TECHNICAL REPORT

ACCELERATING PRECISION AGRICULTURE TO DECISION AGRICULTURE

Enabling digital agriculture in Australia

This Project is supported by funding from the Rural R&D for Profit Program

Australian Government
Department of Agriculture and Water Resources

DATA TO DECISIONS CRC
UNLOCKING THE VALUE OF AUSTRALIA'S DATA

A big data reference architecture for digital agriculture in Australia.

A Skinner, G Wood, E Leonard and T Stollery

Australian Government
Cotton Research and Development Corporation

FRDC
awi Australian Wool Innovation Limited
AMPC
Australian eggs
Australian Pork
LiveCorp The Australian Livestock Export Corporation
GRDC Grains Research & Development Corporation
Horticulture Innovation Australia
Dairy Australia
AgriFutures Australia
Forest & Wood Products Australia
sra Sugar Research Australia
Wine Australia
mla Meat & Livestock Australia

ISBN: 978-0-6481983-0-7 (Hard Copy) 978-0-6481983-1-4 (Electronic)

The project that delivered this report is led by the Cotton Research and Development Corporation and is jointly funded by the Department of Agriculture and Water Resources Rural Research and Development (R&D) for Profit Program and all 15 Rural Development Corporations.

Data to Decisions CRC
Level 1, Base64
64 North Tce
Kent Town SA 5067
Ph: +61 8302 8978
info@d2dcrc.com.au
www.d2dcrc.com.au

Table of Contents

1. Executive Summary

Agriculture is on the brink of an upgrade. An increasing number of producers are becoming skilled at deploying precision agriculture technologies and the amount of data generated on farm is increasing exponentially. Farm machinery, sensors and digital technologies are now generating large volumes of data about the status of soil, water, crops, animals and pasture. This growth in data, currently stored on and off farm on laptops, in spreadsheets, with consultants, in machinery and in supply chain data centres means that agriculture is well placed to benefit from what is commonly described as Big Data, or more precisely Big Data Analytics.

Big Data analytics promises significant productivity and profit benefits for Australian producers and the emerging “AgTech” industry is pushing producers to do more with the data they generate on farm. Unfortunately, this has resulted in many competing products and conflicting messages to producers leading to confusion and paralysis, and this is exacerbated by the noise in the market place as vendors compete to gain market share of their products and producers worry about making an expensive mistake.

Through Precision to Decision project research, it has been observed that there is a clear digital skills and capability gap within the Research Development Corporations (RDCs) and in the early stages of the industry value chains. RDCs and producer consultants have, to date, failed to develop the key skills in Data Science and Technology that are required for the digitisation of their industries and to take full advantage of Big Data.

Additionally, this skills gap has resulted in a visible vacuum of digital leadership across the Australian agricultural industries. Historically, producers and value chain organisations have leant heavily on their industry bodies and consultants for advice and direction on key industry topics. For digital, the usual sources of advice are currently coming up short.

The result is that producer businesses lack the key digital skills and capabilities required to benefit from Big Data. Worse, based on the findings from eight producer workshops, a cross industry survey of 1000 producers and individual producer interviews it is clear that they also don't know where to start or where to invest to take advantage of the benefits of Big Data and other profit building Digital technologies. The RDCs, industry bodies and value chain participants are not communicating a common "north star" value proposition for Big Data to their producers.

RDCs must quickly address this lack of digital leadership within their industries as machinery, sensor and technology providers – both existing and new to agriculture – are moving the sector towards digitisation regardless of RDC policy. In the absence of cohesive industry digital strategies, RDCs risk being locked out of future market decisions and unable to influence the digital enablement of their industry as the commercial market consolidates and industry leaders emerge.

There is an opportunity for RDCs to accelerate the creation of industry digital strategies by collaborating on the core components of digital enablement. Particularly through co-investing in and sharing knowledge on common platforms, capability, change process and language. Without a cross industry whole-of-Agriculture digital strategy it will be difficult for RDCs to clearly communicate the benefits of digital enablement. This is currently seen in the number of levy payers missing out on co-investment opportunities that span industries, particularly those with mixed production systems.

Competitors in international markets may also leapfrog Australian Agricultural industries as they execute established digital strategies. While not affecting the Australian domestic market, this has particular implications for Australian Agricultural export activities.

Finally, inevitable, ad-hoc digital and Big Data projects both current and future will be delivered into value chains by the RDCs. As these products and services become established it will become harder for the participating RDCs to collaborate on common platforms to optimise value for their industries and levy payers.

The BDRA provides a framework to assist RDC projects with needs in Big Data collection, storage and analysis. To achieve this, the BDRA guides solution architectures by assisting with requirements definitions and identifying appropriate strategies and design patterns for Agricultural Big Data challenges.

It is important to note that a Reference Architecture for Big Data provides just one of the elements that is required to successfully transform Australian Agriculture into a data driven industry. Other elements such as Strategy, Culture, Governance and importantly the change management of each of these have been found to be equally important.

The reference architecture can facilitate collaboration between RDCs by creating a common language and approach when addressing Big Data challenges.

Research undertaken for the project, has surfaced a number of potential initiatives that can be adopted by the agricultural sector in Australia. These initiatives facilitate collaboration between the RDCs to define digital strategy and increase the velocity of adoption of agriculture decision support tools based on Big Data. These are detailed in the following recommendations arising from this research.

1. Recommendation 1

Federal Government, in collaboration with the fifteen Australian Research Development Corporations should make a long-term commitment to data driven agriculture. It is recommended that this be achieved through the establishment of an Office for Agricultural Data Analytics focused on cross industry collaboration opportunities in the areas of data and analytics.

2. Recommendation 2

Federal Government in collaboration with the fifteen Australian Research Development Corporations should employ a Chief Agricultural Data Officer (separate from a Chief Digital Officer) to oversee the activities of the Office for Data Analytics and to provide cross industry technical leadership in the areas of Data and Analytics.

3. Recommendation 3

The fifteen Research Development Corporations must collaborate to develop a cross industry digital strategy and implementation roadmap for cross industry digital enablement and to support the expansion of shared services. Additionally, policy should be put in place enabling the development of Shared Services (e.g. a common Agricultural Platform) to encourage the sharing of knowledge between industries as the default course.

