Australian Farm Institute

New thinking about agriculture

Reinventing Australian agricultural statistics

Authors:
Katie McRobert; Lucy Darragh; Anne Laurie; Richard Heath
May 2019

The Australian Farm Institute acknowledges the financial assistance of Australian Wool Innovation, Cotton Research & Development Corporation, GrainGrowers and Meat & Livestock Australia to undertake this research.

The funds provided by Australian woolgrowers, cotton growers, grain growers and livestock producers and the Australian government to support the research, development and innovation detailed in this publication are in turn gratefully acknowledged by Australian Wool Innovation Limited, the Cotton Research and Development Corporation, Grain Growers and Meat & Livestock Australia respectively.

Contents

Executive summary

The Australian agricultural statistics system is in a process of evolution, but as yet is still in a far from ideal state. Data for the sector is often incomplete, out of date, and irrelevant or purposeless. There is an urgent need to improve the collection, analysis and timely provision of agricultural statistics by taking innovative, cooperative action to improve the way this information is gathered and distributed. Failure to do so will compound the existing problem of decisions and policy made in the absence of solid evidence.

Data collection is no longer the exclusive domain of Official Statistics Agencies (OSAs) – and indeed these agencies are unlikely to receive the increased resources required to enable them to address these complex data needs as well as meeting their current obligations.

Given the funding limitations, it is fair to say the current agricultural statistics portfolio reflects available OSA resources, rather than the actual needs of the sector. In a time when water management and the impacts of climate change on agriculture are issues of fundamental national importance, this situation is unacceptable. Sound decision-making requires informed understanding, which in turn requires a reinvention of the current Australian agricultural statistics system.

Efforts to improve official statistics in Australia are already underway following recent reviews. The ABS and ABARES have established programs to modernise and streamline operations. An increase in resources would ensure that these work programs maintain momentum, but the OSAs alone cannot reinvent the system.

In order for the sector to leave behind sole reliance on the five-yearly census process and move into a more responsive, accurate and granular system, the agriculture industry has a pivotal role to play in identifying needs and making available existing datasets which could serve those needs. The industry must collectively embrace this role to identify industry problems and opportunities, promote agriculture's social licence, educate policy-makers on trends and requirements and to also build trust in the distribution and responsible use of data.

This report presents a framework for evaluation of data sources which could augment the existing agricultural statistics system, developed from a desktop study of available literature, and investigation of potential alternative data sources and methods of collection.

Recommendations

The authors recommend that:

1. Industry and OSAs cooperatively agree on a **common framework** (such as the example presented herein) for the evaluation and implementation of alternative or additional data sources to augment the Australian agricultural statistics system.

2. Industry recognises and supports ABS in its role as the official statistical *collector* as well as being the ***coordinator*** **in a multi-stakeholder data-gathering process**.

3. An **education program** be established by ABS for relevant personnel in industry bodies which currently collect data to ensure understanding of the considerations for inclusion of non-official data in the national agricultural statistics system.

4. An education program be established by industry for all stakeholders in the agricultural value chain regarding **the industry good benefits of sharing data**.

5. An Agricultural Data Taskforce (as recommended in the P2D report) recognises the **cross-industry need to share learnings** on systems transformation to ensure a harmonised approach to collection of national data.
    a. *Additionally, those recommendations regarding data use and collection in the P2D report which remain unutilised be given priority*

6. A **national register of potential data sources** which could be used to supplement agricultural statistics be established, maintained and regularly revised by the ABS, with support and contribution from industry.

7. Efforts to transform the national Australian agricultural statistics system must **align with global best practice** on data management / improvement (e.g. UNECE Big Data project, Eurostat Strategy for Agricultural Statistics for 2020 and Beyond) and that Australian agriculture is appropriately represented at these forums.

8. A **review of sources of data queries** (i.e. on stakeholder needs and collection processes) be published by ABS to help OSAs and industry identify where statistical gaps exist and where data-gathering activities can be streamlined and shared.

9. Industry supports the **redistribution of responsibilities defined for the data lifecycle** by ABS; namely the delegation of data collection, processing and distribution from non-official sources as required to augment official statistics.
    a. *Technical solutions should be developed to facilitate the documentation, publication and discoverability of data by setting up data infrastructure.*

10. A **prompt review of current methodologies and classifications** used by ABS and ABARES be undertaken to ensure efficacious use of existing data and to enable ease of integration for new data sources.
    a. *This review should be completed in collaboration with industry, representative organisations and RDCs.*

11. Industry supports the goal of OSAs to ensure that data sourced for use in the national Australian agricultural statistics system is both **FAIR** (findable, accessible, interoperable, and reusable) **and TRUE** (trustworthy, relevant, useful and explainable).

1. Introduction

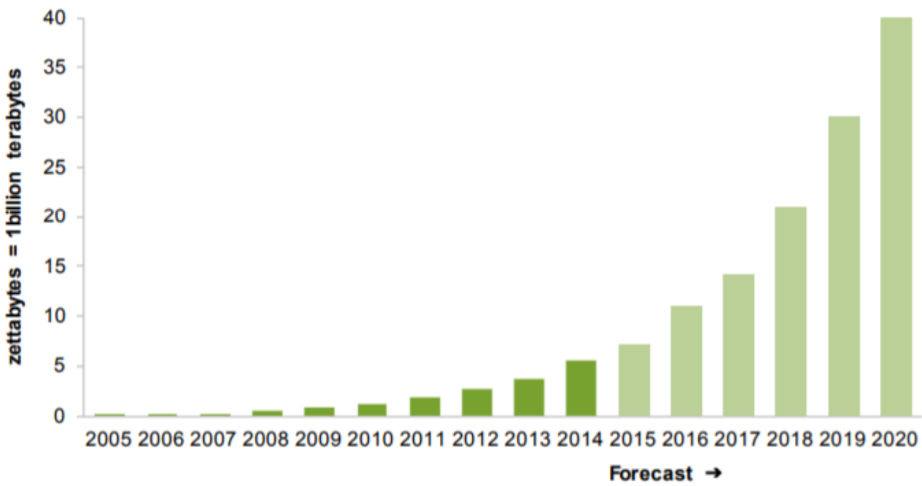
Background

For more than a century, agricultural statistics produced by the Australian Bureau of Statistics (ABS) and more recently the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) and its antecedents have contributed to knowledge and understanding of the Australian agricultural sector and informed decisions regarding agricultural policy, research and development (R&D) funding and activities, and the production, marketing, trade and distribution of commodities.

Agriculture, fisheries and forestry industries make a significant contribution to Australia's economy, society and management of natural resources. The productivity, competitiveness, sustainability and profitability of these industries are enhanced by access to timely, accurate and reliable statistics to inform decision-making by both government and industry (Australian Bureau of Statistics, 2015b).

In the production of such statistics, statisticians have traditionally utilised 'designed' data sources, such as surveys and censuses, and 'found' data sources from administrative and transactional data (Tam & Clarke, 2015b). However, traditional methods of data acquisition are now being challenged by new developments in open data, Big Data, data visualisation and modelled data.

Raw data is now produced by the Internet of Things (IoT) and by digitised transactions at an exponential, almost unimaginable rate (Figure 1). As far back as 2013, IBM claimed that 90% of the data in the world at that time was created in the preceding two years, and that the volume of digital data was expected to multiply another 40 times by 2020 (Eggers, 2013).



[a] Forecast from 2015. [b] Four zettabytes (in 2013) is equivalent to two quintillion jpg images, 456 billion hours of digitally recorded music, one trillion high definition digital films, or 166 billion 32 gigabyte iPads (Larson 2014).

Source: United Nations Economic Commission for Europe (2015)

Figure 1: Data generated globally
Source: (Productivity Commission, 2017)

These new ways of gathering data - and the sheer volume collected - change the nature of information that is available for use and thus the skills needed to interpret it. Additionally, industry changes such as the deregulation of wheat marketing in 2008 (resulting a lost source of industry-wide information) indicate a relatively dynamic data landscape which requires constant monitoring for relevant sources or gaps.

Unfortunately for the sector, the Australian agricultural statistics system has endured progressive degradation over recent decades as the resources made available to the ABS have diminished. With resources restricted, the ABS has been forced to focus on maintaining existing statistical collections rather than developing adaptive systems in response to changing requirements. Private-sector provision of agricultural statistics could (in some cases) challenge the relevance and responsiveness of public sector data.

Governments and industry must rely upon the current statistics portfolio for policy and planning. However, despite recent efforts at improvement, criticism of the scope, robustness, timeliness and utility of official statistics is still commonplace.

Catalysts for change

In the absence of a robust statistics system, Australian agriculture faces a very real risk that policy and business decisions are made without an informed understanding of how the sector operates and is performing. This is a particular risk for agriculture (compared to other industry sectors) due to its dispersed and heterogeneous nature.

Transparency and trust are also areas of increasing importance for the sector, and a lack of reliable data to both quantify and qualify production practices undermines the industry's social licence to operate or right to farm.

The Australian Farm Institute conducted research into Australia's agricultural statistical systems (Potard & Keogh, 2013) in an attempt to improve data accuracy, coherence, consistency, transparency, objectivity and comprehensiveness. Efforts to address some of the recommendations arising from that research have subsequently been initiated, however key recommendations pertaining to clear, consistent leadership and delegation of responsibility, data accessibility and integration are yet to be implemented.

The subsequent 2015 National Agriculture Statistics Review (NASR) highlighted stakeholder concerns regarding the quality of agricultural statistics, including issues with the relevance, accuracy, timeliness, accessibility, interpretability and coherence of statistics produced across the system. These concerns also extended to gaps in data on industry supply chains, value-adding, productivity estimates for some industries, regional scale data, upstream and downstream employment, domestic consumption and labour market supply and demand.

While the ABS is undergoing a review of strategic priorities in the collection and dissemination of agricultural statistics, there are no indications that significant new resources will made available to enhance the system. The review will likely involve a restructuring of existing (or further diminished) resources and will necessitate new approaches to statistical management.

The persistent degradation of the national agricultural statistics system has contributed to declining industry confidence in the information provided by the system. Without intervention and correction, Australian agricultural statistics face a downward spiral whereby the lack of industry engagement and support for OSAs facilitates further reductions in statistical agency resources, necessitating reductions in service output, thereby further exacerbating the lack of industry confidence.

Statistics and data are universally seen as invaluable tools for risk management and effective decision-making, yet confidence in the official sources of agricultural data ranges from moderate to non-existent. Ongoing issues such as the timeliness of reporting, uncertainty of data quality and inconsistencies in categorisation and definitions will further undermine confidence in the system, unless addressed urgently.

The national agricultural statistics system has reached a point where disruption, innovation and reinvention are not merely desirable but necessary.

Options for improvement

These problems are not unique to Australian statistical agencies. They have also been identified in many other jurisdictions and are the subject of several initiatives underway individually or collectively by member nations of the United Nations (UN).

One area of particular focus is the potential of unofficial, non-survey data[1] (sometimes inaccurately referred to as 'Big Data') to replace or supplement official statistical information (UN Statistical Commission, 2014). The ABS has developed a strategy to advance this concept, for example by undertaking trials on utilising satellite imagery for crop production statistics as part of a broader UN initiative.

As it seems unlikely that additional public resources will be made available to enhance agricultural statistics, the redirection of resources and the use of alternate data sources appear to be the best options for systemic improvement.

Recent attention in this regard has centred on Big Data: a term encompassing the capture, processing, analysis and visualisation of potentially large datasets incorporating both structured and unstructured data (NESSI, 2012), such as nation-wide satellite imagery. However, a range of other relevant data sources which could dramatically enhance national agricultural statistics collection remain under-utilised. These include (but are not limited to) state land title databases, national livestock identification databases, livestock and wool

[1] In the exploration of this avenue, it will be important to differentiate between two types of unofficial survey data. One type is effectively privately held data about agricultural activities, such as that held by farmers and consultants. Another is synthetic data that can be generated using research methodologies, such as satellite imagery, modelling, and extrapolation. This second category has the advantage of providing comprehensive coverage, however it is not 'observed' in the strict sense. The first category is observed but suffers from variable quality and coverage.

market information, livestock slaughter data, local government land rates information, transport movement records, rural research and development levy databases, farm benchmarking datasets, ABARES farm surveys and datasets, taxation records (e.g. Business Activity Statements), commercial data (e.g. grain stocks) and market transaction information.

Use of these sources could potentially improve both the quality and scope of Australian agricultural statistics and reduce duplication in record-keeping and administrative burdens for farmers. The lack of coordination between collection bodies in the existing survey-based system often causes great frustration for survey respondents and thus diminished response rates, further reducing data quality and utility.

However, significant administrative, legislative, and cultural barriers are likely to impede the use of much of these alternative data for official statistics, along with a host of technical issues associated with data collection, transmission, manipulation, aggregation and synthesis, administration and ownership. Those limitations noted, the potential of these sources to supplement the national agricultural statistics system is ripe for investigation.

Scope and methodology

The research presented here aims to identify methods of enhancing and enriching Australia's national agricultural statistics system to support improved research, policy, programs and business decisions.

An initial **literature review** (*Section 2*) was conducted to canvass current effective models and methods of statistical collection in agriculture and to note areas of concern.

The next stage of the research involved the identification of **potential sources of data** currently available in Australia that could be used either to supplement or to replace those agricultural statistics now predominantly derived from surveys. Uses and examples of alternative data sources are discussed in *Sections 3 and 4*.

Following this, a **conceptual framework** was developed to enable consistent evaluation of potential data sources, which is presented in *Section 5*. The framework consists of principal requirements to be met before alternative sources of data can be incorporated into official agricultural statistics, including availability, integrity, consistency and accessibility, as well as legal and regulatory considerations. The research has considered a broad range of information sources and has analysed the suitability of some key examples for inclusion in the national agricultural statistics system to demonstrate application of the framework.

The final stage (*Section 6*) involved detailed analysis of information arising from the research, and discussions were held with personnel representing a cross-section of the agricultural industry to gain understanding of sectoral data needs and gaps in the system. This analysis culminates in recommendations for actions to improve the quality of Australia's national agricultural statistics system via the inclusion of non-traditional data sources.

2. Literature review

Australian agricultural statistics date back as far as 1788 and have been collected more periodically with the emergence of statistical analysis in the 1970s-80s. Since then, the primary source of official statistics for the sector has been the *Agricultural Census* (the Census) which is now collected every five years by the ABS. The Census collates information pertaining to commodity production, natural resource management, farm business management and water use, and is used by governments, industry bodies and agricultural businesses in decision making, policy formation and planning for the future of the industry.

Table 1: Main statistical activities and products of the Australian agricultural statistical system

Agency	Surveys	Comment	Content summary
ABS	Agricultural Survey	Every year in alternate- Mail	Data sets include Gross Value Added for agriculture and data associated with national GDP and Non-Agricultural GDP'.
ABS	Agricultural Resource Management Survey		
ABS	Agricultural Census	Every 5 years-Mail	Farm details, production volume by commodities, state, statistical division and NRM division
ABS	Grain Handlers Stocks Survey (GHSS)		Volume by state
ABS	Wheat Use Survey	Monthly- Phone fax to a list of 'contacts'	Volume by state
ABS	Wheat export Sales survey (WESS)		Volume by state
ABS	Livestock slaughtering collection	Monthly- Phone fax to a list of 'contacts'	Number by state
ABS	Price information	No transparent methodology	Sourced from other organisations/informal sources.
ABARES	Australian Agricultural Industry Survey	Yearly - face to face	Indicators of farm business performance and some physical/ natural resource data for broadacre farms.
ABARES	Australian dairy industry survey	Yearly - face to face	Indicators of farm business performance and some physical/ natural resource data for dairy farms.
ABARES	Survey of irrigation farms in the Murray Darling Basin	Yearly since 2006 - funding from DSEWPC	Indicators of farm business performance and some physical/ natural resource data for irrigation farms.

Source: (Potard & Keogh, 2013)

In addition to the Census, the ABS and ABARES[2] jointly collect agricultural data from 31 different initiatives (Appendix 1 – ABS and ABARES statistics). Data collections on agricultural statistics and information are also conducted by a range of other organisations including other Australian Government state or territory agencies, industry, and academic research institutions (Hodges & Lehmann, 2017).

[2] Note: ABARES (the Australian Bureau of Agricultural and Resource Economics and Sciences) is the research arm of the Australian Government Department of Agriculture and Water Resources, not the official statistics collector.

ABARES, and the Department of Agriculture and Water Resources (DAWR) more broadly, rely on a range of data to inform decisions, including official statistics. ABARES accesses data from industry sources (such as research and development corporations), publicly available domestic and international sources and purchases data from service providers. ABARES and DAWR are also significant collectors of data on Australian agriculture, fisheries and forestry through administrative processes and research programs, including performance and practice management surveys (e.g. Pest Animals and Weed Management Survey, 2016). ABARES undertakes four major agricultural surveys annually: the Australian Agricultural and Grazing Industries Survey (AAGIS), Australian Dairy Industry Survey (ADIS), Vegetable Growers Survey and Irrigation Survey. Supplementary questionnaires and other ad hoc collections are undertaken as policy and program needs emerge.

The reduction in coverage and decline in quality of official statistics over many years has impacted the ability of ABARES and DAWR to service industry and government needs. To address these issues, ABARES has worked with the ABS for several years to explore opportunities for collaboration with industry and government organisations and identify opportunities to improve the efficiency of the statistical system and quality of statistics.

The desire to improve agricultural statistics is not unique to Australia. The Food and Agriculture Organization of the United Nations (FAO) report on agricultural surveys (2010) noted that many governments face pressure to cut costs while simultaneously being confronted with increasing and more complex demands for data. The report also identified growing interest in data topics such as food security, environmental impact, labour and special agricultural practices like organic farming.

While the current system is under serious stress, new initiatives both in Australia and internationally offer potential models for adoption by the agriculture sector and promise for systemic reform, even reinvention.

To provide context for the development of a data source evaluation and assessment framework, a literature review was conducted to ascertain the history, trends and developments of agricultural statistical collection.

Recent reports and reviews have highlighted the need for a significant overhaul of the Australian agricultural statistical system, notably *Is counting farmers harder than counting sheep?* by Potard & Keogh (2013) and the 2015 National Agricultural Statistics Review (NASR). These reviews, along with comparable international reports and related programs, are discussed in the following sections.

2.1 Australian reviews and programs

Is counting farmers harder than counting sheep?

Potard and Keogh (2013) posed the question “is counting farmers harder than counting sheep?” in relation to the difficulty of accessing reliable statistics about agriculture. The Australian Farm Institute (AFI) report, which compared the statistical systems in Australia, the United States and France, questioned the adequacy of the system in Australia. The authors addressed the apparent underappreciation for the importance of agricultural statistics despite the critical nature of statistics to inform government, agribusiness and farm business decision-making for the benefit of the sector's productivity, competitiveness, and prosperity (Hodges & Lehmann, 2017; Potard & Keogh, 2013).

The report contains examples where unreliable or uncertain statistics could lead to ill-informed or unsuitable policy or funding outcomes. For example, according to the ABS in 2015 the number of Australian farmers jumped by almost 25,000, despite a steady annual decline in numbers in the two decades prior. Equally perplexing was the reported area of land used for farming in Australia jumping by almost 10 million hectares during 2010-11, again against the long-term trend at the time.

The report concluded that the Australian agricultural statistics system lacked leadership and integration and was hampered by limited resources and declining support in public funding, which combined would lead to the progressive degradation of the Australian agricultural statistics system. Six recommendations were put forward to address these issues:

1. The Australian Government Department of Agriculture, Fisheries and Forestry (DAFF)[3] and more specifically ABARES should be given responsibility to lead and oversee the Australian agriculture statistical system, and carry out this role under the delegated authority of the ABS.
2. In transferring leadership and responsibility of the Australian agricultural statistical system to DAFF/ABARES, negotiations should be entered into between DAFF and the ABS to ensure suitable long-term funding arrangements are in place for both current operations and future developments so that the system more effectively meets government and industry needs.
3. Appropriate statutory provisions should be implemented to reinforce the impartiality, objectivity and confidentiality of ABARES agricultural surveys. Impartiality should be further strengthened by augmenting the statistical skills of the organisation.
4. To ensure that data and agricultural statistics in Australia are readily accessible to stakeholders, the creation of a unique, interactive data warehouse is recommended.
5. Urgent steps should be taken to ensure the various components of the Australian agricultural statistical system are better integrated, and that there is close cooperation between the ABS, ABARES and state agricultural agencies in order to enhance the value of the agricultural statistical system and improve its performance in the future
6. The ABS and ABARES should clearly identify the costs and resources associated with the Australian agricultural statistics system.

The National Agricultural Statistics Review (NASR)

In an almost immediate response to the research by Potard and Keogh, the ABS and ABARES initiated the National Agricultural Statistics Review (NASR) to assess the system's adequacy for informing decisions, planning and policy (Hodges & Lehmann, 2017). The final report released in 2015 investigated the priority information needs of stakeholders, where

[3] Now the Department of Agriculture and Water Resources (DAWR)

these needs were not being met by existing sources of data, potential overlaps and inconsistencies in data, and opportunities to improve efficiency in the system (Australian Bureau of Statistics, 2015a).

