FAIR) data is to enable "machine actionability" of data with minimal human intervention. In terms of how practitioners understand this acronym, the concept of interoperability and reusability of data is particularly tangible. This study as one of its key outcomes, looks to see if existing data will deliver new ways of tracking (vine) performance, either through easier access to existing indices, or alternately the ability to apply new indices previously too hard to measure in a production vineyard environment; this likely achieved by use of new IoT devices to sense vineyards in new ways previously only possible by researchers and/or by breaking data out of silos and "reusing" it beyond what it was initially collected for. Appendix 3 provides further detail on FAIR data. The image below has been extracted from www.wikipedia.org

3.4 8 Pillars of agtech and farm data

Wes Lawrence of Axistech in Western Australia, in collaboration with Belinda Lay of Coolindown Farms, developed an 8 Pillar framework to describe the relationship between farm IoT, agtech and how to manage the data they deliver. Axistech strives to convey understanding as it develops proprietary sensor technology, deploys it on clients' farms and then manages the IoT data in conjunction with other farm data streams. Specifically, Axistech looks to realise FAIR data principles for its clients, as it facilitates data collection, breaks data silos, and delivers insights.

Appendix 1 has a full description of the 8 Pillars of agtech and farm data.

Figure 8: AxisTech Managing Director Wes Lawrence demonstrates the 8 Pillars

Chapter 4: How is data stored?

"Data was critical to facilitate continual improvement (Kaizen) and no industry – specific software was available. A custom production system "Hosusaku Keikaku" was developed in 2012 to house (this) critical data " – nabeco Japan.

The way data is stored has been changing and for many in agriculture it has been as subtle as to go unnoticed. However, while the reality in agriculture may continue to be USB transfer, hard drives, and physical servers, increasingly everything from the latest app to business-critical data is being stored in the "cloud." But what is this cloud and what difference has it made?

Cloud storage, while seeming virtual and intangible, has many benefits such as access, sharing and flexible compute depending on requirements. Cloud storage is housed in data centres around the world, offering security, anonymity, and access 24:7 via the internet. Storage of data in these environments provides additional advantages not widely recognised such as: security, backup, and low impact. On this last point, there is an increasing awareness of the carbon footprint of data and its storage.

While there are few individuals that do not use cloud storage in some way, the industry publication "Precision Viticulture, A new era in vineyard management and wine production" has not been updated and advice on storage remains as a:

"…good directory structure from the beginning and naming conventions used for folders and files that are easily recognisable and used consistently…"

It is also recommended that:

"a service provider is used for these activities unless there are the skills and time available to properly manage this process".

Further, Bramley and Lamb in their 2006 report, *Precision Viticulture – Making sense of vineyard variability"* recommended sourcing expert skills given that:

"PV is difficult and requires skills in viticulture, soil science, spatial statistics and GIS that are unlikely to commonly reside in the same individual."

While sound advice, storage of data in a directory structure can contribute to data silos and difficulty in extracting data. Multiple copies of the same, or worse, modified data can easily be saved in the same folder and multiple users cannot work on data at the same time.

As such it is recommended that data is stored in a dedicated database, be it in a proprietary system or custom built managed by a data engineering business. Data should be able to be accessed directly and any updates that are made reflected in all associated dashboards, reports, links to GIS and more.

In terms of sourcing professional management of a business' data, both this research and that of fellow 2021 (New Zealand) Scholar Lucie Douma identified the field of data engineering as a specialised skill and that producers need to consider outsourcing their data management. Much like tax advice is left to accountants, a data engineer should be considered to manage a company – specific database of all critical data.

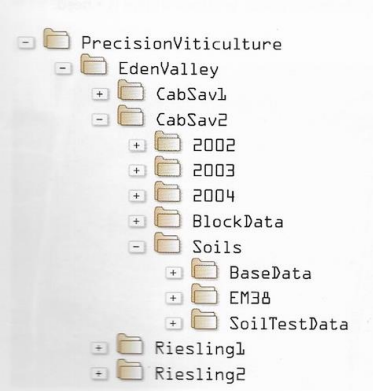


Figure 17. A possible directory structure for storing Precision Viticulture data.

Figure 9: Folder structures risk leaving data in silos and duplication of data.

4.1 Cost of storage

It is of note that data professionals such as Neil Carter from Microsoft have provided comments on data storage as cheap, while wine industry professionals have expressed the cost of data storage as expensive. The volume of data generated by agricultural enterprises is likely underestimated, with an anonymous source gauging the volume of data generated by an average-sized family farm as equivalent to a small-medium size mining operation. On this point, there appears to be a disconnect between a viticulturist who collects data and professionals who work with data.

Are we storing our data cost effectively and in the safest way?

4.2 Rise of the data aggregators

Coincident with my Nuffield journey there has been the increasing rise in awareness around data aggregation and its ability to break data out of silos.