4. Recommendation 4

The fifteen Research Development Corporations must collaborate to create a Change Management Strategy to facilitate the adoption of data, digital technologies and shared services within the Australian Agricultural Industries. The Research Development Corporations must work to win the backing of Producers and Industry Groups by implementing cross industry extension programs, helping levy payers understand the value of data and the small steps they can take to introduce benefits into their businesses.

5. Recommendation 5

It is recommended that the fifteen Research Development Corporations collaborate to establish a set of Foundational Data Sets for use cross industry. Looking internationally it is clear that cross industry datasets such as soils, climate and market data have significant value for developing Agricultural Decision Support services.

6. Recommendation 6

It is recommended that the fifteen Research Development Corporations build Data Science competency throughout their value chains. Competency Frameworks should be established and facilitated by the shared Office for Data Analytics. Industry representatives from the 15 Research Development Corporations can take sabbaticals into the Office for Data Analytics to build industry competency.

7. Recommendation 7

It is recommended that regulation be used to ensure the collection of core industry data. There are a number of data points that, if collected, could significantly advance Australian Agricultural Industries. Presently, collection of this data is discretionary and as such Producers frequently do not participate as they do not see immediate value. Additionally, for those who do, the quality of the collected data is often poor. Light touch regulation would provide the required incentive for producers to collect and submit this data.

8. Recommendation 8

It is recommended that the fifteen Research Development Corporations collaborate to establish an open standards hub and associated principles (see, for example https://standards.data.gov.uk) where standards can be suggested, considered, implemented and catalogued in an open forum. Note, that while standards are recommended, it is crucial that the development of standards should not hinder the Research Development Corporations priority of making industry data available.

9. Recommendation 9

It is recommended that the Research Development Corporations collaborate to develop the mechanisms needed to source and gain access to corporate data sets by creating appropriate legal frameworks, commercial partnerships and building relationships.

10. Recommendation 10

It is recommended that the fifteen Research Development Corporations collaborate to establish common cross industry data security definitions and standards.

11. Recommendation 11

It is recommended that the fifteen Research Development Corporations collaborate to enable public Open Data with clear usage guidelines. The data made available should include the data sets developed by the Office for Data Analytics where appropriate. Research Contracts should be updated to include commercialisation clauses to give industry value chains access to appropriate research and data.

12. Recommendation 12

It is recommended that the Rural Development Corporations increase investment in automating data collection to improve data quality and consistency.

13. Recommendation 13

Revisit the functionality and user experience of existing levy funded foundational models (e.g. Grazing, Weather, etc.) to build capability and drive adoption. Create a database of available models with clear usage guidelines.

2. The scope of this document and how to use it

The Big Data Reference Architecture (BDRA) provides a framework to assist Research Development Corporation (RDC) projects with needs in Big Data collection, storage and analysis.

The BDRA guides solution architecture by assisting with requirements definition and identifying appropriate strategies and design patterns for Big Data challenges. The intended audience is RDC managers, architects, business analysts, solution designers and other roles involved in scoping, requirements definition or solution design of Agricultural Big Data projects.

The BDRA takes some direction from the NIST Big Data Interoperability framework.

The reference architecture for Big Data is one element within a wider digitisation strategy that will enable data driven decision making within Australian agriculture. Other elements such as culture, governance and training are equally important as illustrated below in Figure 1 Big Data Pillars for Success which is adopted from the Defence Industries “Pillars for Big Data Success”.

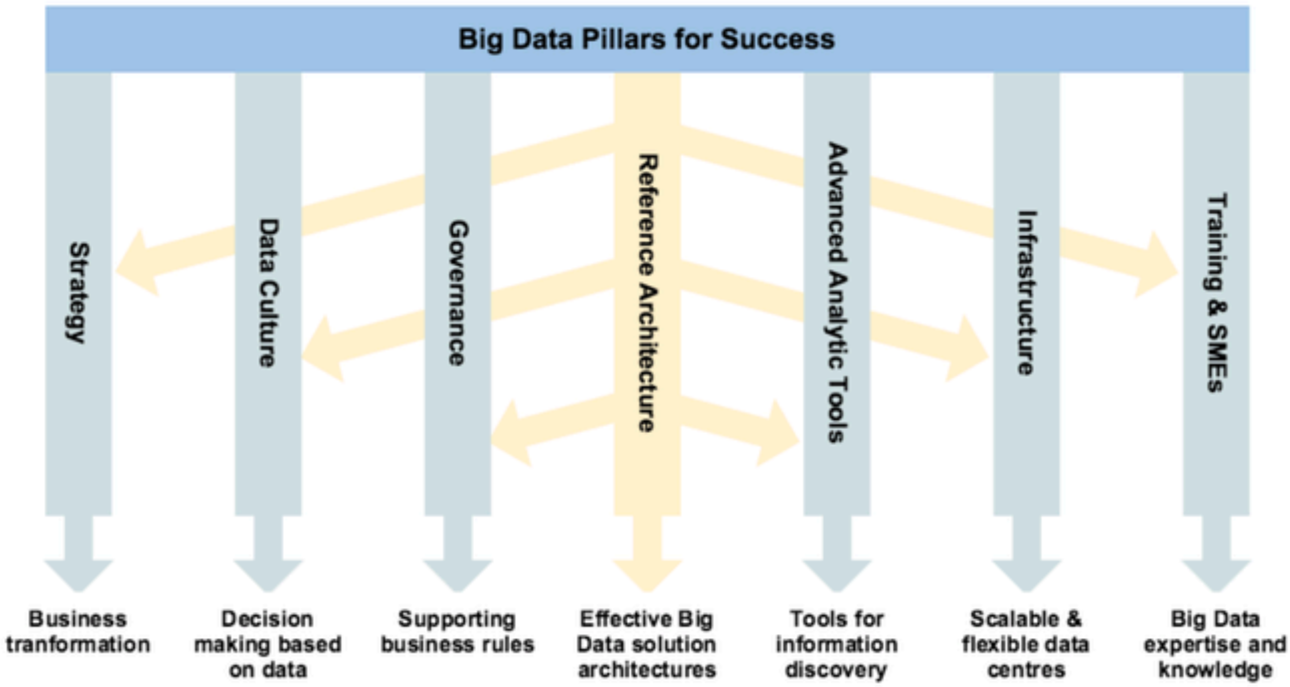

Figure 1. Big Data Pillars of Success

2.1 Scope

The Reference Architecture provides guidance and recommendations for Agricultural projects that have a Big Data component, including addressing many Big Data considerations that are unique to the Australian agricultural industries. The BDRA doesn't provide general guidance for traditional IT solution design. Where considerations overlap with traditional IT considerations, only those that are pertinent to Big Data are covered.