The NASR proposed a framework for agricultural statistics collection and analysis consisting of five high-level goals:

1. Competitive and profitable agriculture, fisheries and forestry industries
2. Prosperous communities
3. Sustainable natural resource use
4. Growing trade and market access
5. The protection of animal, plant and human health and welfare

These goals frame the economic, social and environmental dimensions of the agriculture, fisheries and forestry industries and reflect the impact that changes in these industries could have on other aspects of the economy, including health policy, food safety, renewable energy production, emergency management, rural development and international trade competitiveness.

The review identified the deficiencies believed to compromise the capacity of the system to meet current and emerging information needs efficiently. Key themes raised by stakeholders included:

- managing the ‘red tape’ burden on respondents (primarily farmers) resulting from survey activity,
- improving the quality of statistics produced from the system,
- enhancing the statistical infrastructure underpinning the system,
- improving the coordination and governance arrangements in place to ensure the system functions efficiently and effectively,
- the roles and responsibilities defined by the ABS system, and
- the nature of the contributions to statistical data made by government and industry.

The NASR recommended establishing an agricultural statistics consultative forum aimed at engaging stakeholders and driving effective coordination and improved outcomes across the Australian agricultural statistical system. To guide the planning, construction and conduct of surveys, the review also recommended publication of an annual calendar of planned statistical collections to improve accountability of survey managers and reduce duplication of effort.

Other recommendations included:

- Improving the discoverability, accessibility and usability of agricultural data
- Providing a set of common data standards and governance arrangements
    - to support the adoption of best practice methods for data collection, storage and use and
    - to facilitate sharing and integration through interoperable data

- Reducing duplicative collections, leveraging existing data collections and streamlining resource use to reduce data ecosystem costs and improve quality
- Collaborating with industry and government to deliver a modern and sustainable agricultural statistics system that provides high quality data and statistics

Collaborative partnerships supporting the Australian agricultural statistical system

The NASR has now progressed into its implementation phase. At the Seventh International Conference on Agricultural Statistics in October 2016, Hodges and Lehmann (2017) outlined the progress against actions recommended by the review, and the additional opportunities that have emerged since its publication.

Hodges and Lehmann noted that evolution in agricultural practice has been impacted by growing concerns over land and natural resource use and growing interest in the sustainability of agricultural production. The apparent increased interest in the agriculture sector has amplified the pressure on the national agricultural statistical system to meet a range of emerging information needs relating to environmental performance while maintaining the need for accurate, timely and detailed data for industry stakeholders.

ABARES supports the findings of the NASR, which found that the existing agricultural data ecosystem relies heavily on a range of voluntary and compulsory surveys, undertaken by industry and government. A lack of coordination and collaboration between these collection bodies regularly results in respondents receiving multiple requests to provide similar information. The escalating cost of collection activities, rising compliance burden, and increasing frustration of survey respondents has resulted in a narrowing of survey coverage and falling response rates, reducing data quality and utility.

While the ABS and ABARES consider the NASR recommendations to be aspirational in nature, programs are being developed to address these recommendations directly. The prioritisation of the ‘e-form first’ strategy for the 2016 Agricultural Census was an initial step towards addressing the recommendations. There are plans to follow this with focused collaborative engagement to develop a foundation data framework and explore “high-value” administrative sources.

Roadmap to improve the agricultural statistics system

In 2017 the ABS and the DAWR developed a Joint Transformation Strategy to coordinate the longer-term changes to the agricultural statistics system recommended by the NASR, which aims to migrate the current system towards a *collect once use multiple times* model (DAWR & ABS, 2017).

The strategy outlines four streams of activity to improve the efficiency and effectiveness of the collection, use and accessibility of agricultural statistics, and increase engagement of data suppliers (such as survey respondents) and data users.

1. Consolidation of collections
2. Alternative data sources and collection methods

3. Future collection methods and data sources[4]
4. Stakeholder engagement

Notably, Stream 4 focuses on the development and communication of a value proposition for stakeholders to participate in a statistics system. The need for a clear and attractive value proposition, addressing issues such as the need, benefit and integrity of the transformed system, should enable the social licence associated with utilising alternate data sources in place of survey collections.

ABS and DAWR plan to work with stakeholders to ensure duplication of effort is avoided, to facilitate collaboration on common issues and to coordinate projects with system-wide implications. As of February 2019, the ABS and DAWR have responded to the NASR recommendations by:

- Engaging major stakeholders through forums and individual meetings to inform the prioritisation of work programs
- Producing an annual survey calendar to highlight opportunities for improved coordination
- Establishing pilot projects to assess administrative datasets produced by industry and government, including levies collection data, to supplement or substitute current collections
- Building partnerships across government and the research sector to assess the potential for big data, remote sensing and farm management systems, to be utilised for statistics

The organisations have also improved cooperation and collaboration through the establishment of reciprocal in-posting arrangements and streamlined data exchanges.

Productivity Commission Data Availability and Use Report

Noting that effective use of data is increasingly integral to the efficient functioning of the economy, the Productivity Commission (PC) has investigated the benefits and costs of options for improving availability and use of data.

The 2017 report noted that while increased sharing of data across the public and private sectors could improve systemic integrity and increase administrative efficiency, privacy, security and intellectual property issues must be considered in the context of a more open data culture (Productivity Commission, 2017). The report also found that much data being generated remains underutilised (even allowing for the fact that some may prove to have no value), despite technical developments enabling use.

Other key points included:

[4] Areas for exploration in Stream 3 include Big Data, remote sensing, standard business reporting, farm management software, advanced analytics and machine learning.

- Data is a form of capital essential to the production of most goods and services. However, some data is non-fungible, meaning that while the overall volume of data is expanding exponentially, some datasets will have significant scarcity value.
- Legislation on data access was formulated up to a century ago, and much is no longer fit for purpose.
- A lack of national leadership has contributed to piecemeal bureaucratic processes for data sharing and release.
- Extraordinary growth in data generation and usability has enabled a kaleidoscope of new business models, products and insights.
- Lack of trust by both data custodians and users in existing data access processes and protections and numerous hurdles to sharing and releasing data are choking the use and value of Australia's data.
- Marginal changes to existing structures and legislation will not suffice.
- Recommended reforms should treat data as an asset and not a threat.
- The likely incremental costs of more open data access and use should be substantially outweighed by the opportunities presented.

Government programs

The ABS Statistical Business Transformation Program (SBTP) represents a significant investment towards transforming ABS statistical infrastructure, systems and processes, and hopefully an essential enabling step towards also improving the agricultural statistics system.

This program aims to enable the ABS to engage better with partners and develop more responsive solutions, for example by leveraging data sources held by partners to expand the scope of official statistics and enhance data integration by improving the accessibility of ABS data and microdata.

The Australian Government has also committed to making public data more open and accessible while encouraging collaboration with the private and research sectors to extend the value of this data.

While governments and organisations have been integrating data for some years, there has been a renewed recent focus on data partnerships; for example, the Data Integration Partnership for Australia (DIPA) builds on data integration progress made by projects such as the Multi-Agency Data Integration Project (MADIP) and the Business Longitudinal Analysis Data Environment (BLADE).

The DIPA aims to improve technical data infrastructure and data integration while preserving the privacy of individuals and ensuring data security. Although focused on data assets in health, education, social welfare and business, lessons learnt from maximising the use of these assets through data integration and analysis will be instructive for the agricultural sector.

The MADIP has linked existing Medicare, government payments, personal income tax, and 2011 Census data, in participation with the ABS, the Department of Education and Training, the Department of Health, and the Department of Social Services. Data analysis projects undertaken in MADIP have highlighted possibilities for improving Australian's access to the health care system.

The BLADE is a core component of the DIPA which combines business tax data and information from ABS surveys with data about the use of government programs, to provide a better understanding of Australian businesses and the economy. BLADE is not a data set but rather a methodology for linking business datasets by using the Australia Business Number (ABN) as the identifier. The ABS is responsible for combining the data, providing approved researchers and analysts with access via highly secure ABS systems, and safeguarding privacy and confidentiality in collaboration with its partners to ensure that no individual person or business can be identified. BLADE demonstrates how using technology to combine existing public data can help to provide the information required for evidence-based policy development.

2.2 Global reviews and strategies

AAS: Improving crop estimates by multiple data sources

The American National Agricultural Statistics Service (NASS) is responsible for estimating acreage, production and yield for most commodities grown in the United States (US). NASS's county-level estimates support the efficient functioning of agricultural markets by providing information about the supply of, demand for, and use of commodities.

The American Academy of Sciences (AAS) report, *Improving crop estimates by multiple data sources* (National Academies of Sciences, Engineering, and Medicine, et al., 2017), was conducted in response to calls for a new NASS system of gathering data, due to an ongoing decline in survey response rates and the perception that county-level estimates were not representative of individual farmer experiences. US farmers have also pushed for a switch from NASS data to insurance data reported to the US Department of Agriculture (USDA) Risk Management Agency (RMA) as part of their Agriculture-Risk-Coverage program, due to the perceived increased reliability and accuracy of these data.

The AAS research aimed to consider how administrative, remotely sensed and ground-gathered data might be integrated into models to yield more accurate and timely estimates. The resulting NASS vision for 2025 to stay relevant to agriculture and Congress by adopting "the most robust technologies for data management and to utilise non-traditional data sources and statistical methods" includes evolving the role of the Agricultural Statistics Board (ASB) from one of integrating multiple data sources to one of reviewing model-based predictions; macro-editing; and ensuring that models are continually reviewed, assessed and validated. It was also suggested that the NASS should improve transparency and reproducibility by utilising model-based estimates that combine survey data with complementary data in accordance with Office of Management and Budget standards and by delivering predictions through 'dashboards' to facilitate use.

The current NASS approach to integrating multiple data sources is through the ASB. While the current process follows specific steps and guidelines, it is inherently subjective and neither transparent nor reproducible. The development and use of models which accurately and reliably combine multiple data sources have been investigated by NASS. Four modelling approaches that either provide the same information as current ASB process or have the potential to provide such input include:

1. The composite indication – a linear combination of basic indications prepared from direct survey estimates, administrative data, remote sensing data, and estimates made for the past year.
2. Remote sensing – independent indications of plant coverage etc.
3. Berg, Cecere & Ghosh (2014) model for cash rents which uses two univariate area-level models, one for the average of cash rents in two survey years, and the other for the difference; using least squares regression to estimate the model parameters.
4. Approaches to end-of-season model-based crop estimation.

StatCan: Agricultural statistics program review

The Agriculture Division of Statistics Canada (StatCan) collects and analyses extensive statistical information, including the five-yearly Census of Agriculture (CEAG), crop and livestock surveys, farm economic statistics, agri-environmental statistics, tax and other administrative data, research and analysis, and remote sensing.

The Agriculture Division regularly reviews its program, however the 2012 Agricultural Statistics Program Review was expanded to look at the entire program in the context of alternative data sources (Statistics Canada, 2012) - a context that is common to all reviews listed here.

The review included evaluation of the CEAG's efficacy and options to streamline the program to reduce response burden and costs while continuing to meet priority data requirements. The report noted that increasingly complex demands for data are being matched by increasingly diverse and potentially automated data sources and analysis.

The high levels of interdepartmental cooperation and support that will be required to integrate diverse data sources were recognised, however, and in the short to medium term the review suggested that the traditional census should continue in parallel with efforts to develop processes for the use of alternative data sources.

EU: Strategy for agricultural statistics for 2020 and beyond

The EU strategy for agricultural statistics (European Commission, 2015) provides relevant guidance for refining Australia's current agricultural statistics system. It aims to improve coherence between the agricultural statistics sub-domains, reduce respondent burden, clarify and streamline concepts and definitions as well as improve the integration between agricultural, land use, and environmental statistics. The speed of the statistical system is also identified as a priority area.

Agricultural policy priorities, structures and practices are changing rapidly and therefore a great need exists for a system that caters for new data needs and improved systems. New requirements for the type of data collected will also have to be set given the increasing demand to demonstrate the tight links between agriculture, climate, and the environment.

The EU strategy also notes the need for quality improvements and increasing the flexibility and reaction speed of the system for better introduction to address new needs, provide different statistics and take up new approaches. This would need to be supported through a responsive and responsible governance structure that balances the interests of data users and data providers.

Agricultural statistics must cover the economic, environmental and social dimensions of agriculture to a high standard. Special attention was given in this review to the social dimension, noting that it was important for the agricultural statistics system to embrace social statistics to understand vulnerability issues including food security, living conditions and quality in the life of farmers and rural households.

As with the findings of Potard and Keogh (2013), the European Commission also highlights the need for cooperation in the development of agricultural statistics with other statistical domains.

Modelling and forecasting play a large part in statistical processing. The EU strategy highlighted the importance of including all farms regardless of scale and size in statistics and modelling processes. As most agricultural economic activity is concentrated in larger, market-orientated farms, these farms should generally be better represented in the base data of agricultural statistics. Integrating new modelling and statistical techniques will remain a key consideration in the development of a stronger and more representative statistical system, one that is inclusive of all farm businesses regardless of size and location.

Digital technology provides an appropriate platform to provide data at a much faster and more frequent rate. The need to improve the frequency of data collection and timeliness of reporting becomes more critical in the case of unforeseen weather events, where the value of forecast data can be destroyed quickly.

World Bank: Global strategy to improve agricultural and rural statistics

Findings from the European Commission are similar to those of the World Bank (2010) *Global Strategy to Improve Agricultural and Rural Statistics*. The document presents a strategic direction for national and international statistical systems and provides a framework for the production and application of basic data needed to guide decisions. It is the result of extensive consultation with national and international statistical organisations, agricultural ministries and other sections of government (World Bank, 2010).

The World Bank's long-term strategy is dependent on a country's statistical capacity and should be updated on an as-needs basis. It is based on three pillars:

1. A minimum set of core data;

2. Integration of agricultural statistics into the wider national statistics system; and
3. Governance to provide sustainability of the agricultural statistics system.

Although the strategy is broadly aimed at improving agricultural statistics and statistical systems for developing countries, it discusses several shortcomings of statistical systems in both industrialised and developing countries. Statistics to date have had a selective focus, with many countries applying a criterion to minimise costs. The strategy recommends that the omission of data based on size, importance or location should be avoided. The European Commission similarly noted the importance of including all farms in statistical representation, regardless of scale and size (European Commission, 2015).

The division of data collection between government and industry sectors results in the use of different sampling frames and surveys. This compromises the strength and quality of data, makes it difficult to measure the impact of action in one sector to another and can result in the release of conflicting statistics. Where statistical responsibilities are decentralised, it is important to integrate the knowledge of the industry within the experienced modelling of national statistics bodies. The strategy suggests adopting sample designs and synchronising questionnaire/survey design.

Different data producers need to adhere to a common set of standards to mitigate duplication of collection and to prevent the publication of conflicting information. This point corresponds with the discussion of common language for data validation (European Commission, 2015). Common concepts and definitions used across different statistical domains work to reduce inconsistencies between domains and enable users to understand reasons for discrepancies.

The frequency with which data are collected is also important. Data for some units do not change much from year to year, and sometimes data are too costly and difficult to obtain. Data to be collected on an annual basis are those components that account for more than three-quarters of the country's value of production. Data items that fluctuate significantly from year to year should be also adjusted on an annual basis.

3. Utilising 'new' data

The investigation of alternative data sources for the purpose of official agricultural statistics is occurring in many countries and jurisdictions. Drivers that challenge traditional systems of statistics collection include:

- diminishing public resources for agricultural statistics;
- increasingly complex and diverse needs for statistics;
- decreasing participation rates from survey respondents;
- abundance of new data sources;
- rapidly increasing computer power and analysis; and
- possibilities for automated data collection.

However, despite increased global interest, the widespread use of alternative data sources for official agricultural statistics has yet to be achieved. While there is a general recognition of the promise that these data sources hold, there is much investigation needed before accepted processes for inclusion of alternative data sources are developed. Some of these sources and processes are discussed in the following sections.

Big Data

Several international and national statistics organisations have already started to explore the potential for Big Data in the production of official statistics (Eurostat, 2013; United Nations Economic and Social Council, 2013). Big Data in this context is a term encompassing the use of techniques to capture, process, analyse and visualise potentially large datasets in a reasonable timeframe, while incorporating both structured and unstructured data and covering several disciplines and domains (NESSI, 2012).

Following the commentary by Potard and Keogh (2013) regarding the apparent lack of adaptation of official statistics providers to the increased availability and capability of data generated from digital technologies, Tam and Clarke (2015a) discussed the use of Big Data in the regular production of official statistics in Australia. The principal considerations for the inclusion of Big Data were found to be relevance, business benefit and the validity of using the source for official statistics, finite population inferences, or analytic inferences.

Some of the greatest concerns raised by Tam and Clarke (2015a) were whether Big Data provides a benefit compared with existing traditional statistics, whether it fills data gaps, how it compares for cost of data acquisition, the magnitude of error from statistical biases, and how to develop suitable methodologies for the analysis of Big Data sets to produce fit-for-purpose official statistics.

As part of a complementary study, *Big Data, Statistical Inference and Official Statistics,* Tam and Clarke (2015b) discussed the methods of data collection most traditionally used in the production of official statistics and list the issues statisticians can encounter. Their primary focus was Big Data originating from sources such as sensor networks and tracking devices (e.g. satellites and mobiles phones), behavioural metrics (e.g. search engine queries) and online opinion (e.g. social media commentaries). Tam and Clarke expressed recognition for the potential Big Data holds in creating a rich, dynamic and focused picture of Australia and in improving efficiency in the production of statistics, although they argue that adopting Big Data in statistics is not as straightforward as one might immediately imagine or hope.

Despite the ABS's current use of administrative and business data (which are forms of Big Data) Tam and Clarke have expressed relative caution about the potential of these types of sources to augment existing statistical systems. It is argued that exploration of sources beyond administrative and business data would be for opportunistic use at best, and regular reporting of these data without a clear value proposition is unlikely. Therefore, investment by national statistics offices (NSOs) or OSAs would be hard to justify. The authors also investigated whether the use of bigger data sets lead to provision of more accurate information. They stated that the objective of any set of statistics should be to accurately represent a population to avoid hidden biases, rather than just provide a quantity of information.

Tam and Clarke (2015b) contend that Big Data, Semantic Statistics (Clarke & Hamilton unpublished ABS manuscript, 2014), and Statistical Business Transformation currently provide the most promise in the transformation of future business models for OSAs. Their work provides a preliminary statistical framework for assessing the validity of making statistical inference for official statistics.

It could be argued that the Big Data sources identified which have the greatest potential for use by OSAs are, in most cases, directly applicable to agriculture, e.g. satellite imagery and geo-spatial data (discussed in the subsequent section) and remote sensing data. NSOs and OSAs are well positioned to experiment with Big Data, due to their experience and expertise in collecting and processing large quantities of data.

Spatial data

Satellite imagery and geo-spatial data have significant potential to augment official agricultural statistics collection by providing more timely information, reducing the cost and frequency of surveys and by providing data at a more disaggregated level.

The UN Task Team on Satellite Imagery and Geo-Spatial Data is developing a global work plan on utilising this type of data for official statistics and as indicators for post-2015 sustainable development goals. Within this Task Team, the ABS is assessing the feasibility of statistical methodology for classifying satellite surface reflectance data to crop type and for estimating crop production.

The general Statistical Spatial Framework developed by the ABS (Figure 2) is made up of five elements to incorporate population, social and economic data into current geo-spatial analysis:

- Accessible / usable
- Interoperable metadata
- Common geographic boundaries
- Geocoded unit record data

- Authoritative infrastructure / geocoding

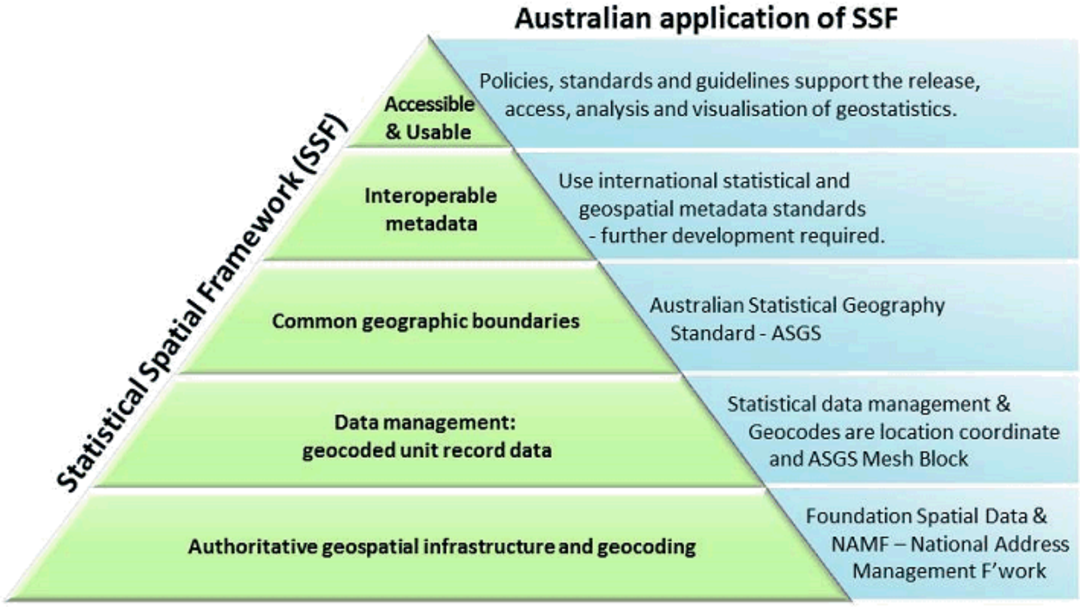


Figure 2. Australian application of the Statistical Spatial Framework

Source: (ABS, 2018)

Quantifiable statistical reliability and computational feasibility are key concerns when integrating geo-spatial data into statistical collections (e.g. using satellite imagery data to estimate crop area statistics), and investigation into the appropriate methodology is ongoing. The lack of adequate ground-truth data available in Australia to inform classification methods has been a major obstacle to date.