This work is undertaken by data engineers, yet the role of the data engineer is little understood and differs from that of a data analyst and other data professionals (Figure 8).

Data engineers focus on the preparation, validation, transformation, upload, download and safe storage of data. This ensures any insights from data can be made with a high degree of confidence, given the validation and high quality of data being assessed. The data engineer helps to overcome a "rubbish in, rubbish out" scenario and prepares data at significantly less cost than an analyst, ensuring analysts can focus on what they do best in the time/budget available.

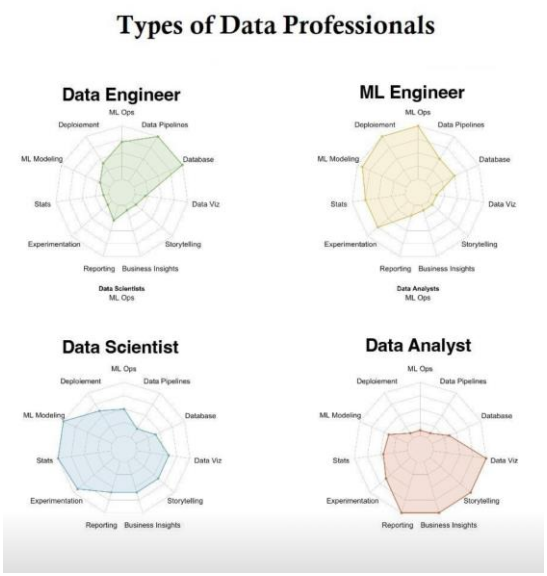


Figure 10: Types of data professionals, general job requirements and weighting. Wang, A., July 2023

The data industry talks of the process of loading data for storage in the context of: ETL (Extract, Transform, Load). As the transform step involves validation, checking of nomenclature and correct data structure, it is recommended that for most agricultural data, a process of ETL should be followed to ensure high quality and reliability of data storage.

Systems themselves should intuitively have the ability to be interoperable.

Just as there are a limited number of programming languages, there is a limited number of database schema which underpin all the data capture/storage/analysis platforms being used. Some may have increased in profile to be no longer noticed, particularly when their ownership and branding are some of the world's largest companies.

Some examples include:

- Proprietary: Microsoft SQL Server, Oracle SQL, Amazon RDS/AWS/DynamoDB.
- Open Source: PostgreSQL, MySQL, MongoDB Atlas.
- Industry Specific (likely built on one of the above): Growdata, AgWorld, Athena.

Data can be transferred to a database using a myriad of file formats or an Application Programming Interface (API), ideally with a data engineer maintaining oversight and acting as a gatekeeper to any spurious data. Once in a common format such as outlined by the Collabriculture Common Data Model, it becomes easier to access data within a database and similarly transfer data between databases.

An example relevant to viticulture could relate to rainfall data, which could be stored in a database to be transferred to multiple agtech platforms such as Athena's IRTech or Sentek Green Brain Platform.

Several data companies visited as part of this research have expertise, a focus, or set the industry benchmark for data aggregation and presentation! While not comprehensive or in any order, these were:

- Microsoft
- AVEVA
- earthSQL
- Pairtree
- The Yield.

Fellow Nuffield International Scholar Lucy Douma has provided a detailed report on data interoperability and as such, this report will not cover this topic further.

Chapter 5: Data ownership opportunities and risks

“But who owns the data?” Is this a question, statement, or full stop? It seems that due to many uncertainties around data in viticulture, it can be all of these. But what does “taking ownership” mean?

5.1 The hero

We all need a hero, and for success with data every business will need a data advocate. This is someone who:

- takes ownership of data;
- understands why it is being collected;
- how it should be collected;
- that collection is done well;
- the numbers seem right; and
- that all the data is safely stored.

Being a hero can be tough, particularly when concepts around data seem abstract, its analysis not intuitive and terminology another language. Support is required and as businesses are becoming increasingly complex, it is suggested that corporate wine businesses consider acquiring roles such as Chief Data Officer (CDO), while smaller businesses either hire or outsource a Data Manager to navigate a business’ increasingly important reliance on data.

“Yield mapping declined due to manager disinterest…” – Anon requested.

“Data is like sand and without the right people it can slip through your fingers. But with the right intent it can be turned to glass” - Jon Young Flores (evokeAg, 2023)

5.2 Investing in data

Before being concerned about who owns data, it is necessary to ask whether one’s own investment in data is sufficient to impart ownership. As in the northern territories of Canada where a human presence is supported to maintain basic sovereignty, if there’s no data hero, insufficient investment and understanding, let alone appropriate storage, can it be said that ownership is already being left to chance? Below are several comments recorded during this research:

“We can get by without all that data…”

“Experience and instinct; I could make the decision before seeing the data.”

“What can I do within the constraints of yield, value and budget?”