The BDRA is a descriptive Reference Architecture. It is designed to aid communication by supporting consistent use of terminology across systems and providing guidance and recommendations for solution design. The BDRA is not prescriptive i.e. it does not define architecture standards, data standards or constrain solutions.

Big Data challenges are defined as technologies and initiatives that involve data sets that are too diverse, fast-changing or massive for conventional technologies, skills and infra- structure to address efficiently. In other words, the volume, velocity or variety of the data is too great.

The main characteristics that define Big Data and force new architectural approaches include:

- Volume:Size of the dataset - farmland accounts for 61% of Australia's land mass,
- Velocity:Rate that data arrives into the system - farm sensors and machinery are recording data every second,
- Variety:Data from multiple sources or types - multiple data providers, manufacturers, agtech companies all competing with data types and formats, spacial data, satellite imagery, soil samples, climate data etc.

These are known as the 3Vs of Big Data and were originally defined by Gartner Analyst Doug Laney in 2001. Some, more recent definitions of Big Data extend the 3Vs to also include additional Vs such as Variability - how the data changes over time or Veracity - data from sources with varying trustworthiness or correctness. However, these additional Vs generally support or clarify the original three.

2.2 How to use the Reference Architecture

The reference architecture component of this report is divided into two main sections, components and considerations.

1. The components define and describe the high level components of a Big Data solution that makes up the BDRA. Each component is a logical grouping of considerations.

2. The considerations describe the considerations for each component in detail, including example strategies and design patterns for supporting the considerations.

The components present in the BDRA represent logical groupings of considerations and don't usually directly map to individual components in a concrete system architecture. The scale of Big Data solutions often leads to separating the solution into implementation components that address considerations of several components in one product. Furthermore, some considerations affect multiple components, these are described in the BDRA as global considerations (security, federation etc).

Readers of the BDRA should become familiar with all of the components, identifiy those that are relevant to their use case and then consult the relevant considerations to identify the suggested potential strategies for how to address them. It is important to note that the strategies covered in the considerations are not exhaustive, instead they provide common design patterns and techniques that can be applied as a baseline.

Specific products are sometimes described within the BDRA where they aid the description of the technique or strategy. They are intended to be used only as examples and not recommendations. Solution architects should evaluate a wide range of products when designing solutions and should not be constrained to the examples in this BDRA. Where products are described, Open Source solutions have been prioritised for discussion - this is not an endorsement for open source.

To aid the reader and to provide consistency, the components and considerations are supported by a documented set of guiding principles, a review of the current state of Big Data in Australian Agriculture and an overview of the Big Data lifecycle. These sections of the BDRA can be used to inform wider Big Data project design decisions.

Also included in the BDRA is a set of example architecture. These provide examples for using the BDRA to create a concrete system design. They show how to use the BDRA to step through the various considerations and to select appropriate strategies and design patterns. These are based on the included use cases.

A decision support tool is provided to aid the reader in identifing and evaluating their current Big Data requirements and to support the evaluation of appropriate products and services.

Finally, a set of recommendations have been made. Data to Decisions CRC consider these to be opportunities for the Research Development Corporations to collaborate to effectivly leverage Big Data. They are based on our research findings and observations thoughout the project.

3. The Big Data Lifecycle

Big Data analytics differs from traditional data analytics mainly because of the challenges introduced by the volume, velocity and variety of the data that is being processed. To address the distinct requirements for applying analytics to Big Data, a step-by-step methodology is often useful to organise the activities and tasks involved with collecting, processing, analysing and making decisions from data.

This section of the BDRA aims to aid the process of approaching and planning a Big Data project within an Australian Agricultural organisation such as a Research Development Corporation. This is achieved by describing a structured lifecycle for approaching Big Data activities and tasks.

There are several established Data Mining standards already published such as the CRISP-DMmethodology and SEMMA that describe in detail common Business Intelligence challenges and offer techniques to manage them. However, in the context of Big Data these approaches fall short. For example, the SEMMA methodology does not consider data collection and pre-processing of different data sources. As many organisations working with Big Data have found, these are often the most time-consuming components that consume most of the resources available to Big Data projects. CRISP-DM also misses some important considerations needed for managing Big Data.

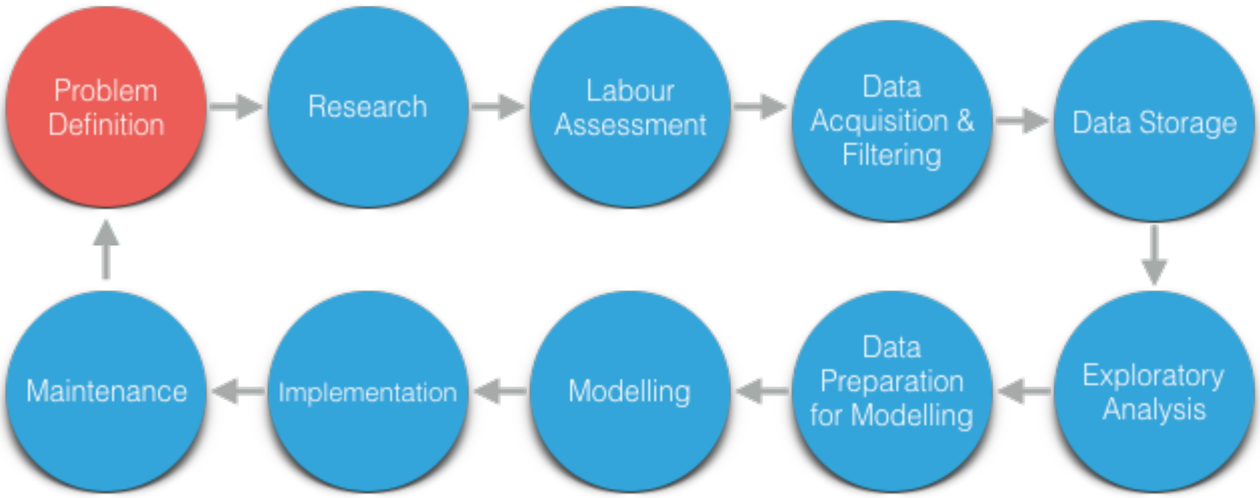

Figure 1. Big Data Application Lifecycle

Building on these methodologies a circular agricultural Big Data Lifecycle (illustrated above) can be organised around the following ten stages:

1. Problem Definition
2. Research
3. Labour Assessment
4. Data Acquisition and Filtering
5. Data Storage
6. Exploratory Analysis
7. Data Preparation for Modelling
8. Modelling
9. Implementation
10. Maintenance

The sections below describe each of these lifecycle stages, the activities that should occur during each and the outputs that should be expected at each stage.

3.1 Problem Definition

While an obvious place to start, problem definition can also be the most critical step in a Big Data lifecycle. This is especially true for the nuanced, complex value chains that exist in Australian Agriculture. Is the right decision being targeted? Is the right question to enhance the decision support process being asked? Are the expected outcomes, costs and risks adequately defined and accepted by all stakeholders?

3.2 Research

Once the problem has been defined, research is required to understand what solutions other agricultural and importantly non-agricultural industries may have already developed. This will also involve looking beyond Australia to international project outcomes and defining a strategy for managing the future stages of the Big Data Lifecycle.

3.3 Labour Assessment

With the problem defined and an approach to managing the Big Data Lifecycle agreed with stakeholders. The current skills available in industry need to be assessed to ensure that there is capability to address the Big Data activity. This will involve reviewing capabilities of staff and also evaluating external SMEs against a capability matrix (see RFP tool). Additionally, the transformative nature of many Big Data projects will also require a change management strategy to be developed to aid adoption and to reduce risks arising from process change etc.

3.4 Data Acquisition and Filtering

Data acquisition is a key component in the Big Data lifecycle. This is a non-trivial step and normally involves ingesting structured and unstructured data from multiple sources often arriving in real-time from sensors or other streaming sources. This data is then subjected to filtering for the removal of corrupt data or data that has been deemed to have no value to the analysis objectives. For example, this could include ingesting weather data, satellite or drone imagery, batches of data uploaded ad hoc by a producer/property who has intermittent internet connectivity, or retrieving data from supply chain organisations.

3.5 Data Storage

Once the data has been acquired and filtered, the data often needs to be retained or stored. There are many Big Data technologies designed to support this and strategies for working with these technologies are covered in detail elsewhere in the Big Data Reference Architecture. This stage in the Big Data Lifecycle aligns with the Labour Assessment as many of these technologies require specialist skills to implement, manage and support. Including cloud offerings from technology providers.

There may also be situations where data may not need to be retained because the specific analytic output is the valuable asset and there is certainty that the input data will not be reused.

3.6 Exploratory Analysis

Once the data has been cleaned, enriched and stored in a format ready for analysis, the data exploration stage is mandatory. The objective of this stage is to make sure that the data is explored and understood ready for the application of Data Science. During this stage, it is necessary to re-evaluate the problem definition to ensure that with the available data the problem still holds true, or if any additional opportunities have been presented.

3.7 Data Preparation for Modelling

The Data Preparation stage involves cleaning and enriching the data and moving it into a format optimised for the analytic. It also involves setting up tests to manage data quality and isolating features for selection and extraction. Invalid data can skew and falsify analysis results. Unlike traditional enterprise data, where the data structure is pre-defined and data is pre-validated, data put into Big Data systems can be unstructured without any indication of validity. Cleaning data is the process of establishing often complex validation rules to remove any known invalid data ahead of modelling.

3.8 Modelling

The data preparation stage will have produced several datasets for iterating the training and testing of a data model or models. For example, a Data Scientist may train a machine learning model to predict the susceptibility of a crop variety to a disease. Ultimately the model output will be verified or ground 21rothed against real world observations to evaluate the quality of the analytic.

3.9 Implementation

The analytic is then implemented into the analytic pipeline of the value chain organisation or integrated into the producer managed decision support tool. The output of the model will need to be tracked over time to ensure the quality of the analytic over time and in order to track it's performance as new data is ingested.

3.10 Maintenance

Perhaps the most important stage of the lifecycle and the stage that is often omitted is the maintenance stage. The maintenance stage is important as activities include monitoring the analytic for accuracy, capitalising on new analytic approaches, the ongoing evaluation of the output against the producer management action and the capture and application of new data sources. These activities often trigger a rerun of the overall Big Data lifecycle to realign the solution to the business problem.

3.11 Summary

There are several published data management methodologies that can be augmented to provide a ten-step repeatable, managed lifecycle for Big Data capable of enabling decision support analytics for use across agricultural industries. The other sections of the BDRA expand and describe the detail behind each step where required.

4. Big Data maturity in the Australian Agricultural industries

This section of the Precision to Decision Project Big Data Reference Architecture (BDRA) evaluates the maturity of Big Data use in Australian Agriculture. The fifteen Rural Development Corporations have been evaluated against the model described in the following sections to establish an understanding of current state.

The maturity model presented below was developed by Data to Decisions CRC and is based on our Big Data Pillars of success as well as observed Big Data maturity in ajacent sectors including mining, utilities and manufacturing.

The resulting model aims to evaluate six of the seven characteristics of the Pillars. It should be noted that based on feedback from the project workshops, industry interviews and the wider producer survey, Data and Culture have been amended and evaluated separately in the model whereas they are grouped in the Pillars. Also, due to the prevalence of established cloud services - infrastructure maturity has not been included in the evaluation as it was deemed unnecessary for the purposes of this report. As a result, the areas evaluated were: Strategy, Data, Analytics, Culture, Architecture, Governance and Training.