However, the unit record level data collected by the ABS as part of the Rural Environment and Agricultural Statistics Program could provide a rich source of reference data on which to train classifiers and base classifications.

3.1 Examples of new data projects

Big Data Flagship Project

The ABS Big Data Flagship Project was established to coordinate research and development to build a methodological foundation for the mainstream use of Big Data in statistical production and analysis. Tam and Clarke (2015b) discuss the project as an opportunity for the ABS to gain experience in assessing the business, statistical, technical, and computational issues related to Big Data collection and utilisation.

The only work package scheduled in the project that directly refers to agricultural statistics is the use of remote sensing data for land use, crop type and crop yield statistics. However, a number of other work packages could also contribute to the enhancement of agricultural statistics, such as those that use predictive modelling, visualisation and analysis of multiple connections in linked data.

Better Practice Guide for Big Data

Some of the challenges and opportunities associated with the use of non-traditional data sources by OSAs in the production of official statistics are canvassed in the *Better Practice Guide for Big Data* (Commonwealth of Australia, 2014) developed by the Data Analytics Centre of Excellence. The guide lists the key issues and challenges for use of Big Data for official statistics as:

- Methodology
- Privacy and public trust
  - Statistical legislation: determines how data sets can be acquired, combined, protected, shared, exposed, analysed and retained
  - Mosaic effect
- Computational efficiency
  - Existing computational models for the most common statistical problems in the typical NSO scale very poorly for the number, diversity and volatility of data elements, attribute and linkages associated with big data sources
  - A potential solution could be to outsource analytics to the data owner – although this would create a new set of challenges
- Technology infrastructure

The UNECE High-Level Group for the Modernisation of Official Statistics (HLG-MOS) has also sponsored a series of international collaboration projects to understand how to harness the power of Big Data and other new data sources, to support the production of official statistics.

The P2D program

The Accelerating Precision Agriculture to Decision Agriculture (P2D) program (Leonard et al., 2017) contained a report by the Australian Farm Institute on *The economic impact of digital agriculture* (Perrett, Heath, Laurie, & Darragh, 2017) which made several recommendations on the use of Big Data in an agricultural context which have relevance for methods and models of data integration.

The authors noted that a ‘data lake’ is generally much more valuable than a ‘data silo’, i.e. that the aggregation of several data sets is likely to increase the functionality of data sets to users. Data silos currently exist within and across Australian agricultural industries, creating challenges in accessing, integrating, and analysing data.

To facilitate the development of new service delivery models, the report recommended major initiatives to improve several data analytics issues, including:

- improving education and training to increase the capability of users of digital agriculture and Big Data
- resolving the issues of different data governance approaches (e.g. inconsistent data systems architectures and ontologies) by government and industry service providers (i.e. RDCs).

In addition, legal and trust issues associated with agricultural data are emerging as a critical challenge for Australia's agricultural industries. Another P2D report found that many Australian producers lack trust in service and technology providers when collecting and sharing their data (Zhang, Baker, Jakku, & Llewellyn, 2017) and are concerned about third parties gaining unauthorised access to their data. Currently, there is little or no legislation in Australia that deals specifically with data. Trust is explored in more detail in Section 5 of this report on considerations for assessing data sources. Clarification on data ownership, control and access issues to address data privacy, safety and security is still lacking in the Australian agricultural sector.

All P2D project groups also recommended that the 15 RDCs work collaboratively to develop consistent approaches to data policies and operational procedures (e.g. data formats, systems architectures etc.) that will improve the interoperability of data sets; to review opportunities to share data sets, data tools, and associated services, and to collaborate on establishing a set of Foundational Data Sets for cross-industry use (Leonard et al., 2017).

CSIRO Digiscape and Conflux

Digiscape is one of CSIRO's six Future Science Platforms which aim to underpin innovation in health and biology, resources, agriculture and manufacturing. Utilising networks of sensor data and smart analytics to match crops to future soil and weather conditions, Digiscape is building a common big data infrastructure that will 'support next generation decision-making' to make agriculture more profitable, lower impact and lower risk.

The Digiscape 'uncertainty analytics toolbox' aims to quantify predictions and forecasts of agricultural systems by identifying how outputs from models can be judged, where models break down, where more monitoring data needs to be collected or where expert information is needed. Digiscape's projects include:

- Aquaculture
- Graincast: forecasting Australia's national grain crop yield
- Digital services for carbon farming markets
- Helping sugarcane farmers protect the Great Barrier Reef
- On-farm experimentation
- WaterWise
- Social dimensions
- Data staging: Conflux
- Uncertainty analytics toolbox
- Improving Australia's digital soil map
- Climate and weather forecasting

Constructed using CSIRO's existing Senaps technology, Conflux is a data staging service specifically designed for agricultural sector applications. It transforms raw data into product-ready information suitable for Digiscape's services and applications and allows stored or real-time sensor data to be combined with the predictive models required by each of the Digiscape applications. Conflux will:

- ensure that the knowledge and information derived from the staged data is trusted and traceable, through quality assurance and provenance services
- guarantee data integrity/security, and the privacy of its sources
- embed existing predictive models (e.g. APSIM) into larger decision support work flows that depend on sensor data
- allow third-party services and applications to have easier access to richer data and better efficiency by re-using common building blocks across application domains (i.e. “develop once, use many times”)
- unite disparate environmental information into common data models.

4. Sources of data

Given the unlikelihood that additional public resources will be made available to enhance the existing agricultural statistics system, the use of other data sources to enhance or replace current statistical information appears to be the best option to both sustain and improve the system.

While recent attention in this regard has focussed on Big Data, a range of additional sources could provide valuable supplements to the current national agricultural statistics collection. These include state land title databases, tax records, national livestock identification databases, livestock and wool market information, livestock slaughter data, local government land rates information, livestock transport movement records, rural research and development levy databases, farm benchmarking datasets, ABARES farm surveys and datasets, and value chain / transaction information (e.g. commercially held datasets).

In total, the ABS and ABARES currently collect agricultural data from 31 different initiatives, 25 of which are conducted on an annual or more frequent basis. Data collections are also conducted by states and territories and by industry bodies, for example Forest and Wood Products Australia, Meat and Livestock Australia, Australian Wool Innovation, the Grains Research and Development Corporation, GrainGrowers, the Australian Wool Exchange, the Cotton Research and Development Corporation and Dairy Australia amongst others. Current ABS and ABARES publications and data portfolios are listed in Appendix 1.

The utilisation of other sources as supplements could reduce duplication in record-keeping and administrative burdens for farmers, while potentially improving the quality, scope and timeliness of Australian agricultural statistics. However, significant administrative, legislative and cultural barriers are likely to impede the use of much of this data for official statistical integration, along with a host of technical issues associated with data transmission, manipulation and administration. Methodological challenges in integrating different administrative or Big Data sources to generate representative population statistics (to address issues of duplication, bias, gaps and social licence) will also impede adaptation to new collection methods. Given the seemingly limitless potential of new data sources and uses, potential trade-offs will naturally occur in the delegation of limited resources for collection and synthesis.

Despite these limitations, there is considerable merit in exploring the full potential of these alternative data sources. As an exhaustive examination of all possible sources is infeasible (with new sources are appearing in the sector frequently), this report has identified a cross-section of sources which can be assessed for suitability to augment or replace current national agricultural statistics.

Potential-use case study: Pork industry

The Australian pork industry includes seven export accredited abattoirs across the country:

- Sunpork Fresh Foods, Kingaroy, Queensland *(Sunpork Fresh Foods, n.d.)*
- Booyong Service Processing, Booyong, NSW *(Booyong Service Processing, n.d.)*
- Rivalea, Corowa, NSW *(Rivalea Australia, n.d.)*
- Diamond Valley Pork, Laverton, Vic *(Diamond Valley Pork, n.d.)*
- Big River Pork, Murray Bridge, SA *(Big River Pork, n.d.)*
- Primo Australia Port Wakefield Abattoir, Port Wakefield, SA
- Linley Valley Pork, Wundowie, WA *(Linley Valley Pork, n.d.)*

These abattoirs slaughter the majority of all Australia-grown pigs, with some smaller, non-export accredited abattoirs processing carcasses to be used for domestic consumption only. They also de-bone carcasses on site and process into retail products. In addition, there a number of companies which are purely manufacturing and do not slaughter pigs on site, such as BE Campbell (BE Campbell, n.d.) and Wilmeat (Wilmeat Cut Meat Pty Ltd, n.d.) both located in Sydney, NSW.

Although some facilities may be processing thousands of animals per day, carcase attribute data is collected quickly and accurately through the utilisation of technology. This is usually through assigning data to a barcode which is attached to each individual carcase through a tag which is scanned along different points of the supply chain. Many facilities have daily reconciliations which occur to ensure all animals are accounted for in processing which relates to timely data availability.

Carcase attributes which may be collected in data can include: Hot Standard Carcase Weight (HSCW); Cold Standard Carcase Weight (CSCW) after shrinkage in chillers and trimming; P2 fat reading; gender; and partial condemns of the animal (such as bruising, abscess or arthritis).

Increased accessibility to the commercial datasets from abattoirs, processors, or a combination of both, would enable Australian pig growers to benchmark their farm performance against national or state averages in areas such as weight, fat and carcase damage and thus improve decision-making and overall business performance. High-performing farmers could also utilise this knowledge as a marketing or price negotiation tool – for example, demanding a higher base price due to the higher consistency in weights compared to other farms supplying the business. Improved individual farm performance will benefit processors via the flow-ons effects of increased efficiencies through higher quality carcasses with consistent weights and less damage.

Australian Pork Limited (APL) collects a large amount of data from abattoirs and manufacturing plants such as slaughter numbers and average carcase weights, which is distributed to some value chain participants but is not widely available in an easily accessible location. If these collated data from APL on national figures and averages were combined with abattoir and processor commercial data sets, aggregated information could be provided back to OSAs to enhance the scope and timeliness of national statistics on the pork industry.

However, the small number of stakeholders dramatically increases the risk of privacy breaches via the mosaic effect. As in many agricultural subsectors, concerns on privacy and commercial sensitivity are a significant barrier to the open sharing of data. Collating several commercial datasets into national averages could potentially address this issue.

Potential use case study: Grains industry

The dismantling of single desk marketing arrangements for the Australian wheat industry in 2008 removed an important source of industry-wide data. Wheat market information is currently provided by both public and private bodies, but inconsistencies and gaps have prompted debate on the transparency of market information and concerns over market functionality.

Reliable information is key to a well-functioning, competitive and transparent marketplace, however bulk handling companies (BHCs) now retain information on the grade, location and volume of stocks that is unavailable to the rest of the market. Some stakeholders believe this has resulted in market asymmetry, an unfair market advantage and lower grower returns (Reading, 2012). There are several bulk handlers in the Australian grains industry located around the country:

- GrainCorp - Fisherman Island, Gladstone, Mackay, Newcastle, Port Kembla, Portland, Geelong
- QBT - Brisbane
- NAT - Newcastle
- Quattro - Port Kembla
- Emerald - Melbourne
- Riordan - Geelong
- Viterra - Port Giles, Port Lincoln, Thevenard, Wallaroo, Port Adelaide Outer Harbour, Port Adelaide Inner Harbour
- LINX - Port Adelaide Outer Harbour
- Semaphore - Port Adelaide
- CBH - Kwinana, Esperance, Albany, Geraldton
- WAPRES - Port of Bunbury

BHCs are unwilling to provide their information, citing commercial in-confidence issues as a restriction, and have claimed that increased transparency could adversely affect farmers by revealing detailed stock and quality information to competing international interests (Heard, 2017). Various suggestions have been made regarding this situation, including:

- Grower groups have repeatedly called for BHCs to be compelled by legislation to supply transparent information on stock levels (Reading, 2012) and be required to put in place mechanisms that avoid vertically integrated companies from advantaging their trading arms through access to information not available to other exporters (AGEA, 2011)
- GrainCorp has noted the need for on-farm storage to be included in any calculations of aggregate volumes (Heard, 2017)
- Aggregated data on the volumes of uncommitted warehoused grain held in commercial bulk storage and handling systems should be made available by grade and location on a weekly basis (Reading, 2012)
- The introduction of a mandatory Grains Stocks Reporting scheme in Australia and
    - free public access to ABS monthly grain export data
    - An ACCC market study into the grains supply chain
    - Extension of the Wheat Ports Code to all grains to improve reporting requirements (“Grain Growers 2019 Election Platform,” n.d.)
- Stock levels could be easily collected based on the warehousing invoices sent out and reported on a quarterly or monthly basis

Potential use case study: Grains industry (cont.)

Growers believe provision of additional BHC-held data will increase competition from traders who would then not need to include a 'risk component' of incomplete information in pricing. In 2012, the potential price-benefit to growers from the provision of additional data was suggested to be $2-3 per tonne (Reading, 2012). Additionally, if current stock levels of different grades were made readily available then growers could more efficiently market supplies accordingly to account for gaps.

The transparent distribution of market information improves price discovery, increases competitiveness, creates more transparent price signals, and reduces transaction costs and variability in markets.

While grower groups maintain that better stocks reporting will increase grower returns and reduce grower risks, concerns on potential erosion of competitive advantage (CBH, 2011) must be balanced if BHCs are to be compelled to release this information.

4.1 Considerations for assessing sources

A number of important issues consistently arose in interviews conducted with Australian agricultural statistical stakeholders and throughout the literature review which should be considered when identifying and assessing data sources for statistical use. These issues – including availability, integrity, harmonisation (national and global), respondent burden, fitness for use, user needs, partnerships and legal issues – are discussed in the following section.

Data availability

Perrett et al (2017) noted that new technologies and business models are driving a rise in private sector collection of data, especially weather data, with the volume of on-farm data collected and stored increasing exponentially in recent years.

However, the availability of appropriate data has been constrained by poor connectivity. Slow data upload speeds in rural areas limiting the velocity of agricultural data transmission constrain the capacity for producers to utilise cloud-based data storage systems, and many producers are still reliant on paper-based data storage systems.

While Big Data is seen as a 'holy grail' for improving agricultural statistics sets, collection is currently haphazard and unreliable. There is an increasing need for investment in foundational public-sector datasets, particularly those pertaining to soils, water and climate/weather, that provide a reference point critical to the validation of privately collected data. In addition, an industry workshop conducted for the P2D project concluded that historical datasets from Rural RDCs and other public research was often very difficult to discover and access.

FAIR and Open Data

The growing movement to publish official information in a way that everyone can access, use and share has spawned several initiatives focused on data sharing and availability, such as the Open Data Charter and the Data FAIRport.

The Open Data Charter principles were developed in 2015 to represent a globally-agreed set of aspirational norms for how to publish data. The six principles, which have since been endorsed by more than 90 governments and organisations, are:

1. Open by default
2. Timely and comprehensive
3. Accessible and usable
4. Comparable and interoperable
5. For improved governance and citizen engagement
6. For inclusive development and innovation

The Data FAIRport initiative started as the practical follow-up of a Lorentz Workshop in The Netherlands in January 2014. Participants agreed that widespread adoption of a minimal set of guiding principles and practices would enable data providers and data consumers - both machine and human - to discover more easily, access and sensibly re-use more easily the vast quantities of information being generated by contemporary data-intensive science.

The term FAIR was coined to describe principles to make data Findable, Accessible, Interoperable, and Reusable. Based on these principles, a set of metrics have been defined to quantify levels of FAIRness (which are detailed in the Appendix).

Despite growing support for open data, there is an expected difference in the level of data available to the public versus that which is kept in-house due to privacy or commercial concerns. For example, while Australian agricultural RDCs collect information as part of grower, industry and government funding mechanisms, and are required to make data available to meet reporting requirements to substantiate that funding, some data still remain unpublished and inaccessible.

The ABS acknowledge that more could be done to institute a culture of open data across the system and notes the challenges lie in improving the accessibility and discoverability of ABS data and asserting a stronger leadership role in encouraging other stakeholders to release more data. Protecting confidentiality in published statistics is important for ensuring the trust and cooperation of businesses.

Primary issues with data interoperability to date have been associated with ANZSCO classifications (Australian and New Zealand Standard Classification of Occupations), data to inform biosecurity and data to capture whole-of-industry information.

The use of different data formats by industry and government (Perrett et al., 2017) is compounding the interoperability challenges and is a critical barrier to systemic change. This includes inconsistencies in operational procedures and policies governing the use of data, e.g. the range of competing data systems architectures and data formats.

Adopting and adapting for Big Data

Big Data (for example from satellite images and remote sensors) are by nature messy and unstructured. It can be hard to separate information from noise, and the gathered data are often not fit for purpose.

Big Data requires no sample, is collected in real time, captures real behaviour and is not based on self-declaration which reduces respondent burden and improves accuracy. On the other hand, the questionable representativeness of the data generated, and potential lack or over-coverage, could distort the reliability of estimates. Big Data has also a comparability problem with data from traditional statistics (Florescu, et al., 2014; Landefeld, 2014; Scannapieco, Virgillito, & Zardetto, 2012).

As Big Data are not collected on probability samples of the target population, there is a need for a fundamental change not only in data collection and dissemination but also in the methods of statistical inference. Models such as Calibrated Bayesian analysis that yield inferences with robust repeated sampling properties are likely to be more appropriate for statistical analysis than the classical methods. Machine-learning techniques can potentially augment traditional modelling methods such as Bayesian techniques.

However, inferences drawn from Big Data need to be carefully validated as machine-learning methods for unstructured data are unlikely to be accurate (Braaksma & Zeelenberg, 2015; Little, 2015).

While the modelling of Big Data may be imperfect, OSAs should decide what Big Data projects to undertake on a case-by-case basis and select projects to complement existing statistics sets. By noting initiatives in other sectors and where possible identifying opportunities for cooperative efforts, the industry's adaptation to the use of Big Data can be accelerated. Moreover, to keep abreast of developments, agencies should proactively search for Big Data sources to address the most urgent research needs.

Investment in data-gathering technology and a stronger incentive for producers (i.e. a clear expression of industry-wide and sector-specific value propositions) to adopt and use Big Data generating technologies and services would enable a greater pool of available data to be made available to enhance existing statistics.

Needs of users

Different users of statistics have very different needs. The substantive value of official statistics as perceived by users of those statistics should be reviewed on a regular basis to ensure fitness for use (Task Force on the Value of Official Statistics, 2017) although it should be noted that some users will not express their data needs, and some requests for data could be mal-intentioned. While non-users may not express specific needs regarding official statistics, these needs nonetheless exist and could arise in due course if the non-user is made aware of the availability and value proposition of the statistical information.

Statistical literacy and knowledge of the information available are key to effective implementation of data to inform decisions.

Some NSOs and OSAs have established user segments or personas[5] to improve customer service and to develop products and services that better meet specific user needs. The European Statistical Advisory Committee (ESAC) classifies users into institutional (international organisations, agreements or initiatives) and non-institutional (divided, based on the basis of their interest in statistics).

Ultimately, statistical offices have a responsibility to consider the relevance of statistics to society and to protect the rights of data users and providers.

Respondent burden

The focus of many national statistics agencies has been to lessen reporting burden. While the belief that increased respondent burden is negatively correlated with survey cooperation is common, some studies support this view and some find the opposite.

An analysis of the relationship between survey burden and non-response – *If we bother them more, are they less cooperative?* (McCarthy, Beckler, & Qualey, 2006) – found that the survey burden as traditionally defined does not uniformly affect future survey response.

Conventional measures of burden have been assumed to be directly related to survey cooperation, but they may in fact have little effect on response. Differences in the feelings potential respondents have about the survey sponsor and the perceived effect of survey statistics on respondents appear to be much more closely related to survey cooperation or refusal than burden. The authors suggested an alternative strategy to address the issue might be to increase the burden on a smaller group of respondents and forgo the objective of making the burden as small as possible for everyone.