Investment should be considered beyond business capital, but also time and expertise. To consistently capture data, get IoT and agtech systems working well and store data with an ability to find it in the future, requires an investment of time along with money. It also requires individuals to recognize the importance of collecting representative data. Che Wightwick from earthSQL remarks that:

“It is not enough to leave a business’ data to a third party, without being closely involved in what is happening with that data.”

5.3 The contract

With data storage increasingly in third party software platforms, just as with any commercial agreement there should be due diligence in interrogating a contract. Aspects to consider are:

- who retains the data and for how long;
- what will the data be used for (by the third party); and
- can data be downloaded and/or uploaded into and from the platform.

Ownership of data becomes particularly contentious when it is being used by a third-party business to build their own model IP and reliability, while charging the user for the privilege! Distinguished Professor Kerry Mengerson, Director of the QUT Centre for Data Science comments that currently it remains difficult to properly value data or prove that all data used in designing a model has been removed. Neil Carter imparted a need to protect data as if it is your own IP while Jason Strong, Managing Director of Meat and Livestock Australia is on the record as saying:

"Data has no value until you use it."

A solution such as a blockchain ledger, embedded metadata or a creative commons licence as an idea discussed with Sjaan Davey from SARDI, may provide avenues for traceability of data and/or provide a means to compensate owners of data. These methods either embed source and ownership details, or as in the third example provide an ability for others to legally share data, while the originator retains ownership.

How long data is retained is a two-way street, with data not always of benefit to a third-party provider and a cost to retain for long periods. Producers should keep in mind that they may need to be prepared to store data on their own terms, should a contract period expire, and a provider is no longer obligated to retain their data or worse, if a provider business ceases to exist.

Opportunities; is anonymity more important than ownership?

Instead of treating data as a pot of gold to hold close due to a constant fear of it being taken, what is the opportunity that fewer are prepared to consider: Could sharing of data be a catalyst to accelerate the sector, promote excellence and maintain profitability?

As part of this study, several comments were captured promoting the sharing of data and whether "anonymity is more important than ownership?"

A particularly compelling example was a subscription called the "Farmers Business Network."

https://www.fbn.com/

This was explained by the brother of one of the founders, Chris Turkovich of Turkovich Family Wines, California.

Figure 11: Chris at Turkovich Family Wines

Chris explained that the network facilitates the sharing of performance data and fertilizer/chemical pricing trends as examples, along with text message surveys that provide feedback to contributors within 24 hours. Speed of intelligence and insights independent of commercial players provide the value, while independence of the third party aggregating the data, provides member farmers with the confidence that their data is secure and not being used for intelligence beyond the network.

Insight into products such as Microsoft Farm Beats was provided by Perth-based Peter Schmidt, as another example of where global insights are being used to optimize site-specific forecasts. Preference is given to selecting IoT devices to feed into this environment to ensure data integrity and ease of integration.

"Aggregated data sets can provide value and insights well beyond the individual organisations' that collected them" Jon Young Flores (evokeAg, 2023)

"Industry scale data contribution has potential to create a "backbone" for modelling for the industry" Roger Boulton UC Davis, July 2022

5.4 Farm data principles

As per the National Farmers Federation website:

The Australian Farm Data Code (the Code) was developed and adopted by the National Farmers' Federation in consultation with industry.

The Code aims to promote adoption of digital technology, by ensuring that farmers have comfort in how their data is used, shared, and managed.

The Code is intended to inform the data management policies of service providers who manage data on behalf of farmers. It is also a yardstick by which farmers can evaluate the policies of those providers.

A link to the updated code (2023) is provided in the references section.

Chapter 6: Managing data – current and future opportunities

Storing data in a structured database with a clearly defined schema can facilitate ease of data integration and ease of transfer, both in and out of a database. The benefits of doing this are real and the use cases are limitless:

"To compile a grower presentation takes a lot of work and time. It is necessary to come up with a system that takes data out of silos, leverages historical data, and builds a prediction from a range of data sets from soils to weather."

Verstappen, M., Jan 2023

Many growers call for single platforms that will fulfil all their requirements and it is possible that a proprietary software platform could achieve this. As such, when reviewing an agtech or a farm management platform, it is prudent to consider what percentage of features will be used and whether this represents value.

However, the likelihood of all agtech firms to build comprehensive platforms is unlikely, given this would stifle innovation and soak up limited financial resources to add, link or duplicate features offered by competitors. This effort would also create distraction from the niche an agtech business specializes in.

Instead, it will be important within a viticulture business to identify the tech that is relevant to a vineyard's production methodology, purchase this and then likely invest in a means to store and aggregate the critical data on their own terms, even if this means paying for some duplication of platforms and/or data engineering services.

David Zadow explains his experience of the South Australia Riverland's "VitiVisor" initiative. He says that as every vineyard is different, every grower will want different information. His conclusion is that even with a platform like VitiVisor that was co-designed by growers, it requires flexibility to accommodate individual preference.