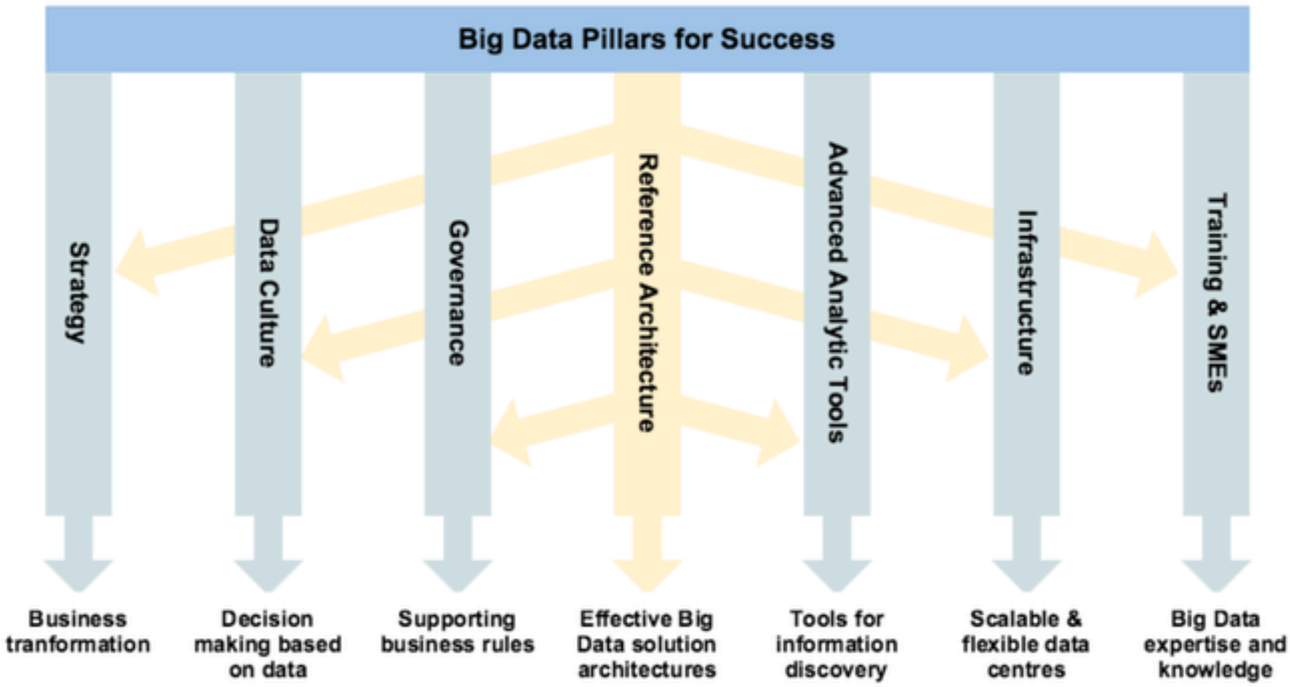

Figure 1. Big Data Pillars of Success

4.1 The Precision to Decision Big Data Maturity Model

Each category of the maturity model is described in detail below. The evaluation outcomes were reached from observations taken during the eight regional producer workshops along with substantive desktop research, producer interviews, interviews with nominated RDC representatives and supportive commercial providers. The model considers not only the technical maturity required for Big Data success but importantly, where appropriate also considers farm business practices.

Inspiration for the model has been taken from similar maturity models created by IBM in their evaluation of other enterprise capabilities such as mobility, governance and service orientated architectures.

The maturity model comprises of seven categories for which five levels of maturity are described.

4.2 Maturity categories

4.2.1 Strategy

The first consideration with any advanced technology capability in Agriculture such as Big Data is to recognise that its use needs to be driven by or support demonstrable producer outcomes. In other words, all activities need to track through to increased profits. While underpinning technology is needed to acquire farm/industry data and to execute analytics, deep industry/production expertise is necessary to derive meaningful insights and then to use them to develop valuable whole of farm business outcomes throughout the value chain. Whole of farm business outcomes can be achieved by enriching value chains and driving operational improvements. This requires producers and their industries to transform their businesses to become more data aware, exploring data for new decision making opportunities and leveraging these decisions to maximise yield and/or profit. Producers and value chains that are mature in the model have developed and communicated strategies that enable them to use available data and apply analytics to innovate, improve their decision-making processes, maximise their value chain and open new market opportunities.

4.2.2 Data

Use of data to inform on-farm decision making is the base capability. However, agricultural industries mature in Big Data Analytics understand that data is a valuable asset. This data comes from many sources such as business records, machinery, data collected on farm, contributed third party data, from sources throughout the value chain as well as other external and open data sources. Industries mature in their approach to Big Data Analytics support producers to provide governed access to their data, sharing it with third parties where valuable to give it meaning and context.

4.2.3 Analytics

Mature use of analytics optimises on farm decision making. The industry survey conducted as part of the Precision to Decision project showed that producers are already using data to manage their finances and to report their regulatory compliance but analytics can also give producers an understanding of why something has happened or to predict what is likely to happen in the future. The resulting insights help inform on farm decision making, maximising yields and/or profit. Agricultural industries mature in their approach to Analytics make data-driven decision making pervasive throughout their value chains, and this requires timely insight in context.

4.2.4 Culture

Data aware industries that promote analytics to make decisions and to derive insights is of little value without the engagement of producers and stakeholders along the value chain. To maximise the benefits, producers need to embrace data as a core tool of the trade, constantly seeking out and utilising decisions derived from their available data. This requires cultural change. To enable this cultural change, trust in these data driven decisions is essential so that the use of data becomes reflexive. To drive this cultural change it is also important that producers have the ability to easily visualise the results and to provide feedback so that their decision support tools become “go to” business tools. Industries mature in their approach to Big Data culture promote data or information first approach to decision making and offer producers diverse, targeted data analytics services aligned to their production processes in order to embed data into modern industry production practises.

4.2.5 Technology

A planned, evolutionary, strategic technology approach to big data and analytics is essential for agricultural industries to establish a thoughtful, stable, scalable capability. This enables a managed approach to data access by end users including producers, researchers, commercial organisations, government and industry bodies. It provides agility to address changes in production processes and across the value chain as well as enabling a considered approach to interoperability. Mature industries will establish architecture that supports the primary three V's of Big Data: Volume, Variety and Velocity. They will support the three V's through the creation of and reuse of shared cross industry architectural patterns, data and standards. This includes supporting cross cutting challenges such as Security, Governance and Service Levels.

