NEW ZEALAND GAP

Future Focus

Technology Report

July 2017

JACOBS

Contents

1. Executive summary

This paper identifies and discusses some of the data management challenges, opportunities and risks related to assurance in the horticultural sector. There is an increasing awareness that if producer led initiatives do not define horticultural data management standards relatively quickly, then major commercial organisations, like Google, will develop proprietary datasets and processes. This would be a lost opportunity to establish data management based on open and global standards that will benefit everyone in the supply system, including growers.

New Zealand GAP is aware that we are at the beginning of a technology revolution, and is anticipating and planning for a phase of rapid change is this area.

This includes:

- Analysing and reporting vast amounts of complex data in a simple and meaningful way
- Customers and certification bodies viewing and auditing records from their office (perhaps overseas)
- Consumers having access to more information about food, e.g. its safety, environmental, or human rights credentials

There are eight areas that influence horticulture data management for New Zealand GAP including:

1. Grower data
2. Harvesting data - The Internet of Things
3. Smart Tools
4. Guiding Data Management Standards
5. The Operational Setting – Contextual Data
6. The Market
7. People and Capability
8. Risk Management

The main challenges and opportunities in these areas are: Big Data, Data Quality, Master Data Management, Data as an asset, The Supply Chain of Data, Data Silos, Dataset Version, Grower Data Lockers and Disruptive Technologies.

Much of the data collected and maintained by any organisation is not actively used. The concept of using "big data" and "data lakes" allows data from multiple sources to be stored centrally allowing it to be analysed for reporting and sharing with relevant stakeholders. The benefits of processing the underlying data include separation of the raw data from the analytical dataset that supports high value business decisions. This also allows better access control and reduced risk of unauthorised access to sensitive raw data.

This report recommends investment of resources and effort to:

1. Develop a technology and data strategy
2. Lead the development of global, open sector wide data management standards
3. Treat data as an asset
4. Actively manage data, discard what is not needed and providing methods to discover data
5. Establish a supply chain of data. Create information from the raw data to help improve understanding
6. Participate in leadership conversations at a national level on topics that include using datasets that fit-for-purpose and network infrastructure
7. Grow work force capability so there are people who have both technical skills and useful horticulture knowledge

There is need for leadership and collaboration within the horticulture sector, including taking an active part in the development of data management standards and to ensure information used is fit for purpose. There are opportunities for data-savvy growers to create joined-up datasets that provide a picture of their entire operation and for brave growers to share this information with others for mutual benefit.

There are also opportunities for New Zealand GAP to help develop a central system to influence how horticulture data is used, visualised and shared, or ensuring smart tools are intuitive to use and are appropriate for the audience.

The risk of doing nothing is that the opportunity to influence, guide and lead the development of standardised datasets and information products is lost. There is also a risk that private technology businesses take control of farm data. By taking a lead, New Zealand GAP can realise the goals of providing high quality and reliable data and reports to those who need it.

2. Introduction

2.1 Vision

New Zealand GAP has a vision where horticultural data is better managed and integrated so that accurate information is available to those who need it and where and when they need it. This paper identifies and discusses some of the data management challenges, opportunities and risks related to assurance in the horticultural sector. This report discusses data and information flows in horticulture and provides a global perspective on new initiatives and trends in data and information science.

2.2 Background

There are increasing demands for better information about horticultural products and services. In New Zealand and globally, there is an increasingly market-led interest in where food comes from and how it is produced. Requests for grower level information from retailers and regulators is becoming more frequent and complex. New Zealand GAP is not alone in facing data management challenges.

The NZ horticulture sector has a reputation as an innovative world leader. It is well connected globally with strong ties into organisations like GLOBALG.A.P. that are also working to improve access to product information. New technology results in more types of data being collected along with an overall growth in the size of datasets. Some new technology will be disruptive in how data is collected, managed and used. Other changes will drive the continued development and evolution of existing capability. There will also be sector lead changes from those who want technology to make their life easier e.g. automating parts of the audit process, reducing the time that auditors spend on farm.

There are opportunities to influence how horticultural data is collected, managed, analysed, shared, used and reused. There is also growing external influences. Regulators and Regional Councils are becoming more involved in assurance systems to monitor the usage of land and water resources.

There is an increasing awareness that if producer led initiatives do not define horticultural data management standards relatively quickly, then major commercial organisations, like Google, will develop proprietary datasets and processes. This would be a lost opportunity to establish data management based on open and global standards that will benefit everyone in the supply system, including growers.

Farmers and growers around the world are leading the way in the development of technology. The U.S. based Farmers Business Network (FBN) is a global, big data-power, farmer-to-farmer network — and they've just raised $USD 15 million to build their system.[1]

1 *http://techcrunch.com/2015/05/19/farmers-business-network-raises-15m-from-google-ventures/*

2.3 The Future Framework

There is no doubt that the world is about to undergo a technology revolution (the Fourth Industrial Revolution). It is likely that we are in the initial phase of this development and we are about to experience a seismic shift in the way we access and use technology over the coming years.

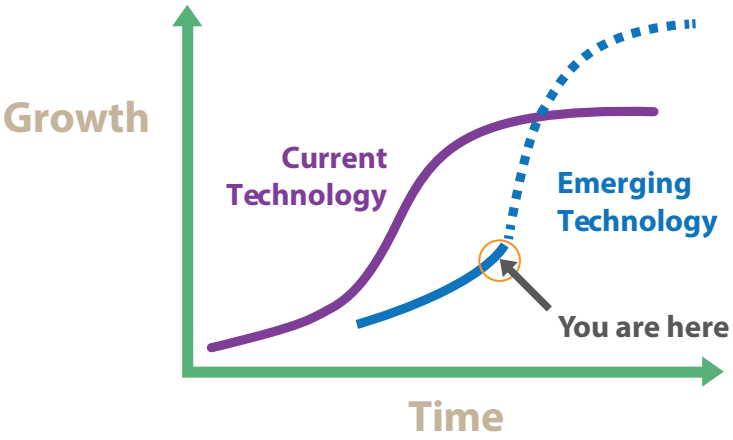

Figure 1: Technology/Innovation Cycle

New Zealand GAP is already using a number of new tools and platforms that are emerging as part of this change. These include geospatial mapping tools (ESRI), e-learning platforms (iSpring), data visualization systems (Tableau), and smart forms (Formstack). Some of these tools are new, while others have been around for many years (ESRI was founded in 1966) but what is new is the way that they are being developed to interact with each other (a system of systems) and to automate simple and complex tasks.

New Zealand GAP is anticipating and planning for a phase of rapid change is this area. This includes:

- Analysing and reporting vast amounts of complex data in a simple and meaningful way
- Customers and certification bodies viewing and auditing records from their office (perhaps overseas)
- Consumers having access to more information about food e.g. safety, environmental, or human rights credentials

3. Key Data Management Topics for New Zealand GAP

There are eight areas that influence horticulture data management for verification as shown in Figure 2. Within the eight topics are overlapping themes, making up a complex and comprehensive data ecosystem. This section drills into each topic, providing information and commentary on data management from the perspective of New Zealand horticulture.