The benefit of this custom aggregating of data within a business can be measured by factors beyond an initial purchase price, with benefits including:

- Better internal communication.
- Data-driven decision making.
- The ability to reference between operational parts of a business.
- Introducing a common language to facilitate informed decision making.

Just as technical and vineyard managers could gain a better appreciation of the financial implications of events in the field, broadly employees of a business will gain greater insight, building transparency and common knowledge.

"Recall of specific data from large datasets can present issues."

David Allen - Pernod Ricard Jan 2023

Settling on how best to store data can seem daunting and the reality is not so simple. This research has identified several platforms that specialize in hosting, referencing, and presenting different kinds and volumes of data.

While not a recommendation or comprehensive, these are listed in alphabetical order to provide insight into a range of database products and their specific "niche":

i. **AVEVA PI System** – PI standing for “Production Intelligence,” the system’s strength is in storage, access and recall of real-time operational data. This sourced from long, continuous records reported many times a second by “Asset Information Monitoring” (AIM) systems such as sensors fitted to plant. This is presented as a digital twin or “single pane of glass” concept for “enterprise visualisations” that drill down from a high start level.
ii. **Axistech AxisStream** – a customisable data management platform constructed through data mapping of an agricultural business. It utilises Microsoft SQLServer and PowerBI to provide analytics, along with directly linking to AxisTech’s proprietary environmental, stock and soil monitoring sensors.
iii. **earthSQL Scion** – built on Microsoft SQLServer database with Collabriculture schema and linked to PowerBI for data visualisation. It is a customizable storage system for all forms of viticulture, environmental and IoT sensor data. Data can be imported, exported, and formatted for transfer between platforms.
iv. **Efficient Vineyard Project MyEV** – developed as a free cloud system for growers to collect, process, store and present structured precision viticulture data and specific sensor feeds. Developed as a collaboration to centralise many precision viticulture tools into a single grower platform.
v. **Growdata** – a data management product for the storage and reporting of spray diaries, labour tracking and field-level input data in horticultural operations. This remains based on a Microsoft Access database structure, which facilitates ease of querying and reporting.
vi. **Integrape & Vure** – built on a Couchbase, AWS and Koordinates platform, this software specialises in integrating field sampling and laboratory-tested results such as soil and leaf petiole, then integrating these with winery and remote sensing data. Vure streamlines data collection and general vineyard scouting. A special focus has been on the user interface design, to ensure ease of use, visualisation, and data transfer.
vii. **Pairtree** – a universal dashboard that is independent and can provide agtech and data integrations to be presented on the “Core” platform. Marketed as providing ‘one stop – operational oversight’.

6.1 What’s the opportunity?

Data is critical to assist smarter workflows, facilitate continual improvement, provide producers with decision support tools, machine learning and AI, which results in greater efficiency while improving oversight.

“How can a grower do things easier and better? Just because we’ve been doing things for 30 years, doesn’t mean we’re doing it right.”

Verstappen, M., Jan 2023

Extracting value from data provides an opportunity for maintaining (if not increasing) productivity from a constrained workforce, just as technology can create greater diversity within roles and increase staff engagement.

Accountability, transparency, and oversight of vineyards can be improved, with data critical in the process of continual improvement and measuring the performance of a business beyond financial metrics. Peter Schmidt (pers. comms.) proposes

considering the value of data in terms of the decisions made daily, weekly, monthly and on a seasonal basis; what data is required to make better decisions and what value can this provide? What are the challenges to obtaining this data? Finally, what is the net value of the many decisions made throughout a season, that by themselves have only a small impact, but that collectively make a huge difference to a season?

Data offers a future of improved internal business communication, particularly if paired with well-designed visualisations that assist in rapid assessment of complex scenarios. This was in part demonstrated by the VitiVisor platform co-developed by Riverland Wine and The University of Adelaide.

Beyond this, machine learning and AI will improve scenario planning with the development of decision support tools. This supported through the creation of digital twins, augmented reality training and presentation in immersive reality environments as is already being demonstrated by businesses such as Think.Digital in Australia.

All however depend on access to quality data.

Data can introduce a common language that facilitates the interpretation of technical facts to non – technical employees of a company, building transparency and common knowledge within a business. Better and automated ways of collecting data will also see more research – level insights, being accessible to farmers for productivity improvements and to respond to ever increasing production, financial and labour force challenges.

Conclusion

Data is critical to assist smarter workflows, facilitate continual improvement, provide producers with decision support tools, machine learning and AI, which results in greater efficiency while improving oversight.

It is suggested that a small to medium sized family farming operation is generating the same amount of data as a small mining company.

The value proposition and need to remain competitive make using data better a compelling case, particularly with terms such as machine learning, computer vision and AI used daily. These tools will provide efficiency, cost savings, decision support tools and an ability for individuals to manage larger areas, most likely remotely, yet still with detailed or even greater oversight.

All rely on quality data, but data is not easy, and it requires investment to extract value.