As recognised by the ABS and ABARES, balancing the need to meet users' statistical requirements with the need to manage the burden on respondents effectively is an ongoing challenge for producers of statistics.

Statistics New Zealand published a Collections Strategy for 2010–20, citing "improving the respondent experience" as the strategy's first theme. Respondent experience is important as it affects both willingness to comply with requests for data and perception of how onerous it is to do so. The strategy explicitly recognises that respondent engagement involves more than managing survey load.

---

[5] User segments/personas are representative profiles or characters created to mimic real users or clients, based on foundational demographic information collected through research.

Trust and legal concerns

Legal and trust issues associated with agricultural data are emerging as a critical challenge for Australia's agricultural industries.

Many questions regarding data ownership and control in Australian agriculture remain unanswered, particularly in terms of online data upload for analysis. Many Australian producers lack trust in service and technology providers when collecting and sharing their data (Zhang et al., 2017) and are concerned about third parties gaining unauthorised access to their data.

Greater transparency around terms of use that govern the aggregation, ownership, storage and dissemination of producers' agricultural data is needed. The Australian regulatory landscape is changing quickly, however there is little or no current legislation in Australia that deals specifically with data.

In *The economic impact of digital agriculture* (Perrett et al., 2017) noted the need for major initiatives to improve several data related issues, including:

- Clarifying data ownership, control and access issues – including potential barriers posed by Intellectual Property (IP) regimes such as copyright, the current contractual practices that regulate data ownership, control and use of agri-data and, importantly, the 2017 Productivity Commission review into data availability and access;
- Improving data privacy, safety and security – which raises issues of privacy, confidentiality and contracts; and
- Ensuring data are transparent and trusted – which raises issues of industry guidelines (e.g. US Farm Bureau, NZ Dairy) and contracts (open and transparent).

As well as providing far more accurate information on industry and sector statistics than is currently provided through traditional surveying methodology, more confidence in data sharing will help inform research gaps and requirements through objective information about agricultural practices and production.

However, before a usable pool of data can be accessed to augment the statistical system, Australian agriculture should work towards developing a clear and consistent national voice in relation to developing understandable, ethical and efficient agricultural data practices.

Fitness for use

Quality of statistics is one of the key features of data influencing the value that users experience. Quality is also described as "fitness for use", which includes dimensions such as relevance, accuracy, timeliness, accessibility, comparability and coherence.

OSAs do not have a monopoly on data quality. Private data providers may even be able to provide users with some better features, for instance improved timeliness or relevance. However, it is undoubtedly true that the value offered by official statistics is supported by their legal and institutional framework which ensures the compilation of objective and independent statistics that are not subject to inappropriate influence.

One of the strongest motivations for gathering data and producing information is its usefulness in evidence-based decision making. Relevant statistics fulfil one or several of the following characteristics: they have many users, are essential to the fulfilment of the mandates of several organisations, facilitate trade or development, their release causes reactions at the markets and their unavailability creates inequities or asymmetric information (Task Force on the Value of Official Statistics, 2017).

A review of the sources of data queries (i.e. from government and industry) could identify where activities can be streamlined and shared. This may require a review of legislation and/or sharing capabilities in order to share the information effectively between stakeholders.

To ensure granular accuracy, coverage of agricultural statistics should be exhaustive and include all farms in statistics, regardless of scale and size.

Strategic partnerships

In 2014, the High-Level Group for the Modernisation of Official Statistics (HLG-MOS) carried out a survey which identified engagement in data partnerships by OSAs and other organisations. HLG-MOS found the most common type of partnership for OSAs to develop was with data providers in the government sector, followed by analytical partners. Other partnerships included those with data consumers, design partners and technology partners. The relationships identified by this study indicate that access to data is currently the main reason for engaging in partnerships.

OSAs such as ABS and ABARES do not have the capability (whether by fiscal constraint or internal systems) to address all high priority data needs alone. An evaluation of the value of statistics, priorities of stakeholders versus those of OSAs, and reporting requirements and capabilities of OSAs would be highly beneficial to determine which data partnerships would be most effective.

Responsibilities need to be defined for the data lifecycle; primarily, what information will be collected, who collects the data and who distributes the data. Technical solutions should be developed to facilitate the documentation, publication and discoverability of data by setting up appropriate data structures.

The Data Integration Partnership for Australia (DIPA) is an example of a data integration project seeking to create new datasets by linking data from different sources in privacy preserving ways. One DIPA case study which could have application for agricultural statistics is the Western Australian Department of Health Data Linkage System established in 1995. It was the first of its kind in Australia to undertake systematic integration of health and administrative data for research and statistics. Over 800 projects from academia, government, and hospitals have used WA’s linked data.

Standardisation

A prompt review of current methodologies and classifications used by ABS and ABARES is required to ensure successful use of existing data and easy integration for new data sources. This review should be completed in collaboration with industry, representative organisations and RDCs. Standardisation of methods and classifications will enhance the accuracy of data collections made by ABS and ABARES, generate value for industry and validate levy investment and reporting requirements.

A common set of standards that different data producers must adhere to would also prevent duplication and the publication of conflicting information. An agreed international approach to agricultural statistics (for example Integrated Farm Statistics) and Big Data source classification would add more value not only to Australian agricultural statistics but also allow better global comparisons and interpretation of data.

Greater transparency is required in the methods and logic behind the development of national statistics initiatives and actions, including descriptions of the methods used for data collections, processing and dissemination.

An objective of both industry and government should be to make agriculture-related data discoverable and accessible for a broad audience of modellers and decision makers in and outside of the sector. Mandatory documentation of datasets with appropriate metadata would enable 'findability' of data in line with the FAIR principles.

4.2 Types of alternative data sources

This section describes and discusses relevant data sources which have the potential to be utilised as alternative inputs to the Australian agricultural statistics system. Alternative data sources can be broadly categorised as:

1. Social information (human-sourced data)
2. Business systems (process-mediated data)
3. Internet of Things (machine-generated data)

The Australian agricultural industry is already making progress on the exploration of using alternative data sources. For example, the *Accelerating Precision Agriculture to Decision Agriculture* (P2D) program resulted in a report by the Australian Farm Institute on the economic impact of digital agriculture (Perrett et al., 2017) which noted industry-led Big Data integration or investigation projects, including (but not limited to):

- the Sugar Research Australia (SRA) Strategic Plan 2017/18–2021/22 (SRA, 2017) considering the acceleration of disruptive Big Data, sensors and smart connected technologies to drive innovation in data analysis and decision-support tools
- Wine Australia's VinSites pilot, a consolidated national dataset for the wine sector that provides critical data to track Australian wine from the vineyard through to consumption

- the Meat and Livestock Australia Digital Value Chain Strategy (MLA, 2016), which aims to enable the integration and interpretation of data generation within the livestock industry, to ensure value chain stakeholders are connected through open data and to develop a user-friendly data platform that will improve decision-making for farmers and business
- the Sheep CRC enhanced sheep wellbeing and productivity program, which aims to produce risk predictions based on Big Data applications that draw on weather data, analysis of the Information Nucleus database and regular monitoring to identify management factors influencing risk to wellbeing and productivity.

As depicted in the *Roadmap to improve the agricultural statistics system*, the Department of Agriculture and ABS (2017) maintain a focus on assessing administrative data collected by industry and government agencies for their potential to substitute for surveys and reduce respondent burden. The Joint Transformation Strategy is continuing to explore administrative datasets such as:

- Levy-payer records
- supply chain data from processors, traders and testing authorities
- export documentation system (EXDOC)
- Australian Tax Office data, such as Business Activity Statements
- Australian Wool Testing Authority classification data, and
- National Livestock Identification System traceability data.

Identifying (or in most cases, accessing and disseminating) alternative sources of data, such as these administrative data held by government agencies and private organisations, offers the potential to improve their ability to satisfy industry and stakeholder demands (Williamson & Nicholls, n.d.).

The linking and improved utilisation of administrative datasets has been a general strategy employed by OSAs for decades, both in Australia and internationally. As such, this report has focused more on alternative data sources not typically classified as administrative data, such as those derived from Big Data.

Technological advances and the subsequent increasing frequency and type of data generated continue to drive the need to identify alternative complementary data sources to improve our understanding of agriculture, particularly in environmental, economic, and cultural contexts (Zipper, 2018). Data required to answer new (and old) questions and increasing demand for additional information, is not easy to find, and increasingly the satisfaction of these demands results in additional challenges for statistical organisations (Williamson & Nicholls, n.d.).

Darnell et al. (2018) predict that much of the important data required about farm business will be collected ad hoc by individuals. However, information bases applicable to multiple industry sectors, such as soils, weather, climate and land use data (much of which could be collected by remote sensing) have the greatest potential for improvement and to inform the most effective policy setting.

Globally, several products and services that offer business solutions could provide options for data preparation and simplification of large and diverse data sets, including data blending, cleaning, integration, scalability, security and self-service management tools. In Australia, most of these products and services are delivered via an online platform or portal, primarily by independent providers or corporations. Additional sources of information more specialised in focus are also available for other areas, such as commodities trading, and provide access to local, domestic and export buyers.

Using these systems, products or services, significant potential benefit can be derived from data associated with supply chain activities and the various interactions that take place at each given stage.

However, as discussed in the previous section, the application of Big Data to national statistics requires careful consideration. For example, the uneven adoption of computing and sensor technologies, (despite being used on Australian farms for more than two decades) limits the effective contribution of these systems as data sources. The application of data science and machine-learning techniques to augment traditional modelling methods (such as Bayesian techniques) can be of benefit in such situations where the data source is not widely representative.

Alternative data sources with potential for integration with official agricultural statistics in Australia are outlined below, based on the classifications developed by the UNECE Task Team on Big Data in June 2013.

Social information (human-sourced data)

This information is the record of human experiences, previously recorded in books and works of art, and later in photographs, audio and video. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned. Sources include:

- Social networks
- Personal documents
- Internet searches
- User-generated maps

An intriguing possibility for aggregating publicly available, user-volunteered data exists in the utilisation of social media. Digital social networks are important to consider, as a farmer's decision-making process contains a definitive social aspect which should be taken into account when considering information to include for national statistics. Data from social media can also provide insight into primary production pain points, including climatic or biosecurity concerns, describe the status of the relationship between producers and consumers, and identify influences on consumer buying behaviour which can provide early indications of market trends or supply chain issues.

Zipper (2018) notes several studies which demonstrate the ways in which social media can be used for environmental research, including species distribution and biodiversity assessments studying animal behaviour, mapping hazards such as wildfire or earthquakes; water resource monitoring; and studying public perception of environmental change.

However, data collected from social networks is usually unstructured and unverified, and specific protocols must be developed in order to utilise this kind of information. A paper for the *International Journal of Information Management* (National Farmers Federation, 2013) notes that social media analytics is a highly complex process, and as such it is necessary to standardise utilisation via a process model such as that depicted in Figure 3. This model by Stieglitz et al demonstrates the need to carefully consider research domains and tracking processes in order to extract benefit from social data for statistical purposes.

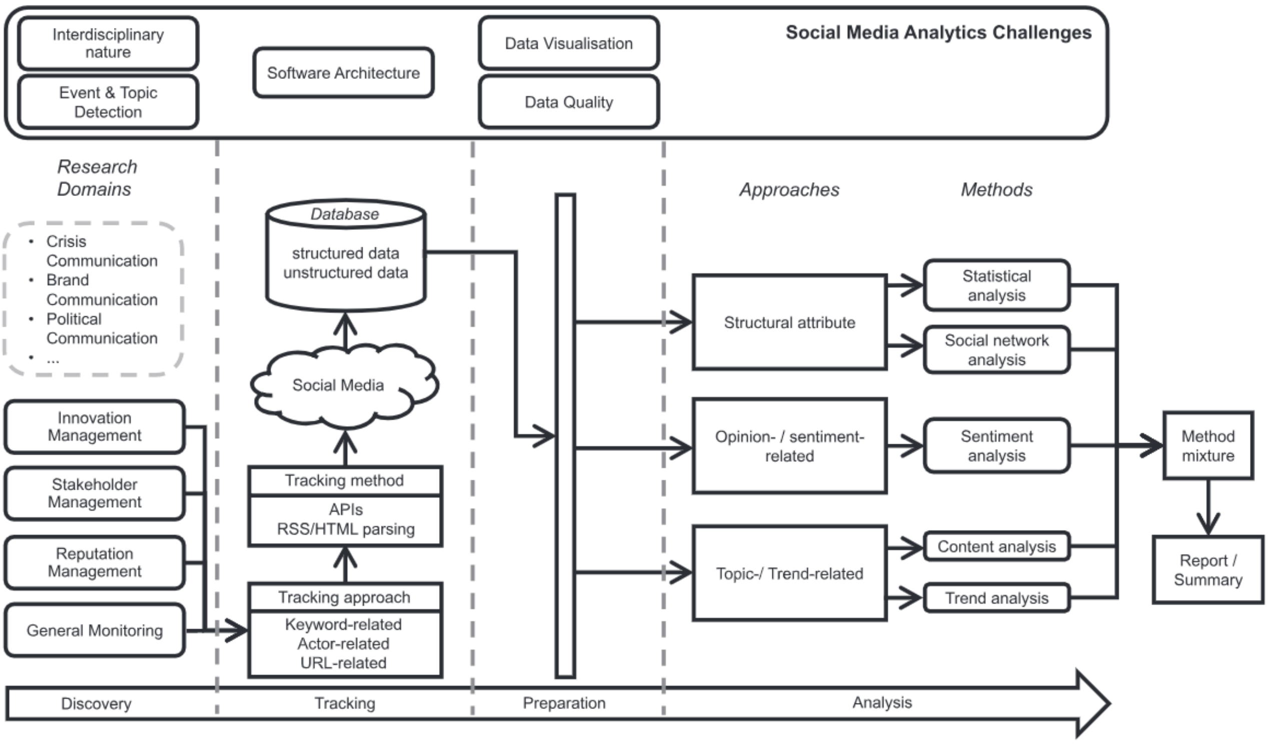


Figure 3. The identified challenges in the context of the Social Media Analytics Framework. Source: (Stieglitz, Mirbabaie, Ross, & Neuberger, 2018)

Business systems (process-mediated data)

Traditional business systems record and monitor business interactions and events, such as procuring materials, registering customers, manufacturing products, completing sales etc. The process-mediated data thus collected is usually highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. However, some data collected from business systems are not always produced in a format which can be easily stored in relational databases, and consideration must be given to the method of, and resources needed, to incorporate this kind of unstructured data (if required).

Business systems data can be produced by public agencies and industry bodies, e.g.:

- Government departments and agencies (e.g. Fisheries NSW creates and maintains a range of significant spatial datasets)
- Grower groups, industry associations, and farm associations
- Citizen surveillance programs (e.g. FeralScan, Pantry Blitz)

And by businesses, e.g.:

- Farming operations
- Agronomy and consultancy services
- Value chain businesses (production, processing, marketing and distribution, and services)

Australian farm businesses currently report the same or similar data in multiple forms to multiple agencies (or even to different parts of the same agency). The National Farmers Federation (2013) suggested that governments must streamline their processes, ensuring that consistency is achieved across both departments and jurisdictions, to redress this duplication.

Internet of Things (machine-generated data)

The Internet of Things (IoT) refers to the network of sensor- and software-enabled devices, vehicles, and machinery that can connect, interact and exchange data. The phenomenal growth in the number of sensors and machines used to measure and record the events and situations in the physical world is creating vast 'data lakes' of information of variable quality and relevance, dependent on the capacity of the sensors to perform as designed and the need for the information beyond the designed purpose.

While the well-structured nature of IoT data makes it suitable for computer processing, as sensors proliferate the volume and velocity of data produced are challenging traditional processing approaches.

Examples of this kind of data include:

- Data from sensors, i.e.
    - Fixed sensors
        - Weather sensors and dataloggers (probes)
        - Scientific sensors
        - Video and imagery
    - Mobile sensors (tracking)
        - Mobile phones
        - Cars/Machinery (all forms of transportation/travel/trade data)
        - Satellite images
- Data from computer systems, i.e.
    - farm management software
    - hardware and production packages
    - simulation modelling

4.3 Examples of alternative data sources

As previously noted, an exhaustive examination of potential data sources is made difficult by the rate of emergence of new data sets. This section presents a small cross-section of potential sources as examples of data available to Australian OSAs that could augment existing statistical collections.

Source	Type	Characteristics	Benefits	Concerns
Agronomic products and resources	*Business systems (process-mediated data) / Internet of Things (machine-generated data)*	Digital agriculture applications and decision support tools such as those described as part of the P2D project report (Perrett et al., 2017) can provide a wealth of structured and semi-structured data on production and value chain inputs, outputs and processes, for example: • Agworld • Agro • APSIM and derivative products (e.g. Whopper Cropper, Yield Prophet) • Back Paddock • CSBP Sampling Pro • CliMate • Connected Farm Field • Farm Records • ProductionWise • FarmGRAZE • iPaddockYield • Stringy Bark Software A more comprehensive list of agronomic decision tools is described in Appendix 3.	Several digital agriculture applications allow communication between service providers – such as agronomists and accountants – and the farmer to improve timeliness and accuracy of data and data analysis. These platforms combine many aspects of farming into one program which can increase robustness and accuracy of farm data and also save time, compared to paper-based record-keeping and data entry.	The large number of discrete farm management software platforms on the market collect and disseminate data differently, meaning the data which can be collected will not always be comparable without significant disaggregation and restructuring. Data robustness is also dependent on the accuracy of user entering the information, which can impact aggregation of data. Privacy breaches are a potential concern.

AskBill	*Business systems (process-mediated data) / Internet of Things (machine-generated data)*	AskBill – named for founding Dean of the University of New England's faculty of Rural Science, Dr Gordon 'Bill' McClymont - is a web-based software tool developed by the Sheep CRC which provides timely and accurate predictions of sheep wellbeing and productivity using weather, stock and pasture information to sheep producers across Australia. The AskBill computer models build on that understanding to predict future events and analyse 'what if' scenarios. Producers enter farm and production data and monitor risk alerts.	The system draws on farm data and weather forecasts to predict production, well-being risks and opportunities for sheep on an individual property level. The combination of individual farm records and big data spot risks can help producers identify opportunities to improve livestock management and may aid in benchmarking. AskBill also allows producers and industry to validate the standard of care animals have received.	Data accuracy is based on the quality of information recorded by the user which may impact the reliability of aggregated data and trends. Weather forecast information may not be granular enough to identify specific risks.
Commercial data	*Business systems (process-mediated data)*	Technology software platforms are utilised by many commercial businesses in consolidating data to perform analysis to improve business performance and decision making, as well as complete day to day operational tasks such as payment to suppliers. Partners in Fast-Moving Consumer Goods (FMCG) supply chains hold large volumes of historic data which can be managed through ERP (enterprise resource planning) software programs such as SAP.	Commercial data can be aggregated and anonymised to provide industry benchmarks for growers to improve their decision-making process and increase farm productivity. The software platforms implemented by commercial businesses to manage their data can hold large volumes and can even perform some analysis as part of the package.	Businesses are usually unwilling to publicise their commercial data due to the increased risk of transparency to competitors in the marketplace who could potentially gain advantage from access to this data. Aggregation of data is often unable to solve this issue due to the small number of participants in some markets. Data accuracy and timeliness of recording also depends on the behaviours of the business which can impact the usefulness of the data to make decisions.

DairyBase	*Business systems (process-mediated data)*	DairyBase was developed by Dairy Australia, in collaboration with DairyNZ, to provide dairy farmers and advisors with a free, web-based tool to help manage their farms, and to provide industry with a national database of dairy farm performance information. It includes data from the Dairy Farm Monitor Project (DFMP) and other validated datasets from consultants and service providers. The DFMP provides a comprehensive physical and financial analysis for dairy farms across Australia. No private user data will be made available through the DairyBase website and benchmarks are based on DFMP data. All data entry, retrieval and reporting are conducted behind a secure login, and all private user or farm information is kept in a secure database controlled by Dairy Australia.	This online system enables farmers to measure their business performance over time and undertake comparative analysis for their business. The tool is free for dairy farmers to use which is likely to increase participation rates and includes the ability of entering historic figures to perform year on year analysis. Reports generated from this data are used by industry and government to inform policy and service delivery to generate economic growth. DFMP data can be filtered based on a range of factors, such as farm size, region, feed input level and other farm system drivers allowing a ‘like for like’ comparison.	While the data presents results and trends, these need to be interpreted carefully as participant farms may not be representative of the industry and not all farms participate every year. Data accuracy is based on the quality of information recorded by the user which may impact the reliability of aggregated data and trends.