4.2.6 Governance

With producers citing a lack of confidence in Agricultural Big Data systems as a significant barrier to adoption, Governance is a critical success factor for Agricultural Big Data projects. Mature Industries will have in place policies that cover ownership, provenance, currency, data quality, foundational data and metadata, lifecycle management, security, privacy and ethical use.

4.2.7 Training & SMEs

Big Data solutions are worth little unless the relevant skills are available in industry to maximise the benefits of investment. A structured approach to building Big Data and Data Science expertise both through training staff and augmenting capability by identifying, evaluating and establishing trusted SMEs is essential for Big Data success. A mature industry will recognise Big Data and Data Science as core competencies that differentiate in the marketplace and build business value and invest in their people and partnerships to maximise the opportunities.

4.2.8 Maturity Levels

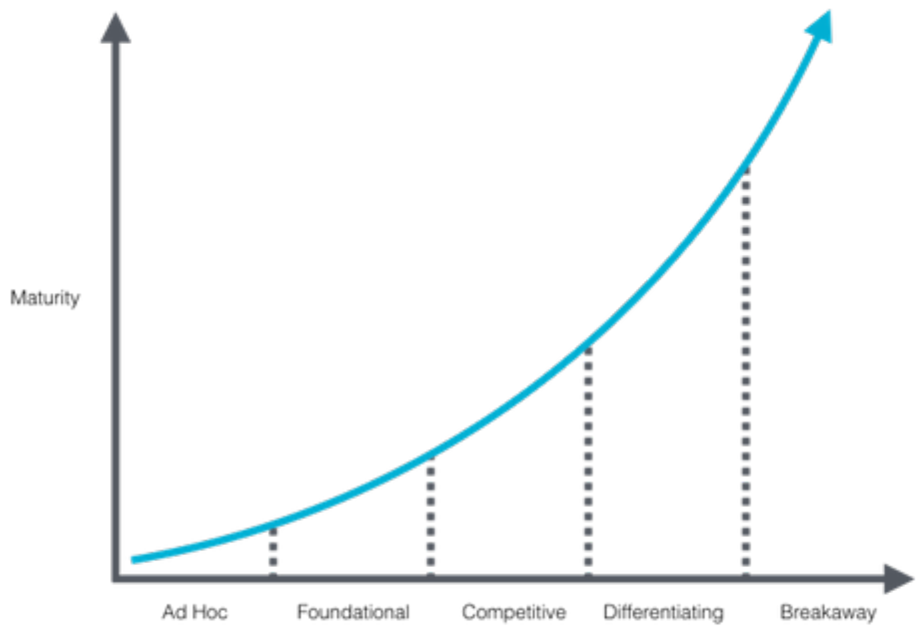

Figure 2. Maturity Levels

Five levels of maturity are defined:

- Ad hoc
- Foundational
- Competitive
- Differentiating
- Breakaway

These maturity levels are intended to illustrate maturity from basic (Ad hoc) through to advanced (Breakaway). The meaning of each is dependant on the category with which it is associated. The following matrix descibes the level of maturity expected for each level for each category.

4.2.9 Evaluation Matrix

	Ad hoc	Foundational	Competitive	Differentiating	Breakaway
Strategy	Big Data is discussed but not reflected in industry or farm business strategy. Use of data is limited to financial management and regulatory reporting.	Producers and Industry bodies recognise that data can be used to generate profit. However, most data at scale is held by larger supply chain entities.	Industry strategy supports producer access to data throughout the value chain to support whole of farm decision making processes.	Industry strategy realises competitive advantage for producers using data to make decisions.	Data drives continuous cross industry and production innovation.
Data	Producers and the industry value chain use historical structured and unstructured data to make decisions.	Data is used to effectively inform producer decisions.	Data is applied across the value chain to improve industry and producer operational processes and decision making.	Data and predictive data driven decision making is used as an industry differentiator.	Data is used as a strategic asset driving producer profits and increasing market access.
Analytics	Analytics is limited to describing what has happened.	Analytics is used to inform decision makers why something on farm has happened.	Analytical insight is used to predict the likelihood of what might happen.	Predictive analytics is used to help optimise producer decision making with linked farm management actions taken to maximize yield and/or	Predictive analytical insight optimizes production processes and management actions are automated where possible.

				profit.	
Culture	The application of analytical insight is the choice of the producer and has little effect on profit or the wider industry.	The producer is aware of the insights available from data produced on farm but is largely resistant to adaptation required to take advantage of the insight.	The producer makes limited farm management decisions using analytical insight to improve operational efficiency and generate additional profits.	Producers are well informed with insight from analytics, and the capable of acting to maximise resulting yield/profits.	The producer and value chain continuously adapt and improve, using analytical insight to support their strategic objectives.
Architecture	Producers and the industry value chain does not have a single, coherent architectural approach to big data within their farm business.	A common Big Data framework exists but does not extend to new data sources or advanced analytics capabilities.	Best-practice architectural patterns for Big Data and Analytics are defined and have been applied in certain areas.	Big Data Architecture and associated standards are well defined and widely adopted covering most of the volume, variety and velocity challenges for structured and unstructured data.	Established Big Data architecture fully underpins business strategies to enable complete market disruption with volume, variety velocity and veracity specifications applied.
Governance	Data governance is largely manual and barely sufficient to stand up to legal, audit	Understanding of data and its ownership are loosely defined and managed in a piecemeal fashion.	Policies and procedures are implemented to manage and protect core data through its	The degree of confidence in data and resulting insights is reflected in making decisions.	Information governance is integrated into all aspects of the business processes.

	and other regulatory scrutiny.		life in the organization.		
Training & SMEs	There is little or no Big Data expertise within the industry	Skills gaps have been identified but no formal strategy to expand expertise to fill gaps has been developed	Training frameworks are implemented and have been applied to key areas of the value chain	Training frameworks are mature and Big Data Analytics expertise is established throughout the value chain	Big Data Analytics is recognised a core competency within the industry generating measured returns

4.2.10 Evaluating the current state of Big Data maturity

The model has been applied across the fifteen participating Agricultural Industries informed by the observations and feedback that data to Decisions CRC obtained during the eight Precision to Decision Regional Big Data workshops. Additional insights have also been taken from the CSIRO run Producer Survey results in order to establish a current state reference for the maturity of Big Data use across the participating industries. While different industries presented differing levels of maturity in some areas. For example, strategy and data. The observations measured against the model been averaged to present a whole of Australian agriculture view of Big Data maturity. Wide variances have however, been noted. This current state is intended to be the starting point upon which the principles, challenges, opportunities and recommendations identified elsewhere in the BDRA are based.

With that in mind the following section describes the observed maturity of Big Data within Australian agriculture.

4.2.11 Strategy

Observed as Ad hoc.

"Big data is discussed but not reflected in industry or farm business strategy. Use of data is limited to financial management and regulatory reporting."

Of the fifteen RDCs, when asked, none could produce a published Digital Strategy that described the roadmap for the digitisation of their industry and where Big Data analytics mapped into the roadmap. Several RDCs have strategies in development and have evaluated data as a source of value to their industry but were not able to assign a dollar value against their industry value chain adoption to support this. Of the producers attending the workshops several of the larger organisations had developed commercial strategies but were unwilling to share the details for consideration. For smaller producers, with the exception of a higher than anticipated adoption of on farm weather stations for climate based decision making most were keeping data for financial management and regulatory reporting purposes only. Smaller farm businesses had no planned digitisation strategy and not considered Big Data or analytic use in their decision-making processes instead they relied heavily on their consultants. Of the consultants who attended the workshops along with those we interviewed as part of our research none had developed a Big Data or analytic strategy or differentiated themselves with Big Data analytics within their industry.

4.2.12 Data

Observed as Ad hoc

"Producers and the industry value chain use historical structured and unstructured data to make decisions."

Through the Precision to Decision project research, it was discovered that volumes of data are collected across all fifteen participating industries by producers. This finding is supported by the producer survey (see figure 3 - Survey results - Data collection by industry).

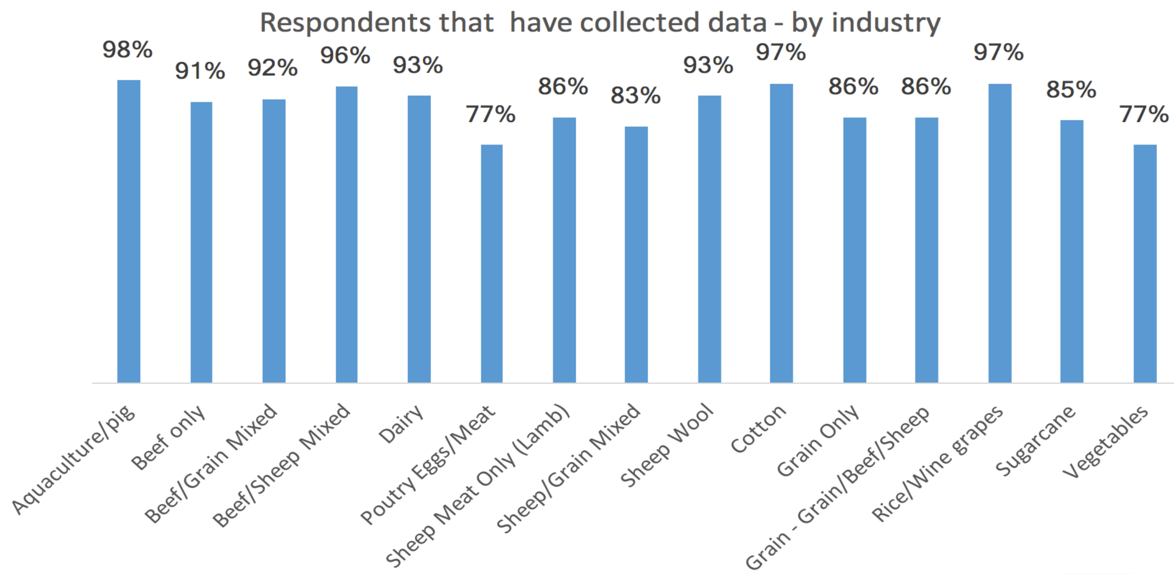

Figure 3. Survey results - Data collection by industry

Data collected on farm and by value chain organisations was found to be diverse and included but was not limited to structured data representing climate, soil moisture, irrigation, yield and farm inputs.

Data is also generated by on farm machinery, labour and a growing volume of other on-farm sensors. In addition to structure data from machine telematics and sensor readings, it also includes data from unstructured sources such as NDVI, Geo Physical, Drone and Satellite other kinds of imagery that is also in use although not necessarily at scale. Additionally, there are many non-digital sources of data such as printed haulage dockets, receipts and log book entries.

Data quality has been observed as an issue as often data is ad hoc or collected from poorly calibrated sensors, yield monitors and other sources. Governance to remedy this is absent or not well established.

It has been observed that currently, despite the growing availability of online farm management solutions, industry consultants and smaller producers generally use basic apps to work with data along with tools such as Microsoft Excel limiting their intuition to evaluate their data to make on farm decisions for their customers. Data representation and interpretation was also a commonly observed issue. Producers commonly reported that they did not know how to evaluate the data that they collected or received (such as soil sample results) to make effective on-farm management decisions. Additionally, the data in use is usually limited to a single property making benchmarking impossible. Finally, many producers cited a lack of interoperability of data between tools and services, again indicating an overall lack of maturity across all industries.

For all industries, it was observed that many value chain organisations working with multiple producers can collect, aggregate and benchmark data enabling them to gain broader regional and national insights. In most cases, this data is not well governed or made easily available to back through the value chain to producers for them to use to inform effective management decisions.

4.2.12 Analytics

Observed as foundational to competitive

"Analytics is used to inform decision makers why something on farm has happened."

The maturity of analytics across the participating industries has been evaluated as foundational to competitive. At present, it has been observed that much of the analysis of data undertaken across industries is used to understand what has happened in the past so that management decisions can be taken to optimise future outcomes.

"Analytical insight is used to predict the likelihood of what might happen."

However, many academic models have been developed through years of industry funded research. In many cases, these analytics also aim to predict future decisions, for example: given these inputs, when will an animal management group reach a weight for optimal eating quality so that they can be sent to the processor? Academics are also now applying Machine Learning techniques in some areas. Other examples of models available to support future decisions include pasture availability, yield and carcass weight. Again not all industries have the same level of skill available to them in this area.

However, while these models exist, the Precision to Decision project research has observed that their uptake within industry is not widespread and adoption by producers is ad hoc.

Reasons for this include:

- a lack of automation – high degree of effort is often required to use them,
- complex user interfaces provide cryptic feedback and no clear management actions and;
- the data producers can make available to the models is limited and of variable quality often limiting their effectiveness.

This has led to lack of take up and distrust of the outcomes, as such producers predominantly return to evaluating past seasons when making their management decisions. This lack of adoption is why maturity has been evaluated as foundational to competitive as first generation models exist but are not yet widely adopted.

4.2.13 Culture

Observed as Ad hoc

"The application of analytical insight is the choice of the producer and has little effect on profit or the wider industry."

The maturity of the participating industries has been placed at ad hoc to foundational as the project found that Big Data use is limited to a small number of motivated producers and while some had seen increases in inefficiencies through the implementation of data driven decision making leading to increased profits overall the ad hoc application of Big Data analytics had, had very little effect on their business. Additionally, due to interoperability challenges though out the value chains the use of data to make management decisions has had very little effect on the wider industry.

Outside of the small motivated groups in each industry, producers are not widely aware of the insights that are available from the data that they produce in their businesses and due to the perceived costs of adoption, many are resistant to take advantage of the available insights.

This is also hindered by a lack of domain expertise within the industry bodies whom are unable to clearly articulate positive, verified outcomes resulting from the adoption of Big Data Analytics technologies.

4.2.14 Technology

Observed as Ad hoc

"Producers and the industry value chain does not have a single, coherent architectural approach to big data within their farm business."

The findings of the eight regional producer workshops along with the follow up interviews with producers, the RDCs and other value chain organisations has shown that across the participating industries there is no common approach to Big Data. The Precision to Decision project, specifically the BDRA will make recommendations based on cross industry research to address this. However, current state of Technical maturity for Big Data technology within Australian Agriculture has been evaluated as Ad hoc. The recommendation made by the project, along with the establishment of well-articulated industry digital strategies and change plans will quickly increase maturity in this area.

4.2.15 Governance

Observed as Foundational

"Understanding of data and its ownership are loosely defined and managed in a piecemeal fashion."

Each participating industry is aware that the data it creates has value. There is also a strong awareness regarding privacy and IP. All the participating RDCs could provide data privacy policies. Many larger producer organisations (such as processors, supermarkets etc) also had established governance policies regarding data use in their organisations.

However, many organisations in the value chain had loosely defined governance in place for data. When challenged most were unable to clearly articulate private vs public data in their businesses and many were not aware of the licencing agreements that they had entered into with their consultants and machinery manufacturers. Other governance activities such as provenance, currency, data quality, foundational data and metadata, lifecycle management, security, and ethical use had not been considered.

As a result, the maturity of the participating industries for Big Data Governance has been evaluated as Ad hoc to Foundational. As the larger organisations were demonstrably more mature in these areas than most of the industry producers.

4.2.16 Training & SMEs

Observed as Ad Hoc

There is little or no Big Data expertise within the industry

When interviewing each participating industry, a common comment was that of a lack of or absence of Data Science and Big Data capability. Of the industries interviewed most had not identified or considered how to address the gap while the two that had formally identified the need had not considered how they may address training. This was also found to be true with producers. Those attending the regional workshops cited a lack of expertise in data, Big Data and Analytics as a significant gap in their businesses. Additionally, they cited a lack of capability in their value chain and that their consultants also often did not have the degree of capability required to help them make decisions from their data that led to demonstrable outcomes in their businesses. The results of the producer survey supported this view. <INSERT>. As a result, Data to Decisions have evaluated the industry maturity in Training and SMEs as Ad hoc when reviewed against the maturity model.

4.2.17 Summary

It is clear that when evaluated against the proposed maturity model, the participating industries evaluate to a low (ad hoc) maturity across all six categories:

- All industries currently lack a clear strategic roadmap for developing Big Data capability both at industry and producer scale.
- Data is being collected across industries however, foundational data sets have yet to be defined, interoperability is an issue and good data management skills are still to be developed.
- The business outcomes of data required to build a data culture, are still to be validated and the time required to do this may in some industries take many years due to crop rotation etc.

- Technology standards are yet to be developed and cross industry collaboration to maximise spend is yet to occur. There are some industries who are developing or have developed data analytics capability but not at petabyte scale. Other industries are building substantial data sets but are unsure of how to query the data for decision support insights.

- Governance is in place at a macro level across industries in terms of managing data privacy however deeper governance of data such as provenance, currency, data quality, foundational data and metadata has not been established in any of the participating industries.

- A structured approach to building Big Data and Data Science expertise is yet to be developed by the participating industries. Capability has been cited as being the leading market failure across industry value chains.

Through the creation of an actionable cross industry digitisation/Big Data roadmap many of the above challenges can be addressed quickly moving the participating industries towards maturity. However, to become breakaway across all categories (if needed) will require focused cross industry effort and collaboration.

5. Attachment

A Big Data Reference Architecture for Digital Agriculture in Australia

- DRAFT - Big Data Technology Report - https://p2d.d2dcrc.net/
- info@d2dcrc.com.au
- d2dcrc

This DRAFT Big Data Technology Report has been developed as part of the Rural R&D for Profit Precision to Decision Project. Compiled in collaboration with the 15 Australian Research Development Corporations it is intended to provide best practice guidance for the development of Agricultural Big Data Systems.

Not sure of the right questions to ask about Big Data? Use the Big Data Reference Architecture Decision Support tool to support your RFP or Proposal Review.

Support Tool

Contents

ACCELERATING PRECISION AGRICULTURE TO DECISION AGRICULTURE - P2D PROJECT RESEARCH PARTNERS

CSIRO

une University of New England

Griffith University

USC Australia

DATA 61 CSIRO

DATA TO DECISIONS CRC

Australian Farm Institute

PROJECT LED BY

Australian Government

Cotton Research and Development Corporation