Having completed this Nuffield research and reflecting on what has been learned, it is obvious that things will not be perfect. There will not be one software platform that does everything; in fact, that would stifle innovation. For now, farmers will still use multiple app's every day, but things will get easier, and integrations will be more commonplace and seamless.

Instead of looking at data in 2D on mobile phones, wearable technology will provide an opportunity to see data feeds directly in the vineyard, inclusive of interpretation. Alternately, it will provide an ability to manage a digital twin of the vineyard from anywhere on the planet.

As such, viticulture and wine businesses are encouraged to consider which data is being captured, which data is important to store securely and which data delivers the most value, when preparing for a data-driven future.

Recommendations

i. Start with mapping the data in a vineyard. Look at this in terms of what is being captured now, what could be captured from existing monitoring and what data is desirable. Does it have to be manually collected or could an IoT device or sensor be installed for automation and easier transfer of data? Is there telemetry and how's connectivity? Can a device feed data to a central point?

ii. Once data is being collected and transmitted, where is it going? Is it to a cloud system and if so, can the data be exported or alternately, comparative data be imported, thus reducing the occurrence of siloed data. Is a custom database required?

iii. Consider the vineyard in the context of the overall business. Is there data from the vineyard that could add value if combined with that from the winery, financial or an external industry body as examples? Could predictive tools help decision making and scenario planning? Consider the business' data maturity to guide the next steps and advance the use of data, business insights and continual improvement more broadly.

iv. Ask for help. Data engineering is a profession involved in the management of data. A relatively small investment to engage a professional could provide significant dividends in terms of time and financial savings as a result of reliable business insights based on securely stored, quality data.

References

Allen, D. & Lattey, K. (January 2023). Personal Communication Pernod Ricard Winemakers

Armour, D., (January 2023). Personal Communication. Bregato Institute, Blenheim, New Zealand

AxisTech, Osborne Park, Australia, Personal Communication

Boulton, R., (July 2022). Personal Communication. UC Davis, California.

Bureau of Meteorology http://www.bom.gov.au/

Carter, N. (July 2021) & Schmidt, P (November 2021), Personal communication. Microsoft, Melbourne & Perth, Australia

Caragata, A. (September 2022). Personal communication. Conexus Cultivator, Regina, Canada

CollabricultureOrg github repository https://github.com/CollabricultureOrg/

Davey, S., (June 2023), Personal Communication.

Douma, L., 2022, Data Sharing to Achieve Data Interoperability, Nuffield New Zealand

Farm Business Network https://www.fbn.com/

Geeks for Geeks online resources and training https://www.geeksforgeeks.org/difference-between-small-data-and-big-data/

Konsek, H., Softkraft.co (2023), Retrieved from https://www.softkraft.co/what-is-data-maturity-and-how-to-climb-the-data-maturity-scale/#:~:text=Data%20maturity%20is%20a%20measurement,all%20decision%2Dmaking%20and%20activities.

LinkedIn: “Data Maturity: Climbing the scale in financial institutions https://www.linkedin.com/pulse/data-maturity-climbing-scale-financial-institutions-athena-sharma

Mancini, Luke, 2016 Nuffield Scholar, South Australia https://www.nuffieldscholar.org/reports/au/2016/understanding-australian-wine-industry-growers-guide-background-and-participants

Mengerson, K., (January 2023). Personal Communication. QUT Brisbane

National Farmers Federation, Farm Data Code https://nff.org.au/programs/australian-farm-data-code/

NetApp Data Products & Solutions https://www.netapp.com/data-storage/unstructured-data/what-is-unstructured-data/

Profitt, T., Bramley, R., Lamb, D., Winter, E., 2006, *Precision Viticulture, A new era in vineyard management and wine production*

Reader, J. (July 2021). Personal communication. Costa Group, Melbourne, Australia

SARDI Naracoorte Personal Communication: Loveland, S. & Lawrence, W., (November 2021).

Saleem, A & Beresford, S., EPISODE 59 – Exploration Radio. “Around the world in 80 projects with Mike Porter.” https://www.aig.org.au/exploration-radio-around-the-world-in-80-projects-with-mike-porter/ Presentation

Saleem, A & Beresford, S., EPISODE 51 – Exploration Radio. “Building An Icon with Dave Lawie.” https://explorationradio.libsyn.com/51-building-an-icon-with-dave-lawie-part-2 Presentation

SoftKraft: What is Data Maturity? Retrieved from: https://www.softkraft.co/what-is-data-maturity-and-how-to-climb-the-data-maturity-scale/

Turkovich, C., (August 2022). Personal Communication. Turkovich Family Wines, Winters, California

United States Geological Service. Retrieved from https://www.usgs.gov/data-management/value-data-management

Verstappen, M. (January 2023). Personal Communication Farmlands co-operative Blenheim

