KELLOGG
RURAL LEADERSHIP
PROGRAMME

Kellogg Rural Leadership Programme

Is open source the best way to manage and share data in New Zealand agriculture?

Course 45 2021/22

Sonja Vreugdenhil

I wish to thank the Kellogg Programme Investing Partners
for their continued support.

Strategic Partners

Programme Partners

Service Partners

Disclaimer

In submitting this report, the Kellogg Scholar has agreed to the publication of this material in its submitted form.

This report is a product of the learning journey taken by participants during the Kellogg Rural Leadership Programme, with the purpose of incorporating and developing tools and skills around research, critical analysis, network generation, synthesis and applying recommendations to a topic of their choice. The report also provides the background for a presentation made to colleagues and industry on the topic in the final phase of the Programme.

Scholars are encouraged to present their report findings in a style and structure that ensures accessibility and uptake by their target audience. It is not intended as a formal academic report as only some scholars have had the required background and learning to meet this standard.

This publication has been produced by the scholar in good faith on the basis of information available at the date of publication, without any independent verification. On occasions, data, information, and sources may be hidden or protected to ensure confidentially and that individuals and organisations cannot be identified.

Readers are responsible for assessing the relevance and accuracy of the content of this publication & the Programme or the scholar cannot be liable for any costs incurred or arising by reason of any person using or relying solely on the information in this publication.

Scholar contact details may be obtained through the New Zealand Rural Leadership Trust for media, speaking and research purposes.

Executive summary

Continual advancement of technology has created an excess number of data creation and management tools for use within agriculture. Multiple tools exist with similar purposes. Many of these are farm management tools or feed directly into the decisions associated with farm management. Reducing the number of new tools being created, would create for drive for information to be shared among the existing tools. Open-source data sharing creates a centralised data pool. Data shared in the pool could be used by other tools for more efficient data utilisation. Combining data from various tools can utilise data more efficiently rather than looking at data from individual tools in isolation.

Many challenges exist alongside creating an open-source platform. Firstly, current tools often do not communicate their data directly and automatically with each other. They rely on an intermediate, often manual step, to transfer the data by downloading and uploading data. This process is time consuming and can result in errors within the data. Secondly, agriculture is not well digitally represented. For a country and a sector which prides itself on innovation, the uptake of technology and data recording has struggled to move past the early adopters until recently. Within the current agricultural data landscape, not many farms have their own digital twin. Under representation of farm digital twins has been driven by a lack of data sharing between tools. Lastly, many farmers are worried that creating an open-source platform may allow for other people to have better access to their data than they do or specific data about location will be easily accessible. However, open source and more public data could help to bridge urban rural divide and demonstrate the work towards environmental stewardship which customers are demanding.

Currently, there is no leader in the open-source space within New Zealand agriculture. A lack of leadership has created a disconnect between the technology developers and those on farm implementing the technology. The evidence of the disconnect can be seen through the workstreams present within the agricultural technology community. These workstreams include data interoperability and targeting management of agricultural data by Trust Alliance NZ (TANZ) and AgriTechNZ respectively. Whilst these workstreams are being undertaken, the implementers on farm do not know these workstreams are occurring. Farmers are still looking for a solution to many of the issues associated with accessing and increasing utilisation of their data. The level of openness is viewed differently between the creators of the data on farm and those who may look to have an insight into what is happening on farm.

To create a successful open-source platform there are several pertinent workstreams to carry out. Firstly, creating a data interoperability standard, where all data can be shared from one tool to another with a small amount of translation. Secondly, defining multitiered data rights and access to the pool based on the contributor granting access. Both creating interoperability standards and appropriate data rights, removes the ability for competitor service companies to hold on to farm information as a competitive advantage. Plus, data about the farm could be passed on with a sale electronically rather than as paper records. Lastly, the creation of an open-source platform could ultimately allow for data sharing and automatic population of compliance documents. Open source could help to create one compliance audit on farm for all purposes, ultimately saving the farmer time whilst utilising data better on farm.

Table of Contents

Acknowledgements

Firstly, to The New Zealand Rural Leadership Trust, the work you do to provide continual development of leadership within the agricultural industry is essential to build the network of likeminded individuals. The diversity within each cohort creates a spread geographically and through a range of sectors. This work is appreciated by every past, present and future Kelloggers. Thank you for providing the opportunity to grow and develop through the ups and downs that COVID-19 has provided.

I would like to thank my employer, Leaft Foods, for funding my course and giving me the time off work to complete the Kellogg Leadership programme. The support to be able to complete the programme has been greatly appreciated.

To Dr Patrick Aldwell, thank you for your advice and leadership you provided for my project and pushing me to go deeper into aspects to improve the overall reach of my project. Dr Scott Champion, thank you for facilitating a great programme during the contact time we shared. Lisa Rogers, thank you for always being the cornerstone of the programme through tough times and multiple rescheduling's. Thank you for always ensuring we were always where we needed to be, on time and well presented.

To Cohort 45, you are an amazing group of people to get to know through the journey and duration of the programme. Thank you for sharing the highs and lows, the wins, and the struggles of the past 12 months. Thank you for reinvigorating and inspiring me with your passion throughout this journey.

Lastly, thank you to all the people who took time out of their busy days to help me grow my knowledge on open sourcing, talked to me about what open sourcing means to them, proofread my report, and provided support at multiple times throughout my time completing this project.

1.0 Introduction

As technology continues to advance, there are more tools created for capturing data in a range of different settings including supply chain, health and safety, water quality management and farm management. These tools are both in-field monitoring, for example weather or soil moisture, and tools to assist with decision making (Wolfert, et al., 2008). The number of tools being created for data collection has been driven from two avenues. Firstly, as farming has become more intensive, increasing the knowledge and granularity of production data has created a need for more data around per-animal and per-paddock production. Secondly, environmental regulation and compliance to meet regulation, are driving the requirement to have more information around the decisions being made on farm, particularly when it comes to water quality management. For example, Overseer is the main tool used for environmental compliance on farm. Particularly when dealing with nutrient losses from farms.

Farm compliance requires accurate record keeping on almost all decisions made on farm. Nutrient applications, effluent storage and biodiversity are some of the many activities which are audited as part of a farm environment plan (Environment Canterbury, 2021). As part of a farm environment plan audit, there needs to be evidence to support the decisions made on farm. These can be as simple as a spread map and temperature data from fertiliser applications, to as complex as soil water budgets to support irrigation application. Optimisation of this data collection through specific tools can aid in ensuring all the data required for the farm environment plan is collected. Data collected real-time and in-field, can save time and be more accurate than data collected retrospectively.

The information collected on farm can be underutilised or only stored within one tool. Sharing the data to another tool can allow for it to be better interpreted and utilised (Wolfert, et al., 2017). Currently, there is limited communication among the various tools used to collect data on farm. To be able to utilise all the various tools together often requires data to be manually pulled out of one and entered into another. There can be multiple tools which collect all the information required for the compliance audit. Hence, there will need to be information pulled out of multiple different tools and manually translated into the same format for compliance purposes. The format of the data may need to be changed again to be entered into a different tool to increase the utilisation of the data.