DEXA technology	*Business systems (process-mediated data)*	Dual Energy X-ray Absorptiometry (DEXA) is an enhanced form of x-ray technology being implemented as an objective carcase measurement (OCM) tool and is fast becoming the global standard for chemical lean (CL) measurement. It is significant to the meat and livestock industry as it offers greater precision and has the capacity to eliminate fat claims and lean giveaway. Unlike other methods of CL testing, DEXA technology is not limited to boneless, ground meat and is capable of being applied to all uncooked meat, whether fresh or frozen, bulk, blended or packaged in cartons. As ongoing R&D enhances the application of OCM around all conceivable measures, DEXA can provide timely, accurate, transparent and objective information on the lean meat, bone and fat composition of each carcase.	DEXA technology allows users to substantiate claims made about specifications marketed to consumers. Other potential benefits of the implementation of DEXA and the corresponding data to the supply chain include: • increased carcase value • better understanding of supply based on the carcase composition of individual herds • increased boning room efficiency (less labour required to trim fat, automation) • increased compliance to market specifications • more informed on-farm management decisions around breeding to optimise feed utilisation and turnoff times • transparency of carcase composition data	DEXA data cannot be traced back to a point of origin, meaning the feedback provided to producers would be reliant on the effectiveness of communication and data handling in the supply chain. DEXA technology only provides information on lean meat, bone and fat composition on individual carcasses. Analysis on effective farming practices and herd selection would need to take other factors into account and would be reliant on the record-keeping of husbandry practices for accurate comparisons. For example, two carcasses of the same breed and farm may not be comparable as they had different vaccination treatments. Aggregate data would need to consider geological differences, exposure to adverse conditions such as drought and variations in feeding regimes.

NLIS	*Business systems (process-mediated data)*	The National Livestock Identification System (NLIS) enables the lifetime traceability of Australian cattle, sheep and goats along the supply chain via: • an animal identifier (a visual or electronic ear tag known as a device) • identification of a physical location by means of a Property Identification Code (PIC) • a web-accessible database to store and correlate movement data and associated details All animals in the system must be identified with an NLIS accredited tag or device before leaving a property and each movement they make to a location with a different PIC is recorded centrally on the NLIS Database. Status may be assigned to individual tagged animals or to properties to record pertinent information, for example to note that an animal has been vaccinated against a particular disease, or that all of the animals on a property with a particular PIC may have been exposed to a contaminant.	NLIS has the ability for producers to record farming and animal husbandry practises on a individual animal basis and then correlate this data with processor feedback on carcase attributes, creating an enriched source of information for better decision-making. The system is endorsed by major producers, feedlots, agents, saleyards and processors and underpinned by State and Territory legislation.	The level at which producers use NLIS as a farm management record/keeping tool can vary, which can impact on the useability of aggregate data. Data accuracy and robustness levels are dependent on the user entering the information, which can impact aggregation of data. Privacy breaches and mosaic risk are also a potential concern.

Satellite imagery	*Business systems (process-mediated data) / Internet of Things (machine-generated data)*	Historical satellite imagery of global crop production zones provides a rich source of data which – when processed via data analytics and modelling – can be converted into invaluable information about agriculture. The U.S. Department of Agriculture (USDA) uses Landsat imagery for Global and Domestic Agricultural Monitoring in various ways, including: • Estimating crop production • Monitoring consumptive water use • Zone mapping • Developing the annual Cropland Data Layer • Foreign Agricultural Service—Global Agricultural Monitoring The adoption of Landsat imagery and other higher-precision technologies is increasing with private producers. For statistical and agricultural agencies, the value of satellite imagery as a data source has been well established, and future use is primarily dependent on the continuing availability of multi-spectral imagery.	Remote sensing technology, such as Landsat imagery, offers one of the most effective and efficient means of collecting timely and objective data on agricultural activities, e.g. the ability to monitor crop conditions during the growing season, such as vegetation stress from nutrient and (or) water deficiency and fungal or pest outbreaks.	The primary challenge is low global repeat frequency, which continues to result in data gaps for global agricultural monitoring. Occlusion of the land surface by clouds remains the main challenge. In high rainfall or tropical areas, generally only one cloud-free image is acquired each year. Usage of satellite imagery varies substantially. For example, use by field agronomists, where what is planted in a specific paddock is well known, differs from spatial identification of different crops across broad production regions, where what is planted is completely unknown. The second use of satellite imagery has greater value for statistical purposes but currently has low crop specificity. The utilisation of satellite technology for broad-scale agronomic management depends on many factors, including perceptions of costs and benefits by end users, sufficient education and expertise to apply the technology, and capability of the technology itself to provide timely and accurate information at sufficient detail to inform management decisions.

Technology supplier acquired farm data	*Business systems (process-mediated data) / Internet of Things (machine-generated data)*	Large technology suppliers such as John Deere and Bayer collect vast amounts of process and machine data via IoT and client accounts on the way their products are utilised, for example: **Production data** - *information about equipment and land use:* • field task details, area worked, route travelled • crop harvested, yield data • agronomic inputs applied, soil data **Equipment data** - *information that indicates machine health and function:* • machine settings and readings • machine hours / life and location • diagnostic codes • software and firmware versions • machine attachments, implements or headers **Administrative data** - *information about the client's account and activities:* • data sharing permissions • users, machines, devices, and licenses linked to the account • number of acres • size and number of files	Data from large technology suppliers may aid producers in understanding how they can maximise the efficiency of their equipment and whether they are utilising all functions. Production data held by suppliers can be a useful data storage system for producers who are not recording data themselves or act as a backup is data is missed or lost.	Monopolisation of data pools is a concern. The proprietary design of many of the digital systems means that the longer a farmer uses a particular machinery brand, the more it cost in terms of loss of data value to switch to another machinery brand. This can have the effect of locking farm businesses into one machinery brand, essentially creating a monopoly farm by farm. Additionally, Australian farmers are currently not adequately protected from their farm data being collected and used without their knowledge or consent. Guidelines are needed to encompass legal, social and ethical rules to protect the agricultural sector, the interests of farmers and farmer privacy.

Trimble navigation	*Internet of Things (machine-generated data)*	Though best known for GPS technology, Trimble integrates a wide range of positioning technologies including GPS, laser, optical and inertial technologies with application software, wireless communications, and services to allow customers to collect, manage and analyse complex information which is especially useful in the application of precision agriculture. Over the past 10 years, the company has secured a number of patents in data management on methods to estimate plant growth and develop corresponding field prescriptions.	Data from ground and satellite-based sensors can be combined to develop accurate plant growth predictions.	Even when the microdata have been gathered and summarised to a high standard, further analysis is usually needed before the compiled data can be shared with, or communicated to, different audiences, or used as the basis for policy-making. The continual cycle of moving sets of data through a process that creates information presents challenges for determining clear ownership of aggregated data and the information from which it derives.

Twitter	*Social information (human-sourced data)*	As an online news and social networking site where people communicate in short messages ('tweets'), Twitter is a rich source of timely, unstructured information. Twitter's raw data can be accessed via a developer portal using the Twitter Application Program Interface (API). A wide variety of functions are available within this API, including information-gathering functions such as searching for and downloading tweets and metadata for a specific topic or user; and interactive functions, such as following users or retweeting. Search capabilities are limited temporally to tweets occurring within the past week (approximately), and therefore longitudinal studies must also have a separate system for storing search results. Basic information that can be gathered from Twitter includes but is not limited to: user name and profile description (including mentions of other users); user location; tweet text (including hashtags, website links, images); number of replies to tweet, retweets by other users; and likes by other users.	Using this information in conjunction with the appropriate methodology for the question at hand has been demonstrated to assist in the quantification of spatiotemporal dynamics of multiple activities. For example, Zipper (2018) estimated crop planting progress in the US and (following verification of the results) hypothesised that this information could eventually be used to monitor planting progress in real-time. It has value in offering improved contextualisation of context existing, expert-collected data.	Key challenges of using Twitter as a data resource include: availability; data storage; semantically driven; geolocation capabilities; representativeness; limited number of tweets accessible for free (API program). There is currently no widely accepted methodology or support for using Twitter to replace traditional sources.

VinSites	*Business systems (process-mediated data)*	Wine Australia's VinSites pilot project aimed to create a consolidated national dataset for the wine sector providing critical data to track Australian wine from the vineyard through to consumption. Western Australia was chosen as the pilot state for the insights system. Each vineyard in WA was mapped using satellite technology. Maps previously developed by DAFWA (that are freely available to producers) were used to verify the satellite imagery. Vineyard owners were invited to 'claim' their vineyard and entered their own specific varietal, harvest, rootstock, block and water information. This information was also linked to production data to provide accurate and timely information about yields by variety and by region. Wine Australia is now assessing the learnings from both the technology implementation and use of the system to determine the next steps.	Currently the wine sector has access to many disparate data repositories, all with slightly different content or out-of-date information; VinSites was designed as a single source of truth for all wine producers to access. VinSites synthesises farm data, public data (DAFWA maps) and commercial (production) data to enrich information sets.	Voluntary provision of data can result in skewed collection as those most likely to participate are usually the more successful and innovative producers.

WoolQ	*Business systems (process-mediated data)*	The WoolQ portal is a digital hub for the Australian wool industry that delivers information, data, selling choices and trading opportunities to woolgrowers, their agents and other industry professionals, in a single online destination. The project is owned and managed by Australian Wool Innovation. WoolQ allows growers to: - capture, share and store clip details digitally - monitor and analyse the clip, test results and current estimated value - use data for informed decision-making - identify opportunities to maximise grower returns - more easily engage directly with the wool industry.	WoolQ allows wool industry partners data to better inform decision-making processes through market transparency. The software provides a platform for industry networking and communication between growers and brokers/buyers. WoolQ is free for wool industry participants and can be used off-line and then synced once internet is available once more. These factors may encourage program participation and hence increase representativeness of the data.	Stakeholders may have concerns over privacy of data, which AWI addresses through privacy settings in the technology platform and aggregation / anonymisation of information and data provided by WoolQ users for the benefit of the industry.

5. Assessment framework

Frameworks to assess and report on the quality of statistical data and to assist those developing data collections to produce quality outputs (such as the ABS Data Quality Framework and the FAO Statistics Quality Assurance Framework) have already been developed and adopted by OSAs domestically as well as globally. These frameworks must account for numerous dimensions which affect statistical quality – for example, the characteristics of the data, the institutional environment in which the data are produced, and the practices and processes used.

Following assessment of different examples, a **high-level framework** (Figure 4) and **evaluation scorecard** (Table 2) for assessing Australian agricultural data sources has been developed for this report, modelled on the framework for best practice of big data architecture created by Auburn University for data management in health care systems (Wang et al, 2015) and the Generic Statistical Business Process Model (GSBPM) shown in Figure 5.

The framework presented here can be used to assess data sources for their suitability within a revised national agricultural statistics system via the scoring system based on the potential for actualisation and efficacy to enhance decision-making processes. It includes the **data characteristics** of the '5 Vs' Big Data model – i.e. velocity, volume, value, variety, and veracity (Ishwarappa & Anuradha, 2015) – and **principal requirements of governance and production** which must be met before potential alternative sources of data can be incorporated into official agricultural statistics collections. These include:

- availability
- integrity
- consistency
- accessibility
- harmonisation (national and global)
- respondent burden
- fitness for use
- user needs
- timeliness
- relevance
- legal and regulatory issues.

These requirements have been selected primarily from the High-Level Group for the Modernisation of Official Statistics, the Eurostat *Strategy for Agricultural Statistics for 2020 and Beyond* (European Commission, 2015), the *Precision Agriculture to Decision Agriculture* report (Perrett et al., 2017) and the *National Agricultural Statistics Review* (Hodges & Lehmann, 2017). Additional questions informing the development of these tools is presented in Appendix 4. Users of the framework also need to consider the environment in which our OSAs operate, i.e. modernisation and efficiency programs currently underway, when assessing data sources for suitability.

5.1 Data assessment framework

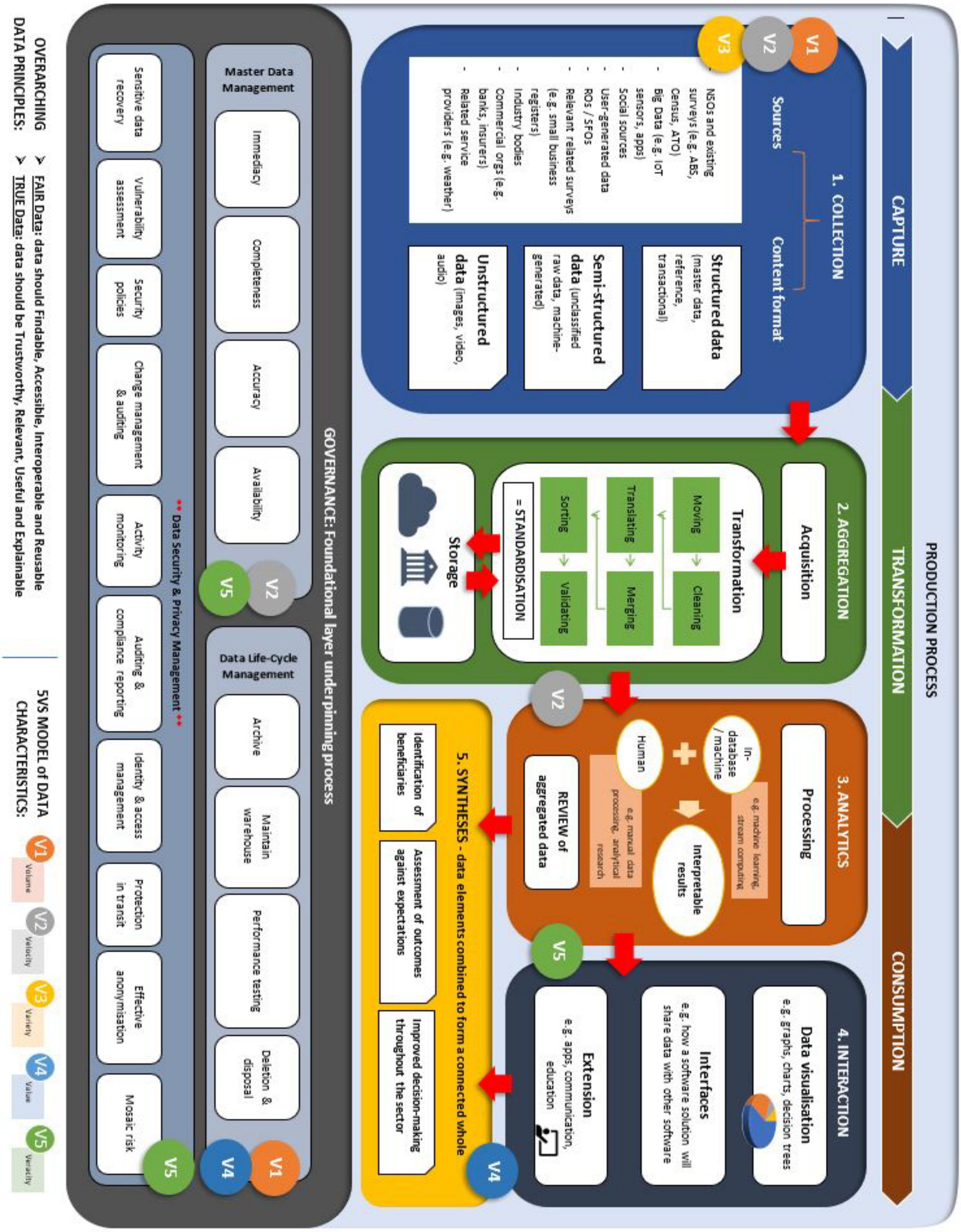


Figure 4. Framework for assessment of data suitability for inclusion in Australian agricultural statistics.

Using the framework

When assessing data within this framework, it is important to consider:

- **Where the data enter / exit the framework**
  - *Do the data have a path to implementation?*
- **What pieces of the framework are required to make the data fit for use**
  - *Is the effort to obtain and analyse the data greater than the benefit gained?*
- **Where the resources to enable the data to progress through the framework are concentrated or missing**
  - *Will the data become lost in the system before it can be made valuable?*
- **Cost of utilising the data**
  - *Are existing systems and processes sufficient, and if not does the benefit of acquisition and integration outweigh the cost?*
- **The overarching open data principles**
  - *Is the data findable, accessible, interoperable and reusable (FAIR)?*
- **The place of the data within the underpinning foundational layer**
  - *Have all governance issues been accounted for and addressed?*

While the FAIR principles for open data provide guidance for data sharing, management and stewardship (particularly within the digital ecosystem), we would add another overarching common-sense principle to assess fitness for use: is it **TRUE**? That is, the data sourced should be **trustworthy, relevant, useful and explainable**.

Potential data sources can be evaluated using a set of questions related to the framework presented here in Table 2. For this process, a series of questions (Appendix 4) from both production (i.e. fitness for use) and governance (i.e. management) perspectives were tabulated into a scorecard. Users of the scorecard assign a rating of good, fair or poor to specific characteristics under these criteria, and a score is calculated which evaluates the source's overall suitability for inclusion in official statistics *in its current state.* It should be noted that a 'poor' assessment does not rule out a source for inclusion; rather that consideration is required to identify if appropriate solutions to the source's weakness can be developed.

These evaluations are separated into production and governance perspectives, as data characteristics for each use can differ dramatically – that is, a source may be ideal for use in official statistics in terms of architectural integration (production) but fall short of required standards of verifiable accuracy (governance). In turn, the evaluation can be further broken down into subsets of requirements to enable the user to identify weaknesses and strengths of a particular source, for example a source which rates poorly overall for production integration may be strong in the 5 Vs but require allocation of management responsibilities to enable efficient use.

While not exhaustive, these questions provide a general (albeit subjective) view of a data source's suitability for inclusion in the national agricultural statistics system.

Table 2: Data source evaluation scorecard

GOVERNANCE CRITERIA	GOOD	FAIR	POOR
1. How well does the source meet the requirements of **master data management for:**			
a. Immediacy			
b. Completeness			
c. Accuracy			
d. Availability			
e. Relevance			
Master data mgt (Score x/5)			
2. How well does the source meet the requirements of **data lifecycle management for:**			
a. Archive availability			
b. Warehouse maintenance			
c. Performance testing			
d. Responsibility of governance review			
e. Deletion & disposal			
Data lifecycle management (Score x/5)			
3. How well does the source meet **security and privacy requirements for:**			
a. Sensitive data recovery & vulnerability assessment			
b. Activity monitoring & protection in transit			
c. Change management, auditing & compliance reporting			
d. Mosaic risk (anonymisation)			
e. Security policies in place			
Security & privacy (Score x/5)			
Data source governance requirements (Score x/15)			
PRODUCTION CRITERIA	**GOOD**	**FAIR**	**POOR**
Data **structure** rating (structured, semi-structured, unstructured)			
Data can be **readily acquired**			
Data can be **readily processed**			
Data can be **transformed & standardised**			
Data can be **integrated** with other systems/sources			
Applicability for use (Score x/5)			
Responsibility for **acquisition** is clear			
Responsibility for **processing** is clear			
Responsibility for **extension** is clear			
Responsibility / applicability for **synthesis** is clear			
Responsibility for **review of functionality** is clear			
Responsibility for use (Score x/5)			
Data characteristics as per the 5Vs model - Does the source have …			
V1: Volume			
V2: Velocity			
V3: Variety			
V4: Value			
V5: Veracity			
The 5Vs scorecard (Score x/5)			
Data source production requirements (Score x/15)			

In addition, any evaluation of data sources must consider more specific data quality management frameworks such as the GSBPM (Figure 5) employed by the UNECE and revised by ABS as the Statistical Production Activity Model (SPAM).

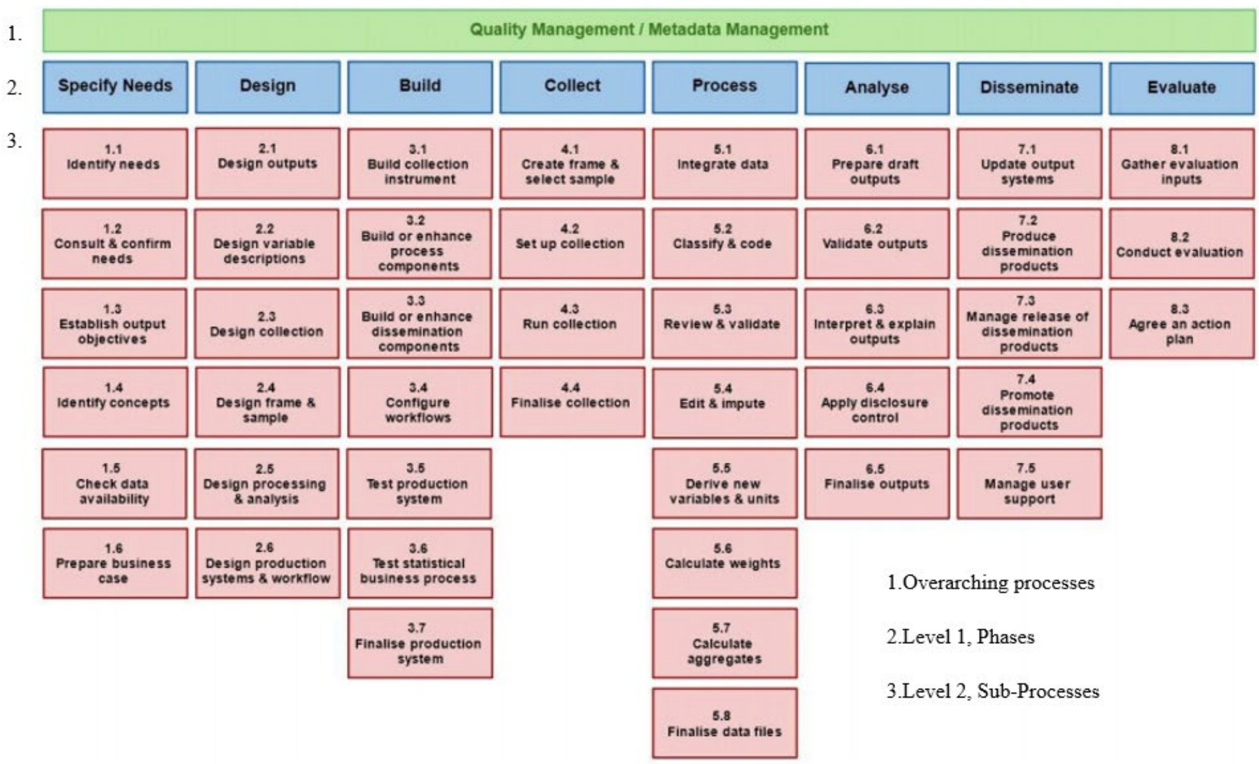


Figure 5. Generic Statistical Business Process Model (GSBPM) - Source: (Muñoz, 2017)

5.2 Evaluation

The scorecard depicted in Table 2 should be used to broadly evaluate potential sources, provided that the evaluation is carried out with a specific question in mind.

As an example, the question was posed: is Twitter a suitable data source to provide additional insight into national crop planting intentions? When assessed using this framework (Table 3), Twitter offers velocity (speed and timeliness) but the data cannot be easily verified or standardised. Overall its suitability for use as a source which fits production and governance requirements *for this purpose* is thus rated as poor to fair.

In contrast, the same question applied to levy payer records results in assessment of mostly fit for use (from a production perspective) and as a good source regarding governance requirements, particularly given recent amendments to legislation that enable sharing of data collected through the levies system with the ABS (but not ABARES).

As with all assessments, subjective analysis is required to decide whether a rating of ‘fair’ or ‘poor’ instead of ‘good’ across certain criteria is acceptable for the intended purpose, and these scores will likely change in response to the purpose in question as well as the user’s own bias.

The scoring system presented here should not be used to definitively assess a source as good or bad, but to identify areas worthy of further investigation. If a source is rated 'poor' in production but 'good' in governance, what steps can the evaluator identify or take to improve the production issues in order to utilise the source in official agricultural statistics?

The scoring system presented in Table 2 is binary – i.e. if the evaluator decides the question should be answered with 'fair' then a 1 is placed in the fair column and 0 or blank in the other columns, providing a total rating out of 5 for subsections and 15 for each criteria (governance and production) as shown in the example presented in Table 3. Totals from this example are displayed as percentages in bar columns (Figures 6-9) to provide a visual score of the source as a good, fair or poor fit *for the purpose in question*. The aggregated views (Figure 7 and Figure 9) provide an overall assessment of source fitness, while the three subset scores for both governance (Figure 6) and production (Figure 8) are also displayed separately in the example as these subsets may have different weightings for an evaluator (e.g. privacy may be a higher concern than lifecycle management).

Example of evaluation method

Question A:

Is Twitter a suitable data source to provide additional insight into national crop planting intentions?

Table 3: Example of evaluation scorecard based on Question A.

GOVERNANCE CRITERIA	GOOD	FAIR	POOR
1. How well does the source meet the requirements of **master data management for:**			
a. Immediacy	1	0	0
b. Completeness	0	0	1
c. Accuracy	0	0	1
d. Availability	1	0	0
e. Relevance	0	1	0
Master data mgt (Score x/5)	2	1	2
2. How well does the source meet the requirements of **data lifecycle management for:**			
a. Archive availability	0	1	0
b. Warehouse maintenance	0	0	1
c. Performance testing	0	0	1
d. Responsibility of governance review	0	0	1
e. Deletion & disposal	0	0	1
Data lifecycle management (Score x/5)	0	1	4
3. How well does the source meet **security and privacy requirements for:**			
a. Sensitive data recovery & vulnerability assessment	0	0	1
b. Activity monitoring & protection in transit	0	1	0
c. Change management, auditing & compliance reporting	0	1	0
d. Mosaic risk (anonymisation)	0	0	1
e. Security policies in place	1	0	0
Security & privacy (Score x/5)	1	2	2
Data source governance requirements (Score x/15)	3	4	8

PRODUCTION CRITERIA	**GOOD**	**FAIR**	**POOR**
Data **structure** rating (structured, semi-structured, unstructured)	0	1	0
Data can be **readily acquired**	1	0	0
Data can be **readily processed**	0	0	1
Data can be **transformed & standardised**	0	0	1
Data can be **integrated** with other systems/sources	0	0	1
Applicability for use (Score x/5)	1	1	3
Responsibility for **acquisition** is clear	0	0	1
Responsibility for **processing** is clear	0	0	1
Responsibility for **extension** is clear	0	1	0
Responsibility / applicability for **synthesis** is clear	0	1	0
Responsibility for **review of functionality** is clear	0	0	1
Responsibility for use (Score x/5)	**0**	**2**	**3**
Data characteristics as per the 5Vs model - Does the source have …			
V1: Volume	0	1	0
V2: Velocity	1	0	0
V3: Variety	0	1	0
V4: Value	0	1	0
V5: Veracity	0	0	1
The 5Vs scorecard (Score x/5)	**1**	**3**	**1**
Data source production requirements (Score x/15)	**2**	**6**	**7**

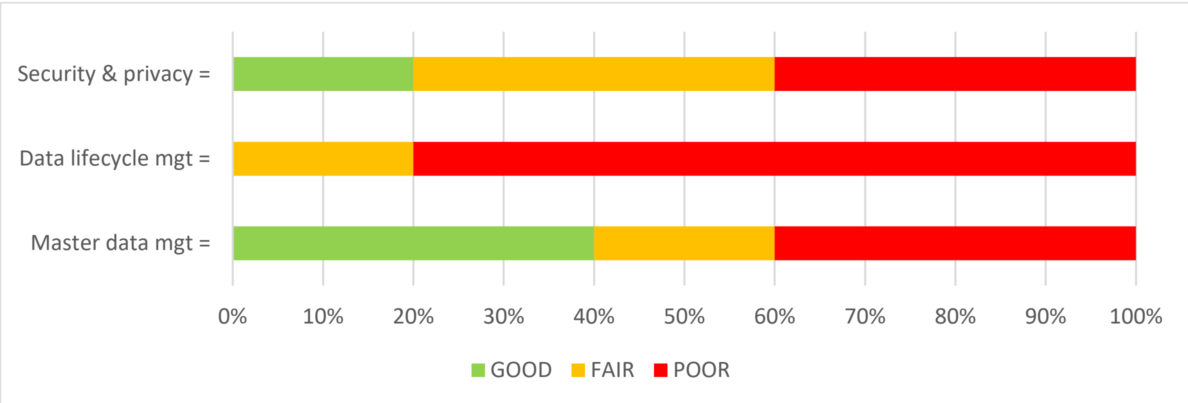


Figure 6. Twitter's suitability with governance subsets for Question A

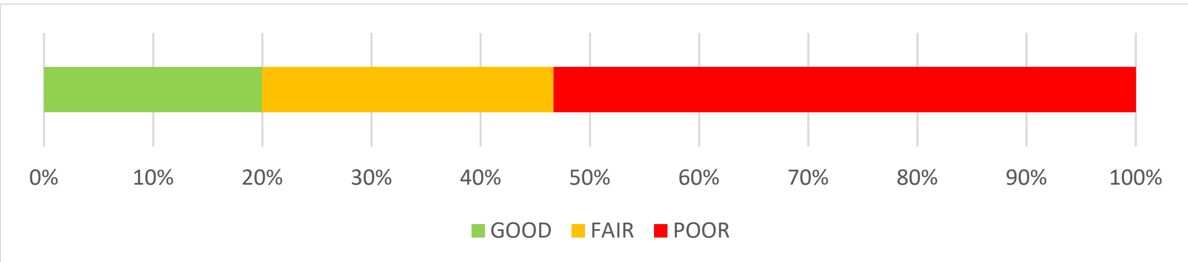


Figure 7. Twitter's suitability with governance requirements overall for Question A

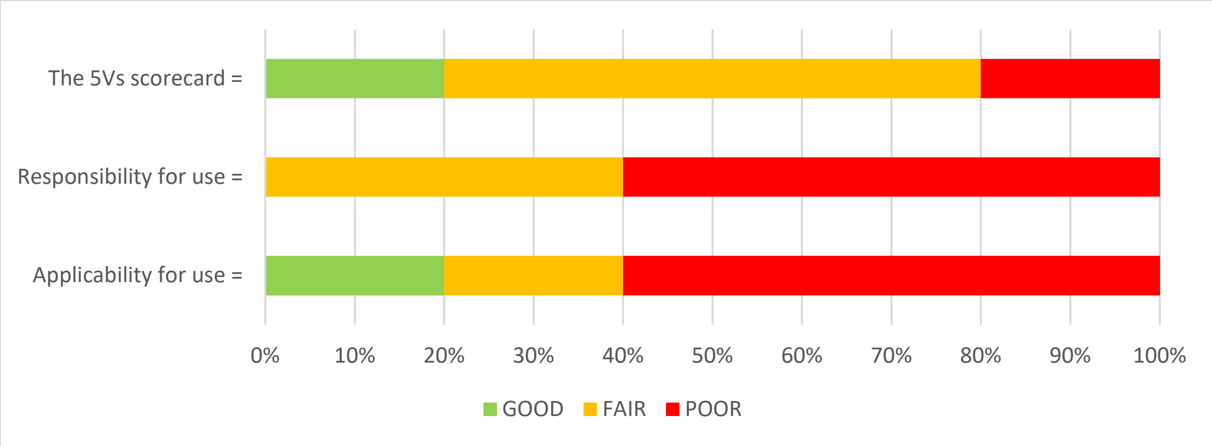


Figure 8. Twitter's suitability with production subsets for Question A

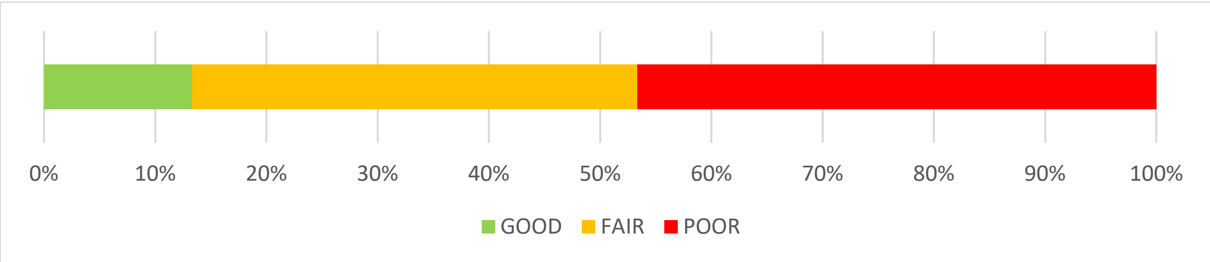


Figure 9. Twitter's suitability with production requirements overall for Question A

Evaluation examples

Using the scoring system presented here, a range of available sources were evaluated for suitability on a common industry-wide purpose (*Question B*) and for a sector-specific purpose (*Question C*), with aggregated scores in governance and production criteria displayed as bar charts for comparison (Figures 10-13):

1. Twitter
2. DairyBase
3. Levy payer records
4. NLIS
5. VinSites
6. Satellite imagery[6]

[6] e.g. historical satellite imagery of global crop production zones such as that held by the USDA

Question B:
Does this source offer viable general information which could enhance the national agricultural statistics system in Australia?

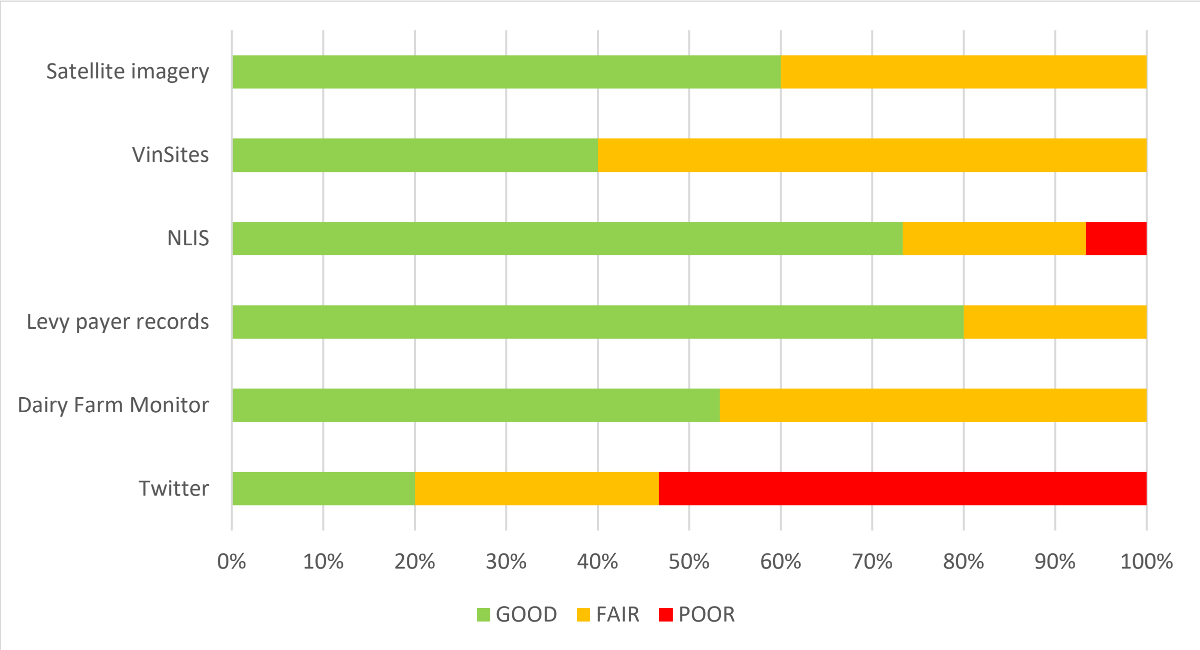


Figure 10. How well the source addresses governance criteria overall for Question B

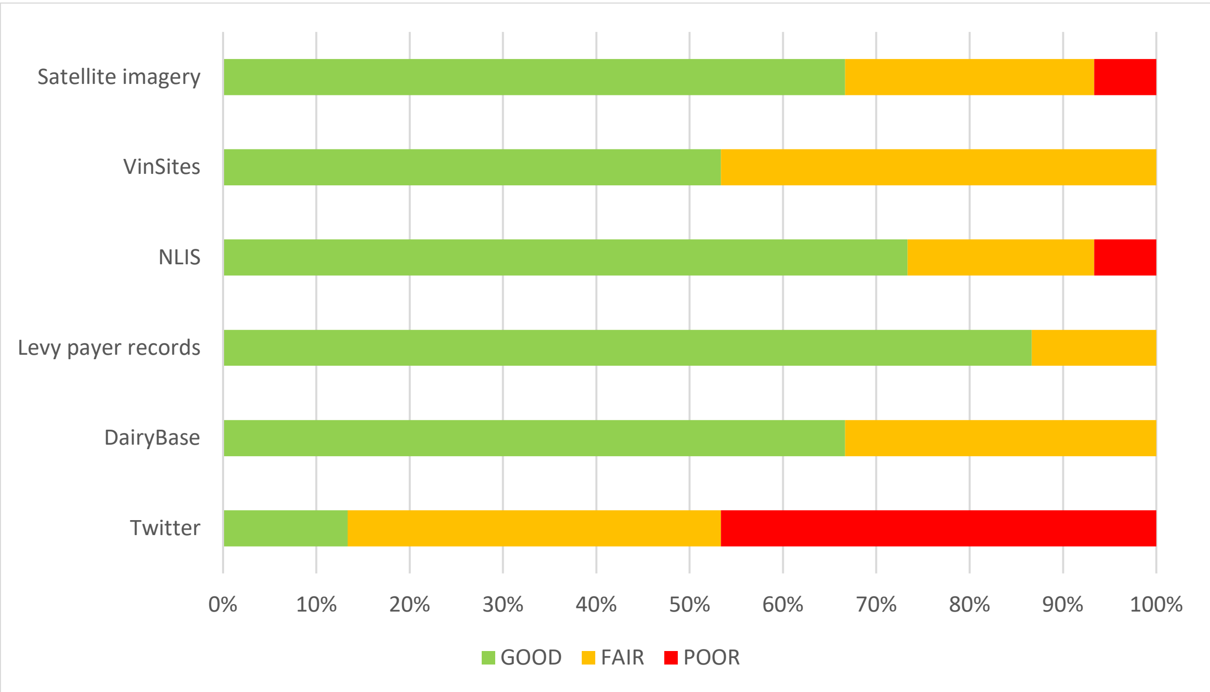


Figure 11. How well the source addresses production criteria overall for Question B

Question C:
Can this source help determine trends in dairy producers' decision-making?[7]

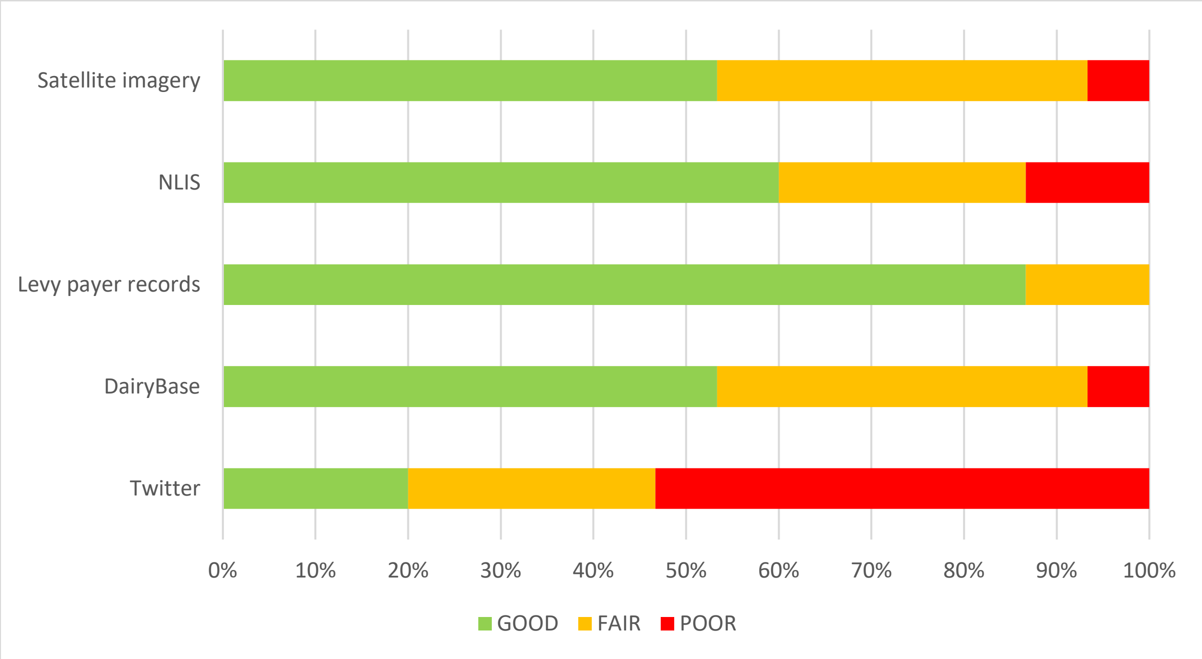


Figure 12. How well the source addresses governance criteria overall for Question C

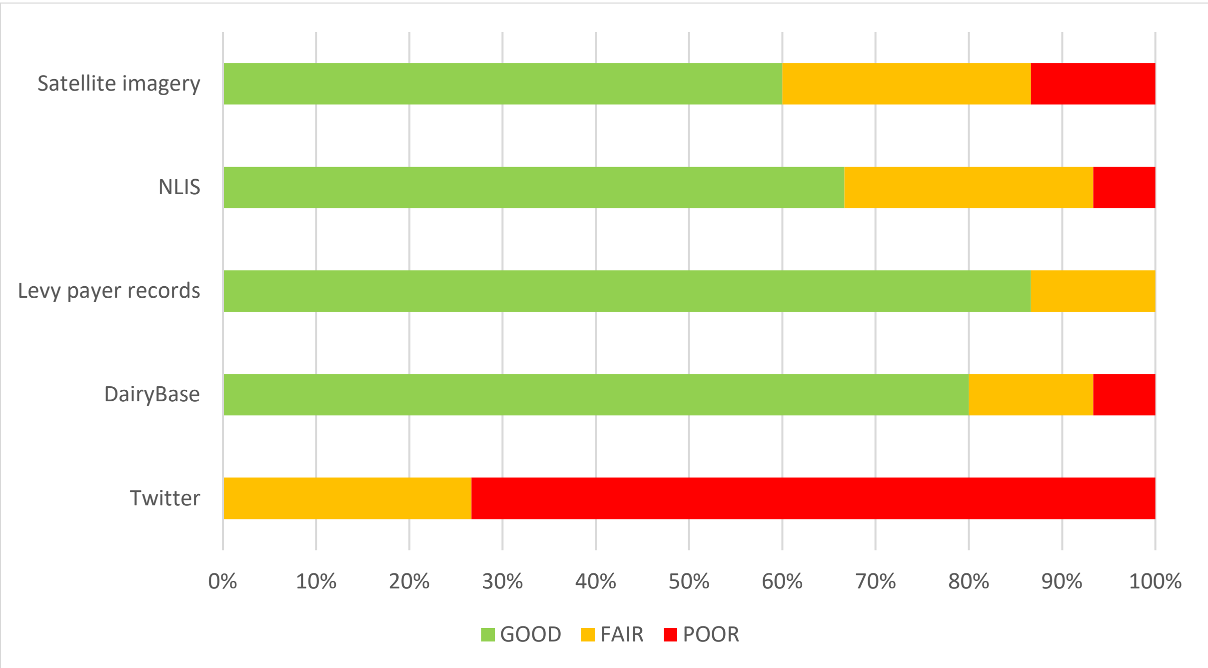


Figure 13. How well the source addresses production criteria overall for Question B

[7] (NB – as VinSites is evidently not a relevant source for this question, it was excluded from the results.)

As many of the values of the sources are fixed, such as whether the data are available or accurate, only slight differences in scoring or rating the sources are observed.

The greatest differences are noted in the areas of relevance and privacy, as mosaic risk becomes a greater factor when data are used in a more specific context. This demonstrates the importance of the intent of the evaluator and the need to give due consideration to subjectivity in the evaluation process.

In addition, this process considers generic data source characteristics without weighting, which should also be taken into account by the evaluator. For example, had 'relevance' been weighted in the examples presented then sources assessed for the third question (regarding suitability for determining trends in dairy producers' decision-making) would have returned a greater percentage of 'poor' scores.

This evaluation system should be used to assess those aspects of data source characteristics which require further investigation or remedial action in order to utilise the source for official purposes. Questions that users of this system need to consider include: What percentage of 'fair' scoring is acceptable for a data source to be considered for integration into official statistics? If a source returns a 'poor' rating, what (if anything) should be done to improve the specific production or governance issues?

To reiterate, the process is intended not to categorically define a source as good or bad, but to provide a framework within which a dataset can be considered according to a set of criteria that enable a user to evaluate the quality of a potential data source and how it is managed.

6. Conclusion and recommendations

6.1 Discussion

It is clear both in Australia and internationally that the need to adapt or transform agricultural statistical systems is urgent, however frustration with the rate of progress in Australia is high. Discussions with industry representatives conducted for this report confirmed the issues noted herein are common across most Australian agricultural sub-sectors and representative organisations.

Opacity of industry reporting for the sub-sectors was a frequently cited concern, as were the difference between industry and OSA classifications of production units and timeliness of data provision. While some interviewees noted that geographic detail was lacking in current statistics, others pointed to the increased risk of mosaic effect[8] and potential subsequent privacy breaches attached to a greater level of geographic granularity and geo-spatial enhancement; i.e. once a dataset is linked, vulnerability exists despite best-practice privacy governance. In addition, once a data set is 'de-identified' it is sometimes also devalued.

Concerns were also raised that the trend away from official statistics towards commercial data production was likely to improve the quality of specific information for some commodities but could also limit the scope and restrict accessibility of information. Ownership of data was an issue discussed by all interviewees, as was the strong need to clearly identify the value proposition associated with data collection. While individuals and businesses could see a need to collect data for decision-making, the broader value to industry is considered somewhat ethereal. It was agreed that education and incentives across industry could improve efforts to provide and share data for the greater good, i.e. to improve forecasting and decision-making and to maintain or strengthen the social licence to operate.

While the 2015-16 Agricultural Census underwent a series of reforms based on government, industry and community consultation to better address the needs of key users, there is strong sentiment amongst industry stakeholders interviewed for this report to further reform the system before the next Census, i.e. urgently.

Additionally, recent research conducted by the Australian Farm Institute (AFI) - particularly into the impacts on the sector of rising energy costs, risk management options and climate change - has uncovered significant gaps in agricultural statistics. These data gaps are impeding the development of evidence-based policy, which the sector needs more than ever in order to maintain its social licence, mitigate an increasing range of risks and ensure a sustainable future.

[8] The mosaic effect refers to reassembling unconnected data like a mosaic puzzle in unforeseen or unplanned ways; that is, as more data become discoverable and machine readable, disparate data threads can be pieced together to yield information that should be secure or private.

Social licence

The AFI report *Australian agriculture: an increasingly risky business* (Laurie et al, 2019) noted that new and emerging institutional risk factors will play a significant role in the future viability of the sector.

Institutional risks associated with competition and supply chain dynamics may be accelerating but are nonetheless familiar to most agricultural supply chain actors. New and emerging institutional factors (e.g. community trust on issues such as animal welfare, potential glyphosate regulation, land use and genetic engineering technologies) have created the greatest uncertainty and increased risk for the entire agricultural value chain in recent times.

As these risks are the product of an increasingly active and engaged consumer base, they are unlikely to diminish. Reliable data is necessary to establish trust via transparency of business practices and decision-making – and while provision of data alone is not sufficient to allay these emerging concerns, there is no doubt that a lack of data exacerbates mistrust and thus enhances the institutional risk.

However, the collection and use of data is a process which is in turn dependent on social licence – i.e. data providers must trust that their data will be used as agreed and accept that sharing of the data will create enough value to make the process worthwhile. For this trust and acceptance to translate to social licence, guidelines on data use and the benefits of sharing must be completely clear to all participants in the process.

As noted by Zhang et al. (2017) 56% of producers surveyed for the P2D project had little or no trust in service/technology providers maintaining their data privacy and not sharing their data with third parties.

Without a clear value proposition on the provision of new, non-official data by agricultural stakeholders at every stage of the value chain, gaps in statistical information will remain. If these gaps are not addressed, the industry risks losing its social licence to operate in many key aspects of production.

Mitigating risk

The research discussed above also specified that improved data availability would have a positive effect on provision of risk management options for the sector. Lack of relevant, accurate and comprehensive data was a consistent theme that arose throughout the compilation of this report, impacting the ability to adequately assess all sectors.

Data limitations are also impeding the ability of the market to develop and provide cost-competitive risk management products. Lack of data was noted as a key constraint on the development of a mature farm insurance sector; that is, the Australian insurance industry lacks detailed climatic information to accurately assess the exposure of and estimated duration for an insured event (Deloitte, 2017).

While technology offers access to new data which will help the development of new risk products, a critical mass of data collected over time is needed before these sources can be useful for this purpose, which may take several years.

Improved access to and interoperability of data would benefit the whole agricultural value chain to better manage risk. The agriculture sector should also learn from other industries/sectors which have improved data collection, particularly where this relates to risk Government and OSAs (i.e. ABS and ABARES) have a role to play in enabling the supply of appropriate data to enable better risk assessment, particularly in the provision of more granular weather data and utilisation of IoT and satellite information.

Sustainable development

Current AFI research into the need for a national strategy for climate-smart agriculture has also identified many data gaps which hinder the development of solutions to pressing issues.

The combined pressures of climate change and population growth pose a real and present threat to the sustainability of the Australian agricultural industry. However, detailed and consistent data on the existing and potential impacts of these pressures is difficult to come by. While agriculture is one of the economic sectors most vulnerable to adverse climate change impacts (Adams, Hurd, & Reilly, 1999; Anwar et al., 2015), the effects will vary across sub-sectors. Agriculture is also one of the drivers of climate change, but information on emission reduction projects is largely siloed within subsectors or unavailable. Granular, uniform data is needed to better assess the priorities for systems change.

The sustainability of the industry is also threatened by rising energy prices. The AFI report on *The impacts of energy costs on the Australian agriculture sector* (Heath, Darragh, & Laurie, 2018) found that as Australian agricultural businesses intensify production systems and utilise additional digital technology, their exposure to energy cost risk is increasing.

Australian energy costs have increased significantly over the past decade, which has important implications for the future competitiveness of Australian agriculture. The sector needs to engage in the debates associated with this issue to optimise outcomes. However, data on energy use in agriculture is inconsistent and piecemeal, which will limit the ability of the sector to engage meaningfully with policy-makers unless this is rectified.

The authors conducted a comprehensive review of literature in an effort to find the best available official sources of quantitative data on energy use in agriculture. They discovered a large variation in data sources in terms of units and scale of reporting, level of segregation of the data, extent of the supply chain the data covered and sampling framework, and many data gaps, particularly in the post-farm and processing components of the agricultural value chain.

The report concluded that lack of data is the biggest barrier to understanding the impact of energy policy changes on Australian agriculture. This research highlighted the need for improvement in the collection and management of energy statistics for the sector, particularly given that the industry needs to urgently construct a compelling case for energy policy change in order to develop sustainably.

6.2 Conclusion

Historically, statistical data have been entirely contained within official statistical agencies (OSAs). The information ecosystem is now changing at a velocity which necessitates an urgent response from industry and OSAs. Cooperative and agile efforts are needed to improve the collection and dissemination of agricultural statistics; a transformative approach is required, not a gradual adaptation.

A vast amount of alternative data is increasingly available to enhance official datasets, but many sources are not immediately suitable for incorporation into the national agricultural statistics system. The need for structure in the data and care in the transformation process is paramount.

Recommendation: Industry and OSAs cooperatively agree on a **common framework** (such as the example presented herein) for the evaluation and implementation of alternative or additional data sources to augment the Australian agricultural statistics system.

Recommendation: An education program be established by ABS for relevant personnel in industry bodies which currently collect data to ensure a consistent and detailed understanding of the considerations for inclusion of non-official data in the national agricultural statistics system.

Many of the recommendations of the P2D report (Perrett et al., 2017) which apply specifically to data use and collection remain relevant to the findings of this report, namely:

- the development of a Data Management Policy for Australian digital agriculture
- the development of a voluntary Data Management Code of Practice and accreditation scheme
- the development of Big Data Reference Architecture and Data Management Implementation Plan
- the establishment, review and refining of foundational data sets, and
- the digitisation and automation of data collection, including for regulatory compliance activities.

Recommendation: An Agricultural Data Taskforce (as recommended in the P2D report) recognises the cross-industry need for sharing learnings on systems transformation and coordinates with other relevant industry bodies to ensure a harmonised approach to collection of national data.

Recommendation: Efforts to transform the national Australian agricultural statistics system must **align with global best practice** on data management / improvement (e.g. UNECE Big Data project, Eurostat Strategy for Agricultural Statistics for 2020 and Beyond) and that Australian agriculture is appropriately represented at these forums.

While alternative data sources can help bridge statistical gaps for agriculture, consistent uptake throughout the sector is required to create a viably diverse data pool. In this regard, trust is still a barrier to data supply. Robust guiding strategy is required to build the trust necessary to facilitate broad, representational data collection.

Shared data could dramatically enhance agricultural statistics and thus improve decision-making at all levels from farm gate to policy tables. To this purpose, a clear value proposition on the provision of data by agricultural stakeholders as an industry good must be explicitly articulated and implemented, not only by OSAs but also by industry. For example, owners of data would be incentivised to contribute to a common data pool to be used for official statistics purposes if they in turn could then access the synthesised data for their own business purposes.

Recommendation: An education program be established by industry for all stakeholders in the agricultural value chain regarding the industry good benefits of sharing data.

The evaluation framework presented in this report demonstrates that the kinds of data sources most appropriate for inclusion in the Australian agricultural statistics system are business / process-mediated and machine-generated data, with a structure, volume, integrity and relevance related to each source. Socially-generated data should not be discounted as a potential source but, at this stage, integration is still problematic, and the value of this source is less obvious.

Recommendation: A national register of potential data sources which could be used to supplement agricultural statistics be established, maintained and regularly revised by the ABS, with support and contribution from industry.

Within agriculture, some sectors are more advanced than others in pursuit of better data and statistical collection, thus cooperative efforts are needed to ensure the industry improves collectively and not in a piecemeal way. In addition, agriculture should learn from other industries/sectors which have improved data collection (such as health and mining). Improved access to and interoperability of stored data would benefit the whole agricultural value chain, particularly through dissemination of datasets with cross-industry value which are used in other sectors.

The gaps within agricultural subsector datasets are notable and of concern. While the extension of official statistics to include alternative data sources will address many of these gaps, further research should be conducted to specifically identify the gaps in existing statistical sets in order to prioritise improvement efforts.

Recommendation: A **review of sources of data queries** (i.e. on stakeholder needs and collection processes) be published by ABS to help identify where statistical gaps exist and where data-gathering activities can be streamlined and shared.

To expedite the reinvention of Australian agricultural statistics, a culture shift must occur in the agriculture sector as well as across industries, whereby the role of OSAs in the new information landscape is acknowledged and supported. Only then can Australia's national agricultural statistics system be enhanced and enriched in order to properly support improved research, policy, programs and business decisions.

Recommendation: Industry recognises and supports ABS in its role as the official statistical *collector* as well as being the *coordinator* in a **multi-stakeholder data-gathering process.**

Recommendation: Industry supports the **redistribution of responsibilities defined for the data lifecycle** by ABS; namely the delegation of data collection, processing and distribution from non-official sources as required to augment official statistics.

Recommendation: A **prompt review of current methodologies and classifications** used by ABS and ABARES be undertaken to ensure efficacious use of existing data and to enable ease of integration for new data sources.

Recommendation: Industry supports the goal of OSAs to ensure that data sourced for use in the national Australian agricultural statistics system is both **FAIR** (findable, accessible, interoperable, and reusable) and **TRUE** (trustworthy, relevant, useful and explainable).

7. References

ABS. (2018, August). Australian Bureau of Statistics - Statistical Spatial Framework. Retrieved from http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Spatial+Framework

Adams, R., Hurd, B., & Reilly, J. (1999). *Agriculture and Global Climate Change a Review of Impacts to US Agricultural Resources*. Washington, D.C.: Pew Center on Global Climate Change.

Anwar, M. R., Liu, D. L., Farquharson, R., Macadam, I., Abadi, A., Finlayson, J., … Ramilan, T. (2015). Climate change impacts on phenology and yields of five broadacre crops at four climatologically distinct locations in Australia. *Agricultural Systems*, *132*, 133–144. https://doi.org/10.1016/j.agsy.2014.09.010

Australian Bureau of Statistics. (2015a). *National Agricultural Statistics Review, Final Report*. Retrieved from http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CCBD44885C436852CA257E900017285B/$File/national%20agricultural%20statistics%20review%20-%20final,%202015.pdf

Australian Bureau of Statistics. (2015b). National Agricultural Statistics Review (NASR).

Berg, E., Cecere, W., & Ghosh, M. (2014). Small Area Estimation for County-Level Farmland Cash Rental Rates. *Journal of Survey Statistics and Methodology*, *2*(1), 1–37. https://doi.org/10.1093/jssam/smt041

Braaksma, B., & Zeelenberg, K. (2015). Re-make/Re-model: Should big data change the modelling paradigm in official statistics? *Statistical Journal of the IAOS*, *31*, 193–202.

Commonwealth of Australia. (2014). *Australian Public Service - Better Practice Guide for Big Data* (p. 35).

Darnell, R., Moore, A., Barry, S., Bramley, R., Grundy, M., & George, A. (2018). *The Current and Future State of Australian Agricultural Data*. *15*(1), 9.

DAWR, & ABS. (2017, November). *Roadmap to improve the agricultural statistics system*. Retrieved from http://www.agriculture.gov.au/SiteCollectionDocuments/abares/ag-stats-roadmap.pdf

Deloitte. (2017, November). Multi-peril crop insurance in Australia: barriers and opportunities | Deloitte Australia | Consumer & Industrial Products. Retrieved September 11, 2018, from Deloitte Australia website: https://www2.deloitte.com/au/en/pages/consumer-industrial-products/articles/multi-peril-crop-insurance-australia-barriers-opportunities.html

Eggers, W. (2013, July 24). Data as the new currency | Deloitte Insights. Retrieved April 26, 2019, from Deloitte Insights website: https://www2.deloitte.com/insights/us/en/deloitte-review/issue-13/data-as-the-new-currency.html

European Commission. (2015). *Strategy for agricultural statistics for 2020 and beyond*. Retrieved from European Commission website: http://ec.europa.eu/eurostat/documents/749240/749310/Strategy+on+agricultural+statistics_Final/fed9adb7-00b6-45c5-bf2c-2d7dcf5a6dd9

Eurostat. (2013). *Scheveningen Memorandum on Big Data and Official Statistics*. Retrieved from European Commission website: https://ec.europa.eu/eurostat/cros/system/files/SCHEVENINGEN_MEMORANDUM%20Final%20version.pdf

FAO (Ed.). (2010). *A system of integrated agricultural censuses and surveys*. Rome: Food and Agriculture Organization of the United Nations.

Florescu, D., Karlberg, M., Reis, F., Del Castillo, P. R., Skaliotis, M., & Wirthmann, A. (2014). Will 'big data' transform official statistics? *Eurostat*.

Heath, R., Darragh, L., & Laurie, A. (2018). *The impacts of energy costs on the Australian agriculture sector*. Retrieved from Australian Farm Institute & Energy Consumers Australia website: http://nla.gov.au/nla.obj-747651278

Hodges, J., & Lehmann, N. (2017). The National Agricultural Statistics Review: Collaborative partnerships supporting the Australian agricultural statistical system. *ICAS VII 2016: Seventh International Conference on Agriculture Statistics Proceedings*. https://doi.org/10.1481/icasVII.2016.h47

Ishwarappa, & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science*, *48*, 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Landefeld, S. (2014). Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues. *International Conference on Big Data for Official Statistics, Beijing, China*.

Laurie, A., Darragh, L., Curtis, M., Heath, R., & McRobert, K. (2019). *Australian agriculture: an increasingly risky business*. Retrieved from Australian Farm Institute website: http://nla.gov.au/nla.obj-1022968574

Leonard, E., Rainbow, R., Trindall, J., Baker, I., Barry, S., Darragh, L., … Zhang, A. (2017). *Accelerating precision agriculture to decision agriculture: Enabling digital agriculture in Australia.* Cotton Research and Development Corporation.

Little, R. J. (2015). Calibrated Bayes, an inferential paradigm for official statistics in the era of big data. *Statistical Journal of the IAOS*, *31*, 555–563.

McCarthy, J. S., Beckler, D. G., & Qualey, S. M. (2006). An Analysis of the Relationship Between Survey Burden and Nonresponse: If We Bother Them More, Are They Less Cooperative? *Journal of Official Statistics*, *22*(1).

MLA. (2016). *Digital Value Chain Strategy Forum Workshop Summary and Report*.

Muñoz, J. (2017, June). *Generic Statistical Business Process Model (GSBPM)*. National Academy of Sciences.

National Academies of Sciences, Engineering, and Medicine, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, & Panel on Methods for Integrating Multiple Data Sources to Improve Crop Estimates. (2017). *Improving Crop Estimates by Integrating Multiple Data Sources* (M. E. Bock & N. J. Kirkendall, Eds.). https://doi.org/10.17226/24892

National Farmers Federation. (2013). *National Farmers Federation Issues Paper Red Tape in Australian Agriculture*. Retrieved from https://www.nff.org.au/get/4228.pdf

NESSI. (2012). *Big Data: A New World of Opportunities* (p. 25) [White Paper]. Retrieved from Networked European Software and Services Initiative website: http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf

Perrett, E., Heath, R., Laurie, A., & Darragh, L. (2017). *P2D Economic impact of digital agriculture* (p. 57). Australian Farm Institute.

Potard, G., & Keogh, M. (2013). *Is counting farmers harder than counting sheep?* Australian Farm Institute.

Productivity Commission. (2017). *Data Availability and Use Inquiry Report*. Retrieved from Australian Government website: https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf

Scannapieco, M., Virgillito, A., & Zardetto, D. (2012). Placing Big Data in Official Statistics: A Big Challenge? *Italian National Institute of Statistics*.

SRA. (2017). Sugar Research Australia Strategic Plan 2017/18 - 2021/22. Retrieved December 10, 2018, from https://sugarresearch.com.au/wp-content/uploads/2017/09/Strategic-Plan-2017-_FINAL.pdf

Statistics Canada. (2012). *Agriculture statistics program review*. Retrieved from http://www.statcan.gc.ca/sites/default/files/agri-report-eng.pdf

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Tam, S.-M., & Clarke, F. (2015a). Big data, official statistics and some initiatives by the Australian Bureau of Statistics. *International Statistical Review*, *83*(3), 436–448.

Tam, S.-M., & Clarke, F. (2015b). *Big Data, Statistical Inference and Official Statistics*. Retrieved from Australian Bureau of Statistics website: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/015937BADB90186BCA257E0B000E428A/$File/1351055054_mar%202015.pdf

Task Force on the Value of Official Statistics. (2017). *Value of official statistics: Recommendations on promoting, measuring and communicating the value of official statistics*. Retrieved from Economic Commission for Europe website: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2017/CES_4-Value_of_Official_Statistics_for_endorsement_for_upload.pdf

UN Statistical Commission. (2014). *Big data and the modernisation of statistical systems*. https://doi.org/10.18356/2db69c02-en

United Nations Economic and Social Council. (2013). *Big data and modernization of statistical systems*. Retrieved from https://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf

Wang, Y., Kung, L., Ting, C., & Byrd, T. A. (2015). Beyond a Technical Perspective: Understanding Big Data Capabilities in Health Care. *2015 48th Hawaii International Conference on System Sciences*, 3044–3053. https://doi.org/10.1109/HICSS.2015.368

Williamson, J., & Nicholls, A. (n.d.). *Challenges to Agriculture Statistics in an Environment of Government Restraint*. 7.

World Bank. (2010). *Global Strategy To Improve Agricultural And Rural Statistics*. https://doi.org/10.1596/978-1-4648-0484-7

Zhang, A., Baker, I., Jakku, E., & Llewellyn, R. (2017). *The needs and drivers for the present and future of digitial agriculture in Australia. A cross industries producer survey for the rural*

R&D for Profit "Precision to Decision" (P2D) project [Technical Report]. Retrieved from https://www.crdc.com.au/sites/default/files/P2D%20producer%20survey%20-%20CSIRO%20Final%20Report.pdf

Zipper, S. C. (2018). Agricultural Research Using Social Media Data. *Agronomy Journal*, *110*(1), 349. https://doi.org/10.2134/agronj2017.08.0495

Appendices

Appendix 1 – ABS and ABARES statistics

Current ABARES statistical publications

SERIES

Agricultural Commodities (including Australian Commodities)

- Agricultural Commodity Statistics (including Australian Commodity Statistics)
- Australian Crop Report
- Australian Dairy
- Australian Energy Statistics
- Australian Fisheries and Aquaculture Statistics
- Australian Forest and Wood Product Statistics
- Australian Mineral Statistics
- Beef
- Electricity Generation - Major Development Projects
- Energy in Australia
- Farm Survey Results
- Fisheries Surveys
- Fishery Status Report
- Forests (State of the Forests)
- Grains
- Lamb
- Minerals & Energy - Major Development Projects
- Plantation Inventory
- Research Reports
- Science and Economic Insights
- Technical Reports
- Wheat Supply Exports Monthly

TOPICS

- Agricultural
- Biosecurity
- Biotechnology and GMOs
- Climate
- Climate Change
- Commodities
- Conferences
- Energy
- Environment and Natural Resource Management
- Farm Performance and Farm Surveys
- Fisheries and Aquaculture
- Food
- Forests
- Indigenous Australia
- Invasive Species
- Land Use
- Minerals
- Models, Risk, Spatial Data and Datasets
- Productivity
- Social Issues
- Trade and Markets
- Water and Irrigation

Current ABS agricultural statistics portfolio

TOPICS AND MEASURE

LAND USE	
• Area	Land for conservation/protection Land not used for agricultural production including houses etc. Land for crops Land for grazing – improved pastures Land for grazing – other Land under forestry – native Land under forestry – plantation Land for other agricultural purposes Land sold/leased – including for continued agricultural use, urban development, mining and other or unknown uses
LIVESTOCK	
• Livestock numbers on holding • Eggs (production)	Dairy cattle – including cows in milk and dry, and various age splits Meat cattle – including various age splits Sheep and lambs – including various age splits, lambs marked and ewes mated Pigs Goats Poultry Egg production All other livestock
HAY AND SILAGE	
• Area • Production	Lucerne cut for hay Other pasture cut for hay Cereal cut for hay Other crops cut for hay Pasture (including lucerne), cereal and other crops cut for silage
FRUIT	
• Trees/area bearing • Trees/area not yet bearing • Production	Grapefruits Lemons Limes Mandarins Oranges All other citrus fruit Apples Pears (including Nashi) All other pome fruit Peaches Nectarines Cherries Plums All other stone fruit Mangoes Olives Avocados All other orchard fruit Grapes for wine production Grapes for all other uses Strawberries Blueberries Other berries Kiwi fruit Bananas

	Pineapples All other fruit
NUTS • Trees bearing • Trees not yet bearing • Production	Almonds Macadamias All other nuts
TOTAL ORCHARD AREA Area	Orchard fruit trees and nut trees
VEGETABLES • Area • Production	Carrots Mushrooms Onions (brown, white, red) Potatoes Tomatoes Lettuces Melons (including rock, bitter and water) Capsicums (excluding chillies) Beans (including french and runner) Broccoli Cauliflowers Sweet corn Peas Pumpkins Cabbages Brussel sprouts All other vegetables for human consumption
CEREALS AND OTHER CROPS • Area • Production	Wheat Oats Barley Triticale Sorghum Rice Maize All other cereals for grain Canola Other oilseeds (including safflower, soybeans, sunflowers, sesame, linseed) Peanuts in shell Lentils Lupins for grain or seed Chickpeas Mung beans Faba beans (including tick, horse and broad) Other pulses Sugar cane – cut for crushing Sugar cane – plant or other, not for crushing Cotton lint – irrigated Cotton lint – non-irrigated All other crops (excluding fruit, nuts and vegetables for human consumption)

ORGANICS	Area of holding certified as organic, bio-dynamic or in-conversion Percentage of fruit and/or nuts certified organic – area and production Percentage of vegetables for human consumption certified organic – area and production
WATER	Sources – total water volume used from various sources Purchase of extra water on temporary basis – total cost and total volume Purchase of extra water on permanent basis – total cost and total volume Annual irrigation water volumetric or usage charges – total cost Use of water – area watered and volume of water used for various purposes
NURSERIES • Area	Nurseries (undercover and outdoor) Cut flowers (undercover and outdoor) Cultivated turf (outdoor)
FERTILISER	Fertilisers containing a nitrification inhibitor – area applied to, tonnes applied, litres applied Slow release fertilisers – area applied to, tonnes applied Nitrogen based and other fertilisers – area applied to, tonnes applied, litres applied, no. of applications Cropping period in which nitrogen based fertilisers were applied – area applied to, tonnes applied, litres applied Crop and pasture types that nitrogen based fertilisers were applied to – area applied to, tonnes applied, litres applied Application methods used for nitrogen based fertilisers – area applied to
CROP AND PASTURE MANAGEMENT • Area	Cultivations – including zero, one, two and three or more cultivations for pasture and cropping Crop stubble and/or trash management – including six crops and eight management practices
DEMOGRAPHICS	Relationship to business Age Gender Length of time involved in farming
FARM MANAGEMENT	Sources of income – percentage breakdown

Appendix 2 – Agriculture Industry Publications

PUBLICATION	DESCRIPTION	FREQUENCY	FUNDING MECHANISM
ABARES NATIONAL WOOD PROCESSING SURVEY	Calculating the production of hard wood and softwood products.	2012-13	Partly funded by FWPA levies
ABS LAND MANAGEMENT PRACTICES SURVEY	The Land Management Practices Survey (LaMPS) is designed to support and inform the Carbon Farming Initiative (CFI) under the Emissions Reduction Fund (ERF) and the National Greenhouse Gas Inventory by providing data in respect to the implementation and uptake of on-farm emission reduction activities.	Every 2 years	User-funded
AGRIBUSINESS BANKING COMMODITIES ANALYSIS	Examples: NAB Rural Commodities Wrap CommBank Agri Insights research Rural Bank Insights Rabobank agribusiness outlook ANZ agribusiness commodity reports Westpac Produce magazine	Usually monthly	Banks (funded by commercial organisations)
AITHER WATER MARKETS REPORT	Provides an overview of water market activity in major water trading zones in the southern Murray-Darling Basin, compares results with previous water years and provide comment on the outlook for future water years.	Annual	Aither
ANIMAL HUSBANDRY SURVEY	To monitor uptake of animal welfare initiatives, Dairy Australia regularly commissions independent surveys of farmers.	2016	NSW Dairy, Dairy Australia

AUSTRALIAN HORTICULTURE STATISTICS HANDBOOK	Comprehensive information covering all sectors of the Australian horticulture industry in an easy-to-read guide; drawn from the ABS, international trade sources, domestic market pricing reporting services.	Annual	Horticulture Innovation Australia
AUSTRALIAN PIG ANNUAL	Pig meat marketing and consumption. Uses APL sources and others such as ABS and ABARES, as well as BPEX and FAO.	Annual	APL
AUSTRALIAN SUGAR MILLING COUNCIL (ASMC) INDUSTRY STATISTICS	Statistics on sugar cane milling, e.g. Original Forecast, Current Forecast, Weekly Crush, To-Date Crush	Weekly	ASMC
AUSTRALIAN WOOL STATISTICS YEARBOOK	The Australian Wool Exchange (AWEX) produces an annual summary of the previous season's wool trading at the beginning of each season: -Annual Statistics for Wool Production -Detailed Analysis of Auction data -Area of Production Analysis -Statistics of Australian Wool Exports by Class and Destination	Quarterly	AWEX
AWI MLA WOOL AND LAMB FORECASTING SURVEYS	MLA and AWI run a joint survey to collect sheepmeat industry livestock numbers and lamb production expectations.	3 x per year	MLA, AWI
BALANCING DAIRY PRODUCTION AND PROFITS IN NORTHERN AUSTRALIA - QUEENSLAND DAIRY ACCOUNTING SCHEME	Improves the understanding of business principles among advisors and dairy farmers by providing farm management accounting and analysis.	2017	QDAS and co-funded by DAF, Dairy Australia
CATTLE MARKET ANALYSIS	Market data analysis and comments.	Weekly	Mecardo

COTTON AUSTRALIA ANNUAL REPORT	Collates data for key areas affecting cotton production in Australia e.g. crop size, forecasts and yields.	Annual	Cotton Australia
DAIRY FARM MONITOR PROJECT	Data collected through the Dairy Farm Monitor Project is **now stored in DairyBase** and provides the high-quality data that allows for comparison of farms, which can be used to identify areas for improvement.	Annual	Dairy Australia
DAIRYBASE	Web-based database that enables dairy farmers to: -Compare their own farm business over time -Create annual reports and forecasts -Identify opportunities to drive profit and reduce risk -Make more informed business decisions -Generate comparative analysis according to farm size, region and production system	Ongoing	Dairy Australia
EGG FARMING - AUSTRALIA MARKET RESEARCH REPORT	IBISWorld identifies key success factors for the Egg Farming Industry: Use of specialist equipment or facilities; Proximity to key markets; Proximity to key suppliers. IBISWorld analysts also discuss how external factors such as Egg production and Demand from supermarkets and grocery stores in the Egg Farming industry impact industry performance.	Published May 2018	IBISWorld
FARM PRACTICES SURVEY	Outlines the adoption of key management practices.	Annual	GRDC
FWPA AUSTRALIAN TIMBER INDUSTRY INVESTMENT REVIEW	A voluntary survey of selected production facilities was conducted to ascertain the total level of investment over the past five years.	2017	FWPA
GRAINS MARKET ANALYSIS	Market data analysis and comments.	Weekly	Mecardo
GRDC GROWER SURVEY	Performance feedback from 1,200 grain growers across all agro-ecological zoning.	Every 2 years	GRDC

IN FOCUS	Comprehensive and credible collection of Australian dairy industry statistics. Uses ABS, ADC and DA sources; ABARES, State Milk Authorities and Dairy Manufacturers.	Annual	Dairy Service Levy, Australian Government
KOALA SURVEY - UAV	Locating koalas in commercial eucalypt plantations using UAVs and ground-based platforms.	5 trials in Oct 2017	FWPA
LIVELINK REPORT	Presents statistical trends on the live export industry.	Monthly	MLA
LOTFEEDING BRIEF	Survey to monitor trends in feedlot sector. Some figure provided my AUS-MEAT Limited.	Quarterly	MLA, ALFA
MERINOSELECT GENETIC TRENDS	National genetic information and benchmarking service for the Merino industry - estimation of an animal's true breeding value based on pedigree and performance recorded information.	Annual	Sheep Genetics, MLA
MLA – CO-PRODUCT REPORT	25 abattoirs across Australia participate, collating average prices and trends for all co-products items.	Monthly	MLA
MLA MARKET INFORMATION STATISTICS DATABASE	Long-term data on domestic market and price information including: Production; Export volumes and values; Consumption; Livestock prices and indicators; Retail prices; Herd and flock numbers; Co-product prices; Feedlot inventory; Slaughter numbers	Weekly	MLA

NATIONAL PADDOCK SURVEY	Identifies the major constraints to production in different regions and soil types. Data collected throughout the cropping cycle in two production zones on 250 paddocks. GRDC joint project led by the Birchip Cropping Group in association with CSIRO, farming systems groups and agronomic consultants.	Cropping cycle (pre-sowing to harvest)	GRDC, BCG & CSIRO
NATIONAL RESIDUE SURVEY (NRS)	Facilitate the testing of animal and plant products for pesticide and veterinary medicine residues, and environmental contaminants. Provides an estimate of the occurrence of residues in products (using systems based on sampling and statistical probability).	Annual	Specific industry funded (eg. AWI funds national wool residue sampling)
PESTICIDES IN PLANTATIONS	Extending an understanding of the use of chemical pesticides in the plantations forest industry.	2006	FWPA
PRICES & MARKETS E-NEWSLETTER	The weekly statistics summary encompasses saleyard and over-the-hook price indicators, feeder and live export quotes, as well as slaughter and throughput information across the cattle and sheep markets.	Weekly	MLA
SHEEP CRC SURVEYS	The Sheep CRC is an independent organisation that performs research and develops technologies to enhance the Australian sheep industry. It conducts research on wellbeing and productivity, the meat value chain and genetics.	5 x per year	CRC with grant funding from the Australian Government
SHEEP MEAT MARKET ANALYSIS	Market data analysis and comments.	Weekly	Mecardo

STATISTICS FOR THE AUSTRALIAN GRAINS INDUSTRY (SAGI)	The SAGI project was a GRDC program under the leadership of Prof Brian Cullis, University of Wollongong, which is currently growing through independent regional nodes and the Southern node is presented by the University of Adelaide, School of Agriculture Food and Wine.	Ongoing	GRDC and local universities
THE AUSTRALIAN EXPORT GRAINS INNOVATION CENTRE (AEGIC) ECONOMIC ANALYSIS AND REPORTS	Ongoing strategic and economic market analysis into the grains market.	Periodical / ad hoc	AEGIC
UNIVERSITY OF NEW ENGLAND BENCHMARKING AUSTRALIAN SHEEP PARASITE CONTROL (IPMS SURVEY)	Measures change in sheep parasite control practices and attributes, provides a new benchmark against which to measure change in parasite control practices and attitudes into the future.	2012-13	AWI
WOOL MARKET ANALYSIS	Market data analysis and comments.	Weekly	Mecardo

Appendix 3 – Review of Decision Agriculture Products

The following provides a list of decision agriculture products available in Australia as noted in *The economic impact of digital agriculture* (Perrett et al., 2017). Although the list is comprehensive it is far from exhaustive, with new products emerging almost on a daily basis.

Crop Management: Multi-Use

- Hay vs. Grain calculator
- MyCrop
- ARM (agricultural risk management) tools:
- -Climate ARM
- -FallowARM
- -CropARM
- -Nitrogen ARM
- NVT Yield and Disease App
- CropPro
- LiveFarmer
- Farm at Hand
- Conservis
- AgGuide software
- Yield Prophet
- eFarmer
- Agridigital
- ProductionWise
- iCropTrak
- Climate Fieldview
- SAP Farm Management (Vistex)
- FarmFlo
- APUNGA
- Satamap
- Cabbige
- Yield Gap Australia
- Grownotes Alert
- Cropscan 3000H system
- Sky Squirrel

Livestock Management:

- AgBoost
- MINDA
  - MindaLive
  - MindaApp
  - MindaMilk
  - MindaLand7Feed
  - MindaWeights
  - Minda Pro
- Cattleworks
- Industry Inventories
- Maia Grazing
- CloudHerd
- Pasture io/MilkFlo io
- Herdmaster
- ILR2/ILR Online
- FarmStock/ Financial Manager
- KoolCollect, KoolPerform, KoolStock
- Drought feed calculator
- Stocktake Plus
- Automed
- MLA Market Info app
- Aglive Pro (App/Web)
- Sense T Pasture predictor
- OPPIS (Online property and planning information software)
- FarmGraze
- Stockplan
- BreedCow and Dynamo Software
- GrassGro
- GrazFeed
- MetaFarms
- Moocall

Sector Specific:

Cotton

- CottASSIST
- FastStart Tools

Sugar

- Qcane Select
- Irrigweb

Wine

- Vinsight (winery software)

Industry specific:

Alpaca

- Optimate AlpaFarm

Goats

- EasyKeeper

Beef

- Farmecco
- Beefspecs Calculator

Dairy

- EasyDairy/ Easy ID
- Rumen8
- Countdown Mastitis Toolkit (Dairy Australia)
- Silent Herdsman
- Pro Grass Rotation/ Pro Milk Solids/Pro Dairy Event
- BOVControl

Soil management:

- SoilMate
- Nulogic Sampling Pro (App)
- SoilMapp
- ApSoil

Water/Irrigation management:

- Product
- Crop irrigation requirement calculator
- SoilWater App
- Idroplan
- Farm Connect Software

Pest/Disease management:

- Weedsmart App
- Weed Seed Wizard
- Weed ute guide
- Plantix
- Pest Track
- FeralScan
- PestGenie Ultimate
- Paraboss
    - Wormboss
    - Flyboss (Sheep CRC)
    - Liceboss (AWI)

Climate:

- RAINMAN Streamflow
- Sensing+ for Agriculture/Aquaculture
- Fires near me

General Business/Finance/Operational:

- Agrimaster
- Phoenix Farm Management Software (AGDATA)
- AgTribe
- Agritrack
- Operations Centre/ Mobile Farm Manager
- PS Stockbook/Farmbook
- i-Agri
- 3D Farm Modelling
- Complete Farm Operations/ Farm Chemical Management
- PAM (Process, Analyse, Manage)
- Farmer Mobile and Farmer Plus
- Back Paddock Manager, Mobile, Adviser
- Farm Manager/Contractor App
- Observant Global
- Farmware App
- SenseAg
- Figured
- Smart Elements
- Adventive
- F-Track Live
- AgriWebb Notebook
- Tractor Tracker App (On the Go Farm)
- Agri360
- Farmbrite
- Cultura Farm software solutions
- Farmer Productivity Insight
- AGBIZ Farm Budgeting Tools
- WALI
- Agdraft
- Wi-Sky
- P2P Agri
- Farm Command
- Trimble
- Drone Deploy

Emerging products - Australia:

- Ceres Tag
- RapidAim
- EShephard
- Sense T Viti-App
- Spec-SINFER
- FluroSat

Appendix 4 – Production and governance questions

Questions and considerations guiding the development of the Framework and Evaluation Scorecard presented in Section 0:

1. Production criteria (fitness for use)

1. Define the **form** of the data collected, i.e.
    - Structured
    - Semi-structured
    - Unstructured
2. Clarify the **acquisition** process, i.e.
    - Who / what will collect the data?
    - Who is responsible for its collation and transport?
3. How will it be **transformed?** i.e.
    - Moved, cleaned, translated, merged, sorted, validated
    - Is a linking key (e.g. ABN, location, PIC) available?
4. Can it be **standardised**?
5. Where will it be **stored?**
6. How will it be **processed**?
    - Database / Human / Other
    - Who is responsible for processing?
    - Who is responsible for reviewing?
7. Can it be **extended** and how?
    - Visualisation / Digital interface / Personal
    - Who is responsible for extension?
    - Who is responsible for evaluating the extension?
8. Can it be **synthesised** to become part of whole-of-agriculture picture?

2. Governance criteria (management)

1. Does the source meet the requirements of **master data management**?
    - Immediacy
    - Completeness
    - Accuracy
    - Availability
2. Does the source meet the requirements of **data lifecycle management**?
    - Where will it be archived?
    - Who will maintain the data warehouse?
    - Who is responsible for performance testing?
    - What are the methods of deletion & disposal?
3. Does the source meet **security and privacy requirements**? i.e.
    - Sensitive data recovery
    - Vulnerability assessment
    - Security policies

- Change management & auditing
- Activity monitoring
- Auditing & compliance reporting
- Identity & access management
- Protection in transit
- Effective anonymisation
- Mosaic risk

4. **The 5Vs (data attributes)** – does the source have characteristics of:
    - V1: Volume
    - V2: Velocity
    - V3: Variety
    - V4: Value
    - V5: Veracity
5. **Is it TRUE and FAIR?**