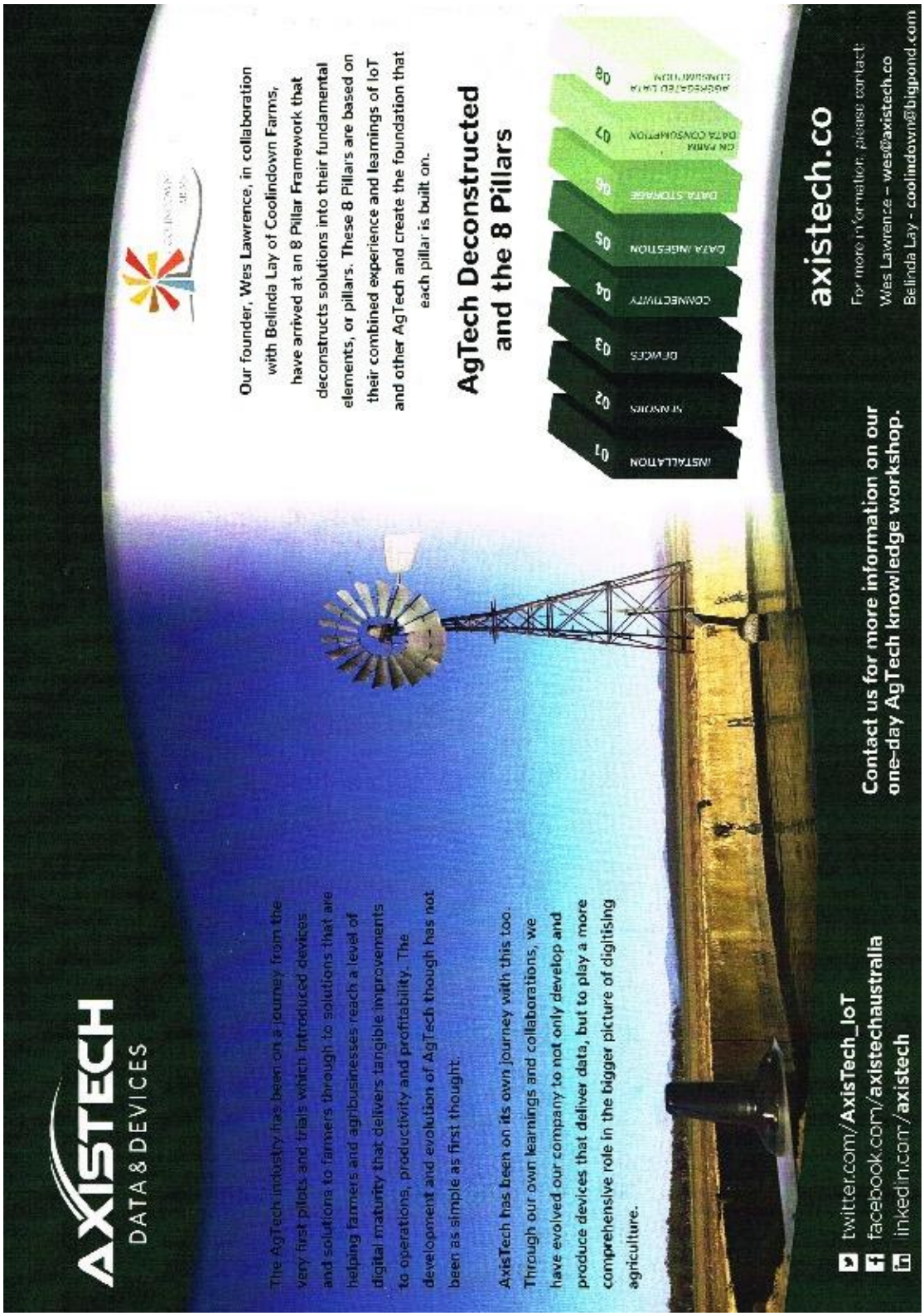
Wightwick, C., (July 2021). Personal Communication, earthSQL Pty Ltd, Melbourne, Australia