One concept which could solve the issue of entering data multiple times is to be able to open source the data collected. The idea of open sourcing data came to the forefront with the Internet of Things (IoT) movement. IoT describes the network of physical objects which are embedded with sensors, software, and other technologies that are used for connecting and exchanging data with other devices and systems over the internet (Wikipedia, 2021). Furthermore, to aid in exchanging data, data pooling can be used. Data pooling is a concept where data is shared into one repository from multiple different collection tools. This information could be useful for both the farmer themselves to increase the utilisation and interpretation of their data and potentially for other farmers to use for benchmarking against their own operation.

2.0 Aims and Objectives

The aim of this project is to define the term open source within agriculture. This will include determining what comprises open source, how it is established and maintained and what it would

mean on farm. Following the establishment of what open source is, there will be an investigation of whether open source is the best way to manage and share data created from New Zealand agriculture.

The desired outcome of this project is to increase the understanding of open source by the agricultural community. Subsequently, providing the initial steps for creating one storage mechanism for data sourced from a variety of data measurement and creation tools used on farm. These outcomes aim to help provide headway into the concept of data pooling into one repository. Additionally, creating data interoperability standards has the ability to have the data transferred among tools would help to assist in achieving the overall goal of this project. This creates the potential to be able to have one compliance audit on farm for various companies and purposes. Such as farm environment plan and meat and wool audits all done at one time rather than having multiple different audits spread across the year. These audits are commonly assessing similar aspects of the farm and farm management practises.

This project reviews the concepts of data pooling and data interoperability which are used within data pooling. These concepts are seen in other industries and their learnings can be pulled over to utilising the potential of open source in agriculture. The review was followed by a thematic analysis and use of the Delphi method. The review and thematic analysis evaluated the potential for utilising open source for information and data sharing within New Zealand agriculture.

3.0 Method

To complete this project, a literature review, thematic analysis and the Delphi method were undertaken. A literature review built on a report produced by AgFirst in 2011 evaluating the farm management tools available to dairy farmers (Allen & Wolfert, 2011). Several other concepts were introduced in the literature review to build on the findings of the AgFirst report. These concepts were digital twins, data interoperability and property rights, which add up to provide the baseline for the concept of data pooling. The information collected through the literature review was tested with industry experts used the Delphi method. There were 12 private conversations undertaken with a range of experts in the technology and agricultural industries. The conversations were analysed using a thematic analysis and trends pulled out of the literature review to understand what is going on within the New Zealand landscape already. The themes identified were used to evaluate the overall goal of creating one compliance audit for the entire farm.

4.0 Literature Review

Farm management software and tools come in many different shapes and sizes. In 2011, there were 127 different tools available to dairy farmers for various aspects of farm management (Allen & Wolfert, 2011). There tended to be more tools in areas which were well known such as finance, and fewer tools on less well understood areas such as labour management. Similarly, May (2015) compiled a list of seven different recording programmes available to the arable industry. They concluded that technology uptake within the arable industry had been slow, due to the decision making of which system to utilise on farm. Unfortunately, the Centre of Excellence in Farm Business Management which commissioned the report carried out by Allen & Wolfert (2011) has not had any new work published since 2017. The work from the Centre of Excellence has primarily focused on the dairy

industry (Lincoln University, 2021). The focus on the dairy industry could be because of more frequent and intensive measurement on farm. This is likely driven by the frequency of animals in the milking shed and tight management of pasture from high stocking rates. However, arable farmers can have detailed yield maps, stimulating tool requirement. There does appear to be a missing workstream for the sheep and beef sector in this space.

Overseas, Australia is a frontrunner in this space. There is a database which is continually updated with the new tools coming into the market at a variety of different levels for on farm (Farm Table, 2021). The Australian Agritech Map demonstrates 18 different groups of tool types available in Australia (Australian Agritech Association, 2021). The number of tools available for farm management continues to increase and the complexity of each of tool varies depending on what they are measuring and the frequency of measurement. These groups of tool types, within Australia, range from whole farm management, waste management, livestock through to mapping, marketing, connectivity and retail. Each of these companies and tools only feature once within the ecosystem demonstrating the choice which each individual has when selecting the best option for them. Australia has become one of the leaders in mapping tools available to farmers. New Zealand could use some of the ideas which Australia has carried out to keep up.

An aspect in the slower nature of New Zealand tool uptake and definition could be associated with the complexity of the various farm systems. Figure 1 shows how complex these interactions are on a farm, hence the number of tools developed. These tools can be specific to areas of the farm, farm types and/or areas of farm management rather than having the breadth to cover multiple of these areas.

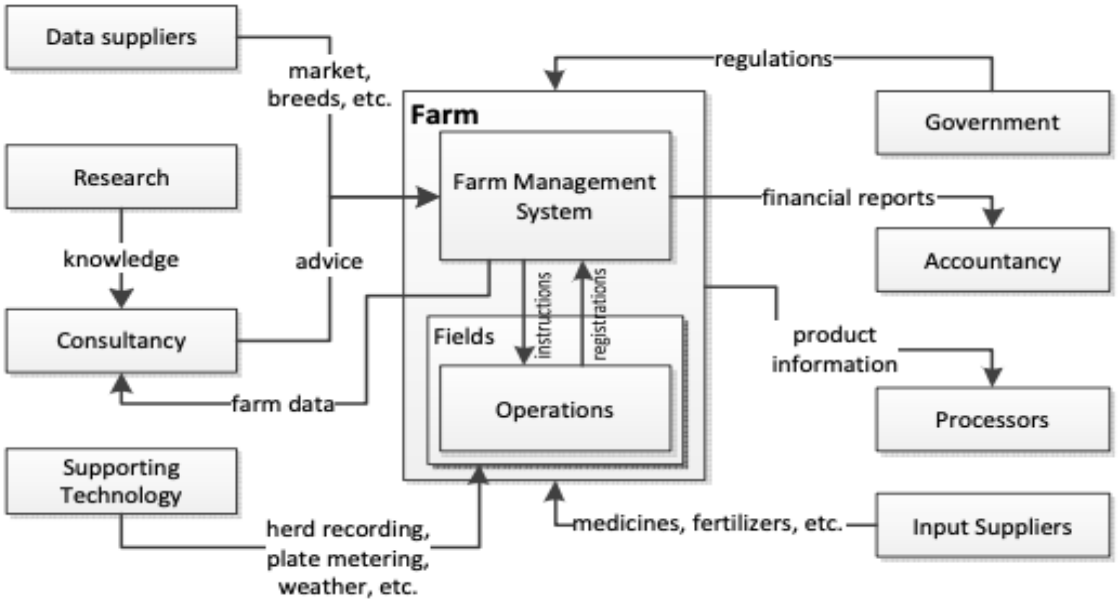

Figure 1: Representation of the information flows within and around the farm (Allen & Wolfert, 2011).

Farm management is complex. There are multiple aspects of farm management which do not tend to overlap. A lack of overlap has resulted in segregation of tools used within agriculture and difficulty to create one tool or agglomerated data for total farm management. Figure 2 highlights the number of

segregated areas involved in farm management. Many of these tools collect and store similar information and selection of which tool to use becomes based on price, ease of use and whether it shares or receives data with the other tools used on farm. These areas on farm and potential for tool utilisation are not unique to each farm system. Instead, they are universal with some small subtleties based on the type of animals grazed and crops grown. This review will focus on the concepts which sit within the ability to share data among tools.

Figure 2: Types of management required for farm management and the tools that could be used (Allen & Wolfert, 2011)

4.1 Digital twins

Digital twins are the virtual representations of a physical process or object. Digital twins are common across the manufacturing, automotive and energy sectors (Pylianidis, et al., 2021). These sectors tend to have diagrams of how equipment or a process works. By having a virtual replication of equipment or process, this allows for a mechanism to be able visualise the inside of what is going on. Additionally, there tends to be engineers involved in the construction of the equipment or processes used within these sectors. Engineers are known for their way of thinking and small margin of error accepted in designing and building equipment and processes. Hence, it makes sense that there are multiple digital twins within these sectors. Comparatively, agriculture does not tend to have processes or systems which can be easily controlled. The majority of variation within an agricultural system comes from the climate, which is difficult to predict, let alone control. Hence why agriculture is relatively new to digital twins at the system level. Pylianidis, et al. (2021) reported 28 digital twin examples within agriculture. These twins are the first steps to expanding the concept and uptake within agriculture. The currently digital twins are simple, such as the irrigation system. With more developed knowledge and technology, the ability to gradually map more complex systems digitally increases.

Digital farm maps are an example of digital twins on farm. However, these are designed to show only the boundaries of the farm and paddocks. There does not tend to be entire farm digital twins which show the entire farm system and all the assets in real-time. To be able to increase the complexity of the digital twin of the farm, there would need to be multiple digital twins feeding into one farm map.

However, the data collected in these various tools may not all end up in the same total management tool. Hence the farmer often must pull these pieces of data together to get the full picture of the digital twin. Allen & Wolfert (2011) highlighted the number of tools and the different types of information that they collect. Yet, none of the tools available in 2011 did the full farm digital twin. Similarly, the tools highlighted by May (2015) did not offer the whole farm digital twin either.

Digital twins are becoming a new tool to be used in the supply chain. A key benefit this could have is reducing the loss of fresh produce during transportation and storage (Defraeye, et al., 2021). Creating a digital twin representing the conditions which the produce is sitting in, combined with the age of the produce would indicate how long it would have until it spoils. The difference this could make to the supply chain is altering the order in which produce is shipped or goes to the shelf. The use of this monitoring within the supply chain is another mechanism by which digital twins could be introduce to agriculture. Specifically, when looking at pasture or crop growth rates within a paddock which dictate the grazing rotation by animals. Paddocks that are growing faster could be grazed more frequently. Poorer performing paddocks could be targeted for regrassing or selective nutrient management quicker and more effectively. Rather than always following the same rotation and adjusting the number of days grazing each paddock.

4.2 Data interoperability

Data interoperability refers to the ability for data, in different forms, to be translated among different tools. This also comes with shared expectations for the contents, context and meaning of the data (Data Interoperability Standards Consortium, 2021). The ability to have all tools that share the same rules for data sharing could help to assist with increasing the ability for data to be shared. Allen & Wolfert (2011) highlighted that some of the key issues for tool uptake were associated with there being no common rules of software design for data sharing. Hence, the requirement for the guidelines of data interoperability.

Many countries already have the foundation guidelines for data interoperability. Due to New Zealand being an export nation, there tends to be more data transformation required before the product reaches the consumer. Data transformation is the process of changing the format or structure of the data to make it usable to others. Due to the amount of transformation required, it would make sense that the creation of guidelines for data interoperability would be quicker. Due to the number of companies and brands within the supply chain, the data is manually translated between systems and likely has additional data added to it before it reaches the next step in the supply chain. DairyNZ & Rezare Systems (2021) have started the data interoperability guideline formation process for New Zealand data management by developing a code of practise. It provides guidelines of what would be required to be accredited as a member of the code of practise. Plus, guidelines are needed about who has rights to the data and how the data is shared and stored with the aim to provide confidence to farmers that their data is safe and managed appropriately. Formulation of a code of practice would be a large step forward for increasing data sharing capability.

4.3 Data pooling

Data pooling combines the concepts of digital twins and data interoperability. These concepts allow for data to be pulled out of each tool and be stored in a central location. The pooled data can be pulled

out by other tools to be incorporated into the data within the tool. The theory behind pooled data methods requires the parameters which encase the data to be identifiable (Jacquez, 1996). Parameter identifiability allows for there to be knowledge of where the data fits within the pool plus which tools can appropriately utilise the data. Similarly, data when pooled, needs to be comparable in the context of the various ways to utilise the data (Friedenreich, 1994). This links to the requirement for data interoperability standards for storing and sharing the pooled data.

Conflict can be associated with pooling data when it comes with data rights and ownership. Farmer and industry interviews carried out by Keenan (2019), regarding digital technologies and the future of farm compliance in New Zealand, showed that there were data ownership questions as technology was evolving in the early stages of farm compliance. With the continuation of technology improvement and increasing access to data, there are likely to be even more concerns about data ownership. The New Zealand Government front footed some of these concerns, releasing an article saying they would establish a consumer data right specifically related to data holders to share data safely and securely with third parties, following consent from the customer (Clark, 2021). Consent becomes a large part of being able to share data with desired parties to ensure the data is handled appropriately. Data created on farm, could also be shared securely to be able to utilise it to the full potential.

Due to the increase in tools available and data collected, the concept of big data can be supplemented alongside pooled data. Big data refers to data that is large, complex and a combination of data types. It is difficult to analyse with traditional methods. Wolfert, et al. (2017) suggested that big data could change the relationship among various tools in the food supply chain. There were two extreme scenarios on a continuum suggested, open and closed, regarding information and sharing platforms. Figure 3 shows the drivers and the challenges involved in big data systems. The preferable situation would be where there is more open information sharing, however this is likely to come with more challenges across network management, stakeholder network, business processes (Figure 3). Comparatively, a closed system could potentially have less challenges, due to lower requirements for stakeholder engagement and data interoperability among the various data creators and holders.

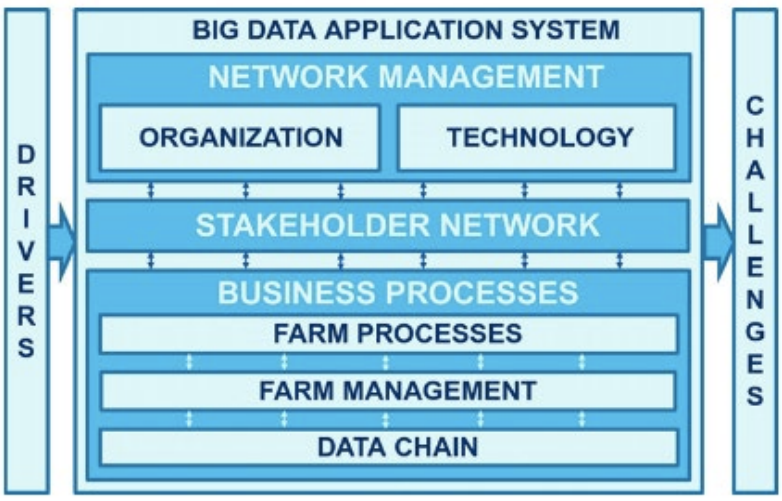

Figure 3: Conceptual framework for big data application system (Wolfert, et al., 2017).

New Zealand has some examples of data pooling. AgYields National Database is a pool of pasture, vegetable and cereal yields grown in New Zealand (Lincoln University, 2022). The data is pulled from published literature and can be filtered by location and species. Ultimately becoming a tool to better predict yield. Comparatively, FarmIQ could also be seen as a data pooling tool. It acts as a platform to bring all the information together on farm (Farm IQ, 2022). Additionally, it allows for a planning, sharing and recording data followed by reporting, compliance and reviewing performance on farm. These examples already available in New Zealand set the framework for other venture which can set up new open-source tools or convert existing tools.

4.4 Property rights

Pooling data from various locations and entities can create conflict surrounding ownership and rights to the individual and pooled data. Ownership of the data can be much more difficult to define for data compared with rights. Ownership would require the data to be more physical. Rights to the data are easier to define when the data sits within a shared storage complex. Data management companies often require access or rights to the data as part of their ability to manage the data (Kpienbaareh, et al., 2019). As part of the ability to pool data, rights would need to be granted to the other tools which utilise the pooled data (Wolfert, et al., 2008). Granting right to the data can create tension between the original creator of the data as those needing access.

Appropriate control must be provided to contributors of pooled data to ensure they can access their data. Figure 4 demonstrates the various levels of control which can be managed within. They can be private, where they are accessible only to the owner, shared, accessible to all group members or public, accessible to everyone (JanBen, et al., 2011). Comparatively, in these models, summary information and aggregated data is often made public while detailed information remains private (Gallacher, 2021). Providing various levels of control allows for data to be shared selectively within the tool and has the benefit of keeping the more personal information, such as where precisely the data was collected, private. Increasing the privacy around exactly who created the data can reduce the fear that the data will be used against them (Rutherford, 2021; Whittington, 2021). Additionally, having the opportunity to keep the data private from others would solve the risk of conflict among the different contributors of data and the tools and users wanting to access the data.

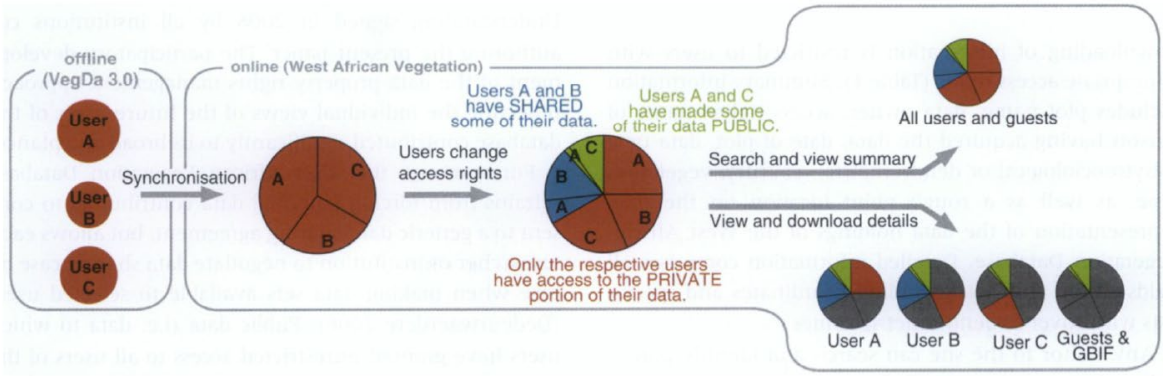

Figure 4: Data property rights management in the West African Vegetation Database. Data sets can either be private (red, accessible only owner), shared (blue, accessible to all group members) or public (green, accessible to everyone) (JanBen, et al., 2011).

4.5 Benefits for the industry

Agriculture could benefit from increasing data availability and utilisation. Through open source, the creation and utilisation of big data could help predictive insights within the farming operations (Gallacher, 2021; Whittington, 2021; Wolfert, et al., 2017). Subsequently, some tools could provide real-time decision making and assist in business models. Progressive farm businesses require forward thinking compliance. Keenan (2019) brought forward the idea of creating one farmer centric plan from FarmIQ which encompasses a farm management plan, decision support tool and stock reconciliation (Figure 5). Additionally, these forward-thinking businesses also use the data they produce on farm and benchmark themselves against others, whether it be industry standards, published literature or conversations with the neighbours (Rutherford, 2021). Increasingly, farmers are being seen to have a duty of care to show how they are managing their environmental footprint (Gallacher, 2021). The ability to demonstrate farm practices which fit within the duty of care and social licence could pull the industry forward when it comes to perception and to a certain extent, bridging the rural urban divide.

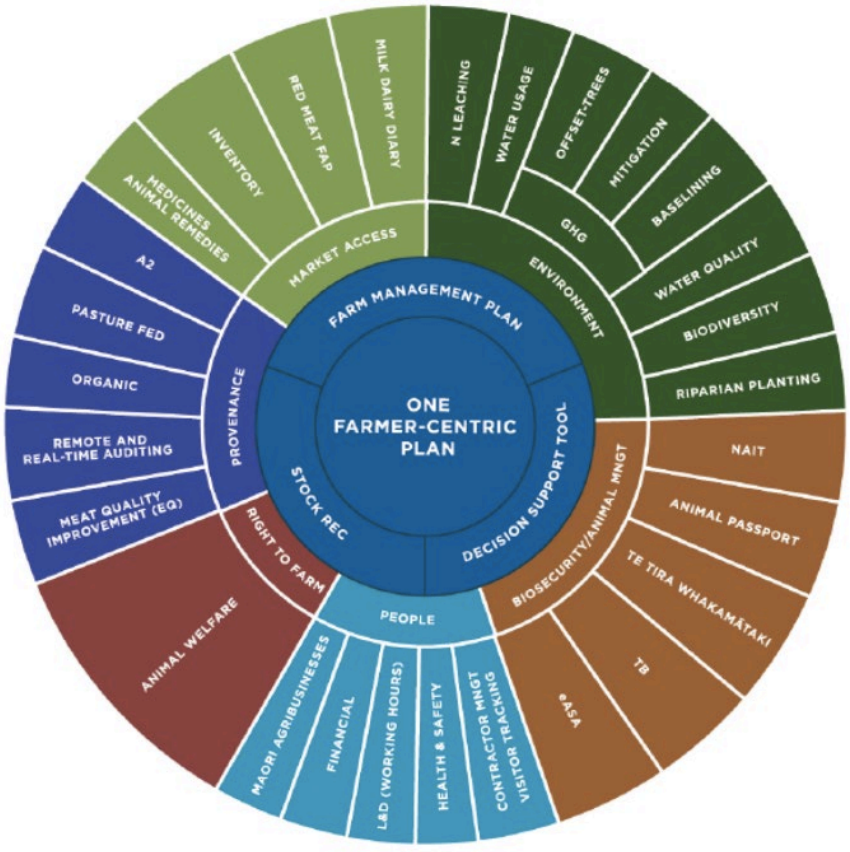

Figure 5: A guide to integrated farm planning (Keenan, 2019).

5.0 Findings and Discussion

Data utilisation within New Zealand agriculture has previously been low. Comparatively, New Zealand is seen as an innovative country. Few of the current tools utilised on farm can communicate with each other creating siloed data. To reduce data silos on farm, open-source data sharing could be utilised. To aid in the development of an open-source platform, there needs to be several developments within

how data is shared and how the farm is digitally replicated. The creation of a digitally well represented farm requires increased data sharing between tools. The benefit would be for each farm to have all their data in one place. Subsequently, creating a better farm data portfolio which could be shared as a full picture of the farm system. This would move the New Zealand agricultural industry one step further to being able to have one audit for all aspects of compliance on farm.

5.1 What does open-source mean on farm?

Open source can be seen as having two definitions, one which is more open than the other. Traditionally, within the technical data handling landscape, open source is referred to as being completely open for everyone to view and access the information (Kpienbaareh, et al., 2019). Internationally, agriculture could be more open but there are concerns from the owners and collectors of the data that it could be too open and at risk of exploitation by extremists. Additionally, having the full story out in the open, may not provide the best picture for the consumer through risk of misunderstanding the processes behind production and collection of the data (Pfeffer, 2021). Data collected on farm is incredibly valuable to the farmer and to the industry and the risk of misinterpretation could be detrimental to the agricultural industry. An example of this could be the already present misinterpretation about shearing sheep. For many, this is an animal health requirement, but for others this could be interpreted as inhumane because of the physical handling and risk of cutting the animal.

Open source, within agriculture, is commonly referred to as farmers having access to all their data that they collect on farm and the data about their farm and animals. Additionally, open source would give them the permission to share the data with whom they desire whether it be a tool, individual and/or entity. In this version of open source, each farm would have their own vault for the storage of their data. Then the farmer could grant access or transfer data to others. True open source goes a step further than one vault per farm. There could be one vault for all data within the industry. Having one vault which contains all the data from all farms and all the data from all the tools. A singular vault would be the first step towards data pooling on farm. Subsequently, data pooling could be the beginning for automatic pre-population of other tools or compliance audit framework.

Open sourcing outside of the individual farm gate could be difficult to navigate. Anonymity and/or strict privacy would be key in assisting in the open-source journey when taking the data outside of the farm (Rutherford, 2021). Keeping farm data anonymous could allow for pieces of information to be pulled together from different farms to go alongside the product marketing to consumers. Anonymity could provide the starting point for data sharing beyond the farm gate. To assist in anonymity, data interoperability standards would be required to share the data easily and the property rights would protect data owners' identity at the individual level. Benefits of pooling the anonymous data, could allow for the original data to be built on through the addition of new information and understanding (Rutherford, 2021). The added interpretation of the data could benefit the farmer and information collector, the tool provider, and the consumer. Ways in which the data could be built on could include the creation of subregion specific benchmarking data based on the mean values within the pool. Specificity in this benchmarking data could increase knowledge of performance without knowing exactly who the benchmark is.

Anonymity works well where the data can be interpreted without the exact location of the collection site. Geographic data is difficult to anonymise as the data is at a very granular level, hence there are only a few farms which the data could be from. Geographical data is likely to be information that would require data interoperability and permissions to share this data with others. Regionally, most farmers know their neighbours and the surrounding farms. However, they do not tend to know their exact production information. Hence, geographical data alongside the in-depth farm metrics in the data pool, could create conflict with being able to see in depth information about every farm. Reducing geographical granularity in the data pool could get around this problem.

Prior to creating the open-source model on farm, farmers need to be able to access all their own data. Storing data online rather than individual spreadsheets would begin to create a vault for the individual farm and increasing utilisation (Pfeffer, 2021). However, not all data is currently easy to import and upload. To assist farmers creating their own vault, data interoperability standards would help with getting all the data into one format for ease of handling. Plus, the data collected on farm would be freely available to be shared with others within the industry.

5.2 Data within the agricultural industry

New Zealand agriculture plays a role in feeding the world. As a country 95% of dairy products, 87% of beef and 95% of lamb produced are exported (Hancock, 2021; MPI 2022). Increasing productivity has been driving the industry forward until recently. Increases in production have shifted from increasing input to increasing output. To increase production efficiently, data is required to compare previous production with current and see the response per unit of input (Pfeffer, 2021). Due to previous low levels of data collection on farm, issues have arisen which have not been detected early on. Examples of these are disease and herbicide resistance, and pasture and crop production yield increase per unit nutrient application. Inefficient nutrient application has been a contributing factor for some of the ground water pollution. Since these issues have been identified, there has been increasing push from consumers and the government to improve the efficiency of production rather than total production.

Tools used for measuring production efficiency can be either continuously recorded, recorded at set intervals or linked to recording certain events. As technology has continued to advance, the tools used for information capture has become more sophisticated. For example, collection of weather data has been occurring for long periods of time. Early on, people could stand outside and see that it is raining, windy, hot or cold. However, these observations were not easy to quantify. Now, weather monitoring is more automated with weather stations carrying out constant measurements. These measurements rely on further interpretation and comparisons over time, both seasonally and annually to get the full benefit of looking at trends. Similarly, milking sheds now have weigh scales linked to cow EID tags and that collect the weight of each cow at every milking. However, this information, needs to be interpreted to identify trends in the data. Identifying trends is a manual task, requiring downloading and uploading the data in an additional processing tool. Due to the manual nature of identifying trends, and lack of additional processing tools which can handle other tools data, there is a massive underutilisation of the data collected on farm.

Due to the slower uptake of data collection and further utilisation of the data on farm, much of what is collected is not in a state or form to be shared. Data collection, until recently was very paper based (May, 2015). Therefore, when a farm is sold, the information collected on farm, such as nutrient

records, soil tests, pasture growth and regrassing history, are not east to pass on with the farm because they are stored in diaries rather than on the cloud (Whittington, 2021). As more information recording moves towards the cloud there are more opportunities to pass this data on through granting access or changing ownership of the digital twin of the farm.

Supply chains require data collection for traceability. Part of the New Zealand provenance story will become the traceability of products from paddock to plate. To increase the uptake of utilisation of technology and better interpretation of data may require more of a lead by example mentality from farmers and industry (Whittington, 2021). Rather than the perception of being backed into a corner and needing to collect the compliance information to be able to keep farming (Gallacher, 2021). Solutions offered on farm need to be specific, and potentially have the capability to interpret the information collected. Therefore, the information and data collected will directly assist with decision making and ultimately storytelling, rather than only validation of decisions.

5.3 Workstreams to assist in open source

New Zealand pastoral production systems are some of the most efficient. Comparatively, agricultural technology (agtech) uptake is slower on farm. Due to slower uptake and less data collection on farm, there has been a lower requirement for data interoperability standards for sharing data in New Zealand. Comparatively, Europe already has data interoperability standards. This could be due to the close link of the supply chain from the paddock to plate. As the New Zealand agtech industry continues to develop, there need to be workstreams on the data management and sharing to assist in data transfer. Assisting data transfer should increase uptake and utilisation of agtech on farm by farmers due to less time commitment required for data transfer.

AgriTechNZ and Trust Alliance NZ (TANZ) are the two leaders in workstreams for data interoperability and data management in New Zealand agriculture. TANZ are a membership consortium for producers, growers, exporters, retailers, and service providers to easily share trusted data. AgriTechNZ are another membership organisation integrating agriculture and technology to strengthen the primary sector. Both are focused on the improvement of New Zealand in their specific contexts.

Workstreams by both providers focus on aspects which will provide benefit of the entire industry. TANZ have workstreams focussing on data interoperability to be able to turn compliance into value. The data interoperability workstream focusses on digital identities, creating a digital farm environment plan, sensor networks for more sustainable farming, optimising the value chain and lastly building on the provenance and consumer story. AgriTechNZ have workstreams which are more targeted at the management of agricultural data. The projects being undertaken are about development and custody of agricultural data standards and creation of a pan-sector baseline for digital adoption in the primary industries (Precision Agriculture Audit). These projects are supported by the Practitioner Working Group continuing the Precision Agriculture Associated of New Zealand's initiative. Additionally, The New Zealand Government have released a discussion document called Towards a Digital Strategy for Aotearoa (New Zealand Government, 2021). Which focusses on enabling all of Aotearoa New Zealand to flourish and prosper in a digital world. There is no specific mention of agriculture in this document, but the goals of trust, inclusion and growth all tie in with maintaining the agricultural industry. Digital adoption baselines within the primary industry could assist with ensuring

that everyone has equal opportunities and access. This is also known as the Precision Agriculture Audit as part of the AgriTechNZ workplan. However, it is also another form of compliance.

Within the data interoperability workstream, is the method of how the various data platforms are identified by the user which is part of the work being carried out by TANZ. Each platform tends to have its own login (identity) (Figure 5). A single user can have multiple identities, each with their own associated tool and data. TANZ have proposed with some of the work that they are undertaking, that a single user has one identity which gives them access to multiple tools (TANZ, 2022) (Figure 6). The benefit of the single identity per user would reduce the number of logins a single person has but should also assist with the 'usefulness' of the data for decision making and not replicating data to be able to use it elsewhere. Comparatively, at a national level regarding Real Me as a unique identifier for each person, is being compared to other national programmes run by the government have had large volumes of money sunk into them (Stuff, 2021). There is a potential risk that some players see the same risk in the development of a unique entity for management of agricultural data due to the cost required for establishment of the standards required for this to happen.

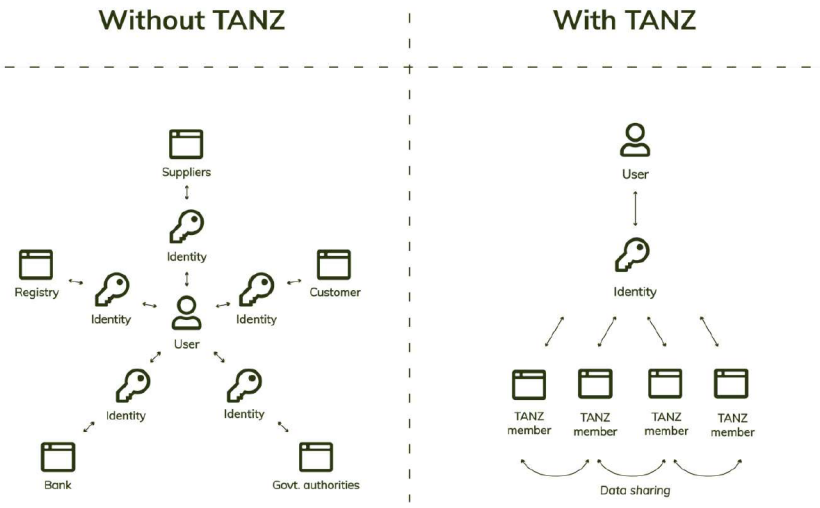

Figure 6: Influence of shared identity on data sharing (TANZ, 2022).

Data interoperability, with a specific focus on agriculture, could assist in improving the knowledge sphere of data sharing off farm. Most consumers have the desire to understand where their food came from but do not have a source where they can ask these questions. Plus, marketing of products is more of a generalised overview rather than specific insights into the actual farm the product was produced on. This could be a driver from both ends to the creation and furthering of the rural/urban divide. The creation of data interoperability standards could help for increased data sharing to give consumers more in depth information about where their food has come from.

Typically, agriculture is closed off when it comes to sharing how the product was produced on farm. Specific production parameters and how the product was grown, and fed are not shared outside of the farmer and those close to them offering advice. Fear of extremists could be a reason why farmers do not want to open their gate to the world to show how they produce food (Gallacher, 2021). However, compliance requires in depth information sharing with on a wide range of topics about how the farm operates, from nutrient management through to biodiversity. Being able to combine the information collected for compliance with marketing for the product could add more value to the

products going off farm. In this sense, using compliance as a tool to not only check in with how things are being done, but to also add value to the product for the consumer through traceability could financially compensate for the time it takes to collect and record the information required.

Some of the early work in New Zealand on agricultural data management has been through DairyNZ and Rezare Systems. The two outputs have been Farm Data Standards (DairyNZ & Rezare Systems, 2021) and DataLinker (Rezare Systems, 2021). Both were developed through separate initiatives to help agricultural data to be more accessible to farmer. Plus, allowing data to be used more effectively by service providers to the farmer. These outputs provide insight about more efficient utilisation of data on farm, yet there are probably few farmers that know these workstreams exist. These projects were funded through levy organisations, the government, meat processors and banks. These groups have tight links to agriculture and are ultimately funded through agriculture. Hence the target audience should have more communication about what is happening in the background to assist with data management.

5.4 What is currently happening on farm?

Data collection and data recorded on farm do not always go hand in hand. There is a common saying that you cannot manage what you do not measure and if you do not measure then you cannot improve. This is particularly true within agriculture. Due to the large area which farms cover, it is difficult to constantly monitor everything going on (Whittington, 2021). Additionally, data collected on farm can be sparse. Where the data collected is averages for periods of time, and the resolution of this data is low. Increasing the resolution of the data by collecting data on a per hectare basis rather than per paddock or a per animal basis rather than per mob of animals. Production measurements like these tend to be driven by the farmer wanting to increase the efficiency of their farm. This data collection may also be used to assist in decision making for the following season rather than the current season as the analysis may happen retrospectively.

Real time data collection can be more difficult to integrate into the farm system. Farms tend to be multi-generational where information regarding production is passed on through word of mouth and learned experience. To have real time assistance in decision making, the information collected needs to build on the learned experience of how the farm and animals are behaving that is different to previous. With the change in climate, monitoring soil moisture is a key tool which can add to the efficiency of the farm. Irrigating at the right times with the right amount can keep pasture growing for longer than without early detection of soil moisture depletion. Within different groups of farmers there are differences in the uptake of soil moisture monitoring. Innovators and early adopters will move towards new technology to assist with decision making. Comparatively, others will be pushed towards adopting technology for compliance purposes. Soil moisture probes provide site specific measurements compared with soil water budgets which rely on measurements from the nearest weather station. Ultimately, specificity creates accuracy.

Farm management compliance has become a large part of the farming operation. Increasingly decisions made need to be validated with data rather than years of farmer experience. Aspects such as soil moisture are required as part of the decision-making process for irrigation to ensure that the soil is not overwatered risking unnecessary nutrient loss with unexpected rainfall. Additionally, proof of nutrient placement, timing of application and total applied nutrients are required for compliance

too. These pieces of information are required in multiple places, including Overseer for nutrient loss compliance, regional compliance for total application and in farm environment plans. This information is generated in the nutrient supplier's management tool which does not communicate with any of the other tools or places that this information is required.

Due to the requirement of additional data collection for validation of decisions made on farm, more tools are being utilised on farm. However, when these tools not communicating with each other, this means that the information collected need to be re-entered every time the data is required. Comparatively, farm management tools should pool information from a variety of different tools to be able to view the big picture across different farming operations. Being able to do this would assist in decision making rather than coming from the validation end of the spectrum. Some large operations would like to see the farm management data at the resolution of being able to see the cost to grow grass in each paddock as an analysis tool. Unfortunately, there are no tools which currently provide this service, either through lack of communication from the other tools used on farm or not having the processing capability within the tool itself. Opportunities to rectify the issue around extra layers of information utilisation would be further building of the tool for processing or developing some rules for information sharing.

Data interoperability rules would assist in the communication of information between tools to reduce the time requirement of the farmer to replicate data among tools. Currently data is downloaded from one tool then uploaded to another in an excel csv file. These files are used to import large lists of data or reduce the risk of import errors. Other excel files are at risk of becoming corrupt during this process. Many companies have an application programming interface (API) which shows a large amount of information but cannot be downloaded or transferred to another tool. The data interoperability rules which could assist in the data being transferred from one tool to another. However, the one benefit of a excel csv file is that many tools use this same format for their data. Some of the early data interoperability is within the use of the same excel csv format for uploading and downloading data.

5.5 Drivers for data sharing between companies

A lack of data sharing could result in a competitive advantage to some companies for customer retention. Data can be described as the new oil, as it has started out rare but is now a necessity to keep the world operating. Information collected and retained by companies required the competitor to have to recollect the same information for them to be able to provide the same service as the original company, reducing the competitiveness of the farmer during this time (Rutherford, 2021). Within agriculture there are no guidelines about sharing information for the benefit of the farm and their time. It tends to be focussed on the company which collected the information or supplied the service.

Data ownership and rights can be difficult to navigate when it comes to sharing data. Additionally, data ownership is difficult to define, as data is not physical (Pfeffer, 2021). Plus, it is possible to have several levels of ownership. Firstly, the raw data often requires some interpretation to become useful whether this be through translation or processing through another tool. This translation may add some intellectual property to the data, giving the translator some rights to the data. Secondly, the data owner could also be defined as the person who put the effort in when collecting the information. Again, another difficult situation as there are multiple occasions where someone else collects data on

behalf of the farmer, paid for by the farmer. (Rutherford, 2021) The exchange of money for the service, creates shared ownership of the data. Lastly, lawyers are obliged to hand over all the information they have on a client if they chose to leave a firm. Comparatively, within agriculture, there is no obligation to share data over to the competitors if a customer leaves. Tying in closely with desires to retain customers through data retention.

For data sharing among companies to become a requirement, including direct competitors, there would need to be a group established to do the data translation or a large player to lead the way. John Deere have been an international leader in this space for the broad service which they supply to their customers. Their expertise and services they provide come from the information they collect in the paddock associated with their machinery, which adds to the agronomic advice they can offer from anonymising and agglomerating the data they collect (John Deere, 2022). Similarly, CropX started out as a water management tool, looking to move into nutrient management progressing to farm management (CropX, 2022). To be able to continue to grow the management tool, more data needs to be collected on farm and processed through the tool. Some farmers already have the appropriate tools on farm, only requiring data to be shared, while others would need new tools to be installed. CropX are in a unique position where with the right partnership, their tool could grow rapidly. There is likely to be push back from other industries as they could be competing for the same data. By being able to share the data collected by one tool on farm with multiple data processor and farm management tools there would be few negatives for the farmer.

5.6 Value add for the farmer

Farmers tend to be time poor when it comes to the day-to-day operation of the farm. There are always multiple jobs which need to be carried out, which may be more important than doing data collection exercises. Data recording can be time consuming to see little benefit in the early stages (Whittington, 2021). To get good data collection and utilisation on farm, data collection needs to be easy and somewhat automated to remove the burden of data collection from the farmer. Additionally, removing the requirement to enter and re-enter data, when transferring between different tools, would be another time saver.

Automation of recording tools makes data collection easier. Currently, data sharing involves entering and re-entering data, it can be a time-consuming task. When the data is only being utilised for compliance data collection can seem like an unnecessary task unless it can be repurposed. Repurposing could include investigating monitoring system efficiency and identifying areas for improvement on farm. Additionally, aiding in decision making from relevant and usable data. The next step within the data utilisation would be to combine data from a variety of tools automatically.

Farmers do not tend to have a particular affinity for spending excess time in the office. Plus, compliance, data recording and transfer require time spent in the office. A key value add for a farmer is to get time back on the farm rather than in office. Mentally, compliance can take a toll on farmers, especially with the volatility of the ruling changes occurring in recent times. By being able to get back outside can help to bring the joy back to farming. Balancing the compliance requirements with being able to be out there doing it, can provide opportunities for visualisation of the changes that would need to occur.

Reducing the number of visits to carry out various compliance activities can also help the farmer to regain time back outside. An integrated farm plan, holding all the information required for all compliance could help to reduce the number of audits. One integrated farm plan may take longer to create in the first instance but should take time off in the long run. An integrated farm plan could also help to highlight the subtleties in the farm system which are often hidden when only looking at one aspect of the farm. Regardless of whether it is the consumer or the manufacturer, there is increasing demand to have more traceability and ability to gauge the authenticity of the product. Neither group determines whether it be one audit or multiple to determine product traceability. Additionally, looking at the full picture can help with the provenance marketing strategy which is used for New Zealand products.

New Zealand's provenance can be seen as a large selling point for exported products. Being able to continue to build on the New Zealand story through improved data sharing could allow for the consumer to have a deeper insight into how the product is produced. The production story is becoming increasingly important to customers and consumers, especially the environmental footprint associated with production, as it adds value (Gallacher, 2021). All this information is already being collected on farm but not being shared further as part of the environmental compliance required. However, the environmental compliance story does need to move further than the farm records and auditing process. Creating a singular integrated farm plan where the information and data on farm could be shared through to the customer and the consumer could be a value-add opportunity for the producer and the manufacturer. To make the added data collection and compliance worthwhile, the value would need to be received at the farm level. Due to the time commitment and potential added cost for the data collection, receiving value at the farm may be the compensation required to aid in uptake.

5.7 Potential issues to uptake

Lack of sharing between companies, competitors or partners, shows that there could be multiple early opportunities for data sharing. Without the holders and collectors of data being involved in the process, it has been difficult to get traction with data sharing. Due to the current climate of data sharing, there is the requirement to have an additional group in the chain to provide the translation of the data. Currently there is no group ready to lead in this space. There needs to be a leader to begin the transition toward an open-source data sharing system. The leader would either need to come from the regulatory side to encourage others to follow or be a large player that could out compete the smaller players if they do not keep up.

Property rights and continued access to data whether it be pooled, or an individual's data could provide a challenge in the future. Maintaining the rights of the groups/individuals to their own data is important to ensure that data can be utilised. Additionally, the data creators need to feel as though their data is being adequately protected. The various levels of data protection can assist with uptake and prevention of hesitancy towards open source. An inability to manage rights to the data has the potential to create a barrier to uptake. The data needs to be in a place where it is well protected even though it is in a position where it can be shared with others.

Connection between different organisations for data sharing requires data translation. The data translation through data interoperability is the biggest requirement to get data pooling happening. The creation of a universal standard, both within the country and internationally, would by default

create a system where the data could move into the pool or another tool. Without these universal standards, creating uptake of data pooling will be substantially more difficult because there will be no autonomy.

6.0 Conclusions

Data sharing has the potential to revolutionise the agricultural industry. Current underutilisation of data by both the farmer and the industry highlights the massive gains which can be made. Increasing the resolution of the data collected on farm can aid in more data being available. Collecting more data, over a longer period and from multiple locations, can create big data. Unfortunately, the data collected tends to be held within the individual tools and the data is difficult to remove. Sharing big data could add to the usability of the information. However, the lack of data interoperability standards and appropriate data rights make access to the data difficult.

Creating better access to data is likely to come from better rules and guidelines around the use and sharing of data. Data interoperability standards and strict data rights are two key pieces of the data sharing puzzle. Creating data interoperability standards and data rights, would form the baseline of the rules of open source. Whilst the access and coding which sits behind the ruling come with costs which must be met by someone. To assist with uptake data sharing and collection needs to be easy for the farmer. By making these processes easier than they already are, will reduce stress prior to compliance.

Agricultural compliance requirements and difficult for farmers when they require multiple audits for similar aspects of different parts of the farm business. The ultimate goal of creating one compliance audit on farm will take multiple step changes in the data sharing process. All these step changes are possible from both the technology sector and the agricultural sector.

7.0 Recommendations and Next Steps

Extension to farmers

1. Increase communication to farmers regarding technology development and digital workstreams by including industry bodies and farmer advocate groups on the steering groups of these workstreams. Having both groups (AgriTechNZ, TANZ, Federated Farmers and a consultancy group such as AgFirst) around the table will enable more effective targeting of outcomes and distillation of outcomes down to the farm level.

Data interoperability standards

2. Create universal data interoperability standards to be used for data sharing within New Zealand by utilising codes of practise used in other countries. Secondly, position an organisation to carry out the data transformation to meet the interoperability requirements between tools. Currently, TANZ is in the best position as an existing organisation.

Anonymisation towards data rights

3. Anonymisation of data will act as a precursor to data rights by creating an early method to pooling data in one place. Use anonymous data to create more diverse industry benchmarks.

Data rights

4. Define strict and variable levels of rights and accessibility to shared data by creating rules which govern the amount of effort required to gain access. Definition of accessibility could be a secondary factor of data interoperability standards. Currently, TANZ is in the best placed existing organisation for working in this space.

Automation between tools and documents

5. Automate data sharing from the tool collecting the information to the tool/document where the data is utilised. Extending the rules for data interoperability plus data rights could be added after these workstreams have been carried out by TANZ.

Data retention by service providers

6. Create regulation for removing retention of data by suppliers after the termination of a relationship. By having this regulation, should give more power to the farmer with whom they do business. Government may be required to aid in the creation of this regulation.

8.0 References

Allen, J., & Wolfert, S. (2011). *Farming for the Future: towards better information-based decision making-making and communication*. Centre of Excellence in Farm Business Management.

Australian Agritech Association. (2021). *Agritech Ecosystem*. Retrieved from https://www.ausagritech.org/ecosystem

Clark, D. (2021). *Beehive Media Release*. Retrieved from https://www.beehive.govt.nz/release/govt-agrees-establish-consumer-data-right

CropX. (2022). *Platform Overview*. Retrieved from CropX Technologies: https://cropx.com/technology/

DairyNZ, & Rezare Systems. (2021). *New Zealand Farm Data Code of Practice*. Retrieved from https://www.farmdatacode.org.nz/

Data Interoperability Standards Consortium. (2021). Retrieved from Data Interoperability Standards Consortium: https://datainteroperability.org/

Defraeye, T., Shrivastava, C., Berry, T., Verboven, P., Onwude, D., Schudel, S., Buhlmann, A., Cronje, P., Rossi, R. (2021). Digital Twins are coming: Will we need them in supply chains of fresh horticultural produce? *Trends in Food Science & Technology*, 245-258.

Environment Canterbury. (2021). *Farm Environment Plans*. Retrieved from https://www.ecan.govt.nz/your-region/farmers-hub/fep/

Farm IQ. (2022). Retrieved from https://farmiq.co.nz/

Farm Table. (2021). *Agtech products, machinery and equipment*. Retrieved from https://farmtable.com.au/accelerate-resources/products-agtech-equipment/

Friedenreich, C. (1994). Methodologic issues for pooling dietary data. *Annual Journal of Clinical Nutrition*, 251-252.

Gallacher, D. (2021). Agricultural data should be a public good - but with exceptions for biosecurity and niche production. *Farm Policy Journal*, 22-29.

Hancock, F. (2021). *Who's eating New Zealand*. Retrieved from RNZ: https://www.rnz.co.nz/news/whoseatingnewzealand/446357/who-s-eating-new-zealand#:~:text=RNZ%20has%20looked%20at%20some,nearly%20%2416%20billion%20in%202020.

Jacquez, J. (1996). Parameter identifiability is required in pooled data methods. *Journal of Pharmacokinetics and Biopharmaceutics*, 301-305.

JanBen, T., Schmidt, M., Dessler, S., Hahn, K., M. Hien, Konate, S., Lykke, A.M, Mahamane, A., Sambou, B., Sinsin, B.; Thiombiano, A., Wittig, R., Zizka, G. (2011). Addressing data property rights concerns and providing incentive for collaborative data pooling: the West African Database approach. *Journal of Vegetation Science*, 614-620.

John Deere. (2022). *Precision Ag Technology*. Retrieved from Technology Products: https://www.deere.com/en/technology-products/precision-ag-technology/?panel=apply

Keenan, L. (2019). *What is the future for farm compliance in New Zealand?* Rural Leaders.

Kpienbaareh, D., Kansanga, M., Luginaah, I. (2019) Examining the potential of open source remote sensing for building effective decision support systems for precision agriculture in resource-poor settings. *GeoJournal*, 1481-1497.

Lincoln University. (2021). *Centre of Excellence in Farm Business Management*. Retrieved from https://researcharchive.lincoln.ac.nz/handle/10182/9557

Lincoln University. (2022). *AgYields National Database*. Retrieved from https://www.agyields.co.nz/home

May, W. (2015). *The value of using digital technology for farm record keeping*. Rural Leaders.

MPI. (2021). *Integrated farm planning work programme*. Retrieved from https://www.mpi.govt.nz/agriculture/farm-management-the-environment-and-land-use/integrated-farm-planning-work-programme/

MPI. (2022). Retrieved from Market insights for the primary sector: https://www.mpi.govt.nz/resources-and-forms/economic-intelligence/market-insights-for-the-primary-sector/

New Zealand Government. (2021). *Towards a Digital Strategy for Aotearoa*. Retrieved from https://www.digital.govt.nz/assets/Digital-government/Strategy/Towards-a-Digital-Strategy-for-Aotearoa-discussion-document.pdf?utm_source=NZTech+News+Mailing+List&utm_campaign=fed7483681-EMAIL_CAMPAIGN_2018_10_01_08_11_COPY_01&utm_medium=email&utm_term=0_

Pfeffer, B. (2021). Harvesting data as an industry asset. *Farm Policy Journal*, 40-47.

Pylianidis, C., Osinga, S., & Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. *Computers and Electronics in Agriculture*, 1-25.

Rezare Systems. (2021). *DataLinker*. Retrieved from https://www.datalinker.org/

Rutherford, C. (2021). Agricultural data sharing for public good - but who pays the price? *Farm Policy Journal*, 16-21.

Stuff. (2021). Retrieved from https://www.stuff.co.nz/business/126484354/realme-faces-questions-as-digital-identity-legislation-nears?cid=app-iPhone

TANZ. (2022). *Trust Alliance New Zealand*. Retrieved from https://trustalliance.co.nz/use-cases/

Whittington, T. (2021). Nowhere to hide - big data in agriculture. *Farm Policy Journal*, 56-60.

Wikipedia. (2021). *Internet of Things*. Retrieved from https://en.wikipedia.org/wiki/Internet_of_things

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. (2017). Big data in smart farming - A review. *Agricultural Systems*, 69-80.

Wolfert, S., Verdouw, C., & Beulens, A. (2008). *Future challenges for information integration in multi-dimensional agri-food supply chain networks*. Wageningen University.