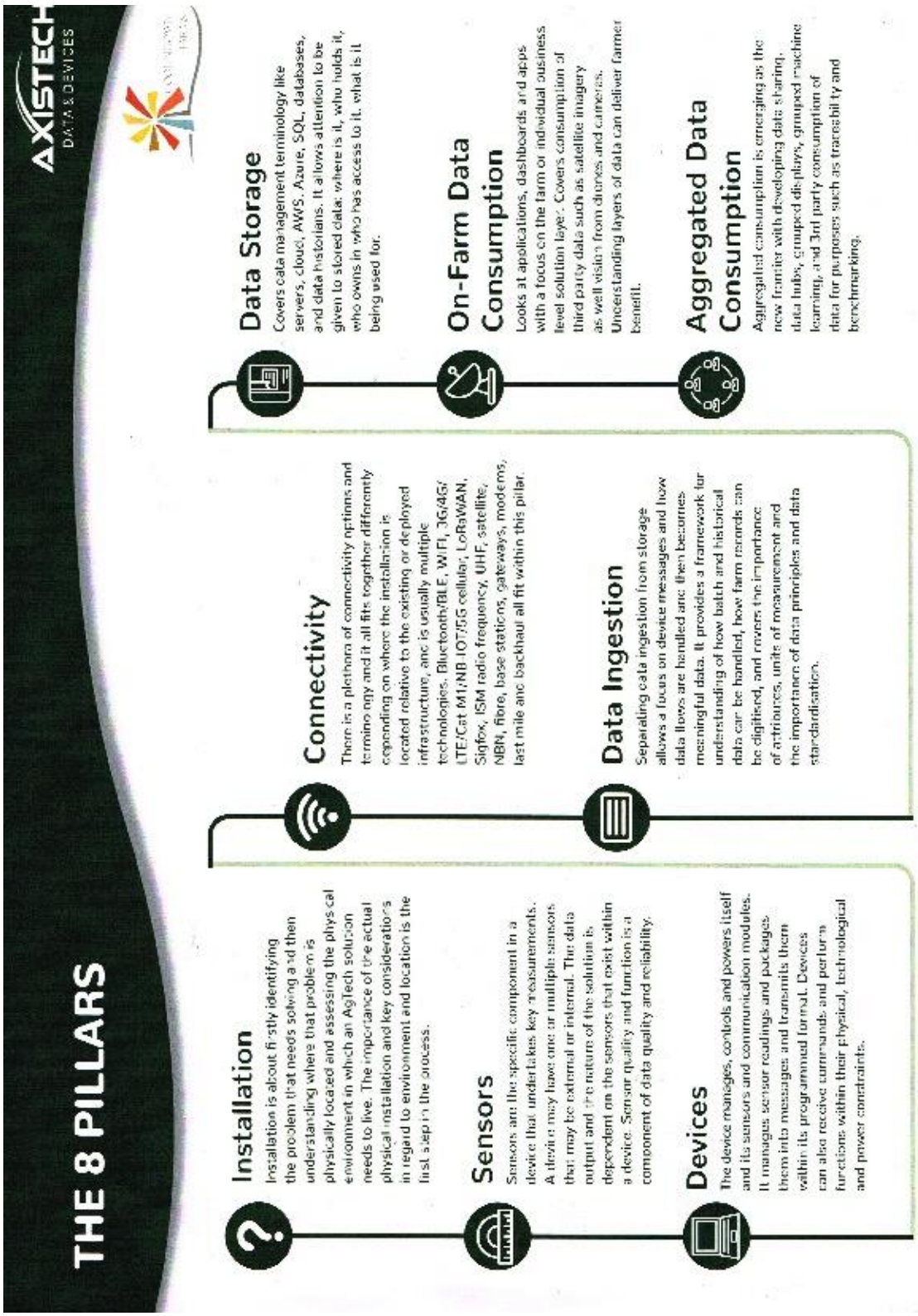
Wikipedia “FAIR data” https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

Zadow, D., (November 2021). Personal Communication. Paisley Hill Vineyards, Australia

Appendices

Appendix 1: AxisTech & Koolin Farms – The 8 Pillars




THE 8 PILLARS
AXISTECH
DATA & DEVICES
Installation
Installation is about firstly identifying the problem that needs solving and then understanding where that problem is physically located and assessing the physical environment in which an AgTech solution needs to live. The importance of the actual physical installation and key considerations in regard to environment and location is the first step in the process.
Sensors
Sensors are the specific component in a device that undertakes key measurements. A device may have one or multiple sensors that may be external or internal. The data output and the nature of the solution is dependent on the sensors that exist within a device. Sensor quality and function is a component of data quality and reliability.
Devices
The device manages, controls and powers itself and its sensors and communication modules. It manages sensor readings and packages them into messages and transmits them within its programmed format. Devices can also receive commands and perform functions within their physical, technological and power constraints.
Connectivity
There is a plethora of connectivity options and terminology and it all fits together differently depending on where the installation is located relative to the existing or deployed infrastructure, and is usually multiple technologies. Bluetooth/BLE, WiFi, 3G/4G/LTE/Cat M1/NB IOT/5G cellular, LoRaWAN, Sigfox, ISM radio frequency, UHF, satellite, NBN, fibre, base stations, gateways, modems, last mile and backhaul all fit within this pillar.
Data Ingestion
Separating data ingestion from storage allows a focus on device messages and how data flows are handled and then becomes meaningful data. It provides a framework for understanding of how batch and historical data can be handled, how farm records can be digitised, and covers the importance of attributes, units of measurement and the importance of data principles and data standardisation.
Data Storage
Covers data management terminology like servers, cloud, AWS, Azure, SQL, databases, and data historians. It allows attention to be given to stored data: where is it, who holds it, who owns it, who has access to it, what is it being used for.
On-Farm Data Consumption
Looks at applications, dashboards and apps with a focus on the farm or individual business level solution layer. Covers consumption of third party data such as satellite imagery as well visible from drones and cameras. Understanding layers of data can deliver farmer benefit.
Aggregated Data Consumption
Aggregated consumption is emerging as the new frontier with developing data sharing, data hubs, grouped displays, grouped machine learning, and 3rd party consumption of data for purposes such as traceability and benchmarking.

Appendix 2: Farm Data Principles – extracts Edition 1

Farm Data Principles

In order to comply with this Code, Providers must commit to and uphold the following Principles:

a. Transparent, clear and honest collection, use and sharing of farm data

Providers will:

- Be transparent (ie up-front and open) about what Farm Data is collected, as well as the purpose/s for which it is being collected, used and shared (including the use and sharing of Aggregated Farm Data).
- Provide Farmers with plain-English information detailing:
  - the identity of the contracting party/ies, and any affiliates who may receive Farm Data pursuant to the contract;
  - what Farm Data will be collected about them and their business;
  - how that Farm Data will be used and managed;
  - processes and conditions for data retrieval and service termination; and,
  - any risks or detriments that may adversely affect Farmers who share data with the Provider.
- Provide appropriate updates about any changes to its Farm Data practices, and where a change to terms and conditions is proposed, obtain consent or provide an avenue for the Farmer to reject the change without incurring a financial penalty.
- Provide a mechanism for Farmers to enquire about the collection, use, storage, security and sharing of their Farm Data.
- Notify Farmers of the legal jurisdiction in which their Farm Data is stored.

b. Fair and equitable use of farm data

Providers must:

- Only collect, use and share Farm Data for the purposes that they have made clear to the Farmer.
- Recognise that Farmers should benefit from the use of Individual Farm Data.
- Ensure that the privacy of Farmers is appropriately protected and that Farm Data is collected, stored and used in an ethical way.
- Ensure Farm Data is not deleted without authorisation.

c. Ability to control and access Farm Data

Providers will:

- Only use Farm Data for the purposes specified in the terms agreed by the Farmer.
- Preserve the ability of the Farmer to determine who can access and use Individual Farm Data.
- Protect sensitive data, such as personal/financial information, confidential information or intellectual property.
- Comply with obligations imposed by the *Privacy Act 1988* (Cth), including the Australian Privacy Principles.
- Explain to Farmers what reasonable steps are being taken to ensure any affiliates permitted access to Farm Data do not contravene the provisions of this Code.

Farm Data Principles

d. Documentation and Record Keeping

Providers must have a record keeping system in place to ensure that all processes and decision making related to Farm Data are documented in a clear and comprehensive manner.

e. Portability of Farm Data

Providers will:

- Provide Farmers with the ability to retrieve their Individual Farm Data – in both a processed (cleaned) and unprocessed form – for storage and/or use in third party systems (this includes during any Data Retention Period).
- At the request of a Farmer, delete any Individual Farm Data or Private Data relating to that Farmer.

f. Keeping Farm Data Secure

Providers will:

- Take all reasonable and prudent steps, in line with industry best practice, to ensure Farm Data, Private Data and Public Data are protected at all times from unauthorised access, damage or destruction.
- Promptly notify a Farmer where an attempt (successful or otherwise) has been made to gain unauthorised access to, or damage or destroy their Farm Data or Private Data.
- Implement a backup and recovery regime that is appropriate for the scale, sensitivity and timeliness of the Farm Data.
- Ensure that contingency plans exist to return or delete data (as per Principle 'e') in the event of insolvency.
- Ensure all staff and sub-contractors are trained to comply with the terms of this Code.

g. Compliance with National and International Laws

Where Providers are required by law to provide information to a third party, the Provider will:

- Avoid disclosing any Farm Data or Private Data; or,
- If Farm Data or Private Data must be disclosed, where legally permissible the Provider must promptly notify any Farmer whose identifying information will be (or has been – if prior warning is not possible) disclosed.

Appendix 3: FAIR Data

FAIR Principles

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship'[1] were published in Scientific Data. The authors intended to provide guidelines to improve the **Findability**, **Accessibility**, **Interoperability**, and **Reusability** of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

F1. (Meta)data are assigned a globally unique and persistent identifier.

F2. Data are described with rich metadata (defined by R1 below).

F3. Metadata clearly and explicitly include the identifier of the data they describe.

F4. (Meta)data are registered or indexed in a searchable resource.

Accessible

Once the user finds the required data, she/he/ they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.

A1.1 The protocol is open, free, and universally implementable.

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.

A2. Metadata are accessible, even when the data are no longer available.

1 See https://www.nature.com/articles/sdata201618

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles.

I3. (Meta)data include qualified references to other (meta)data.

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.

R1.1. (Meta)data are released with a clear and accessible data usage license.

R1.2. (Meta)data are associated with detailed provenance.

R1.3. (Meta)data meet domain-relevant community standards.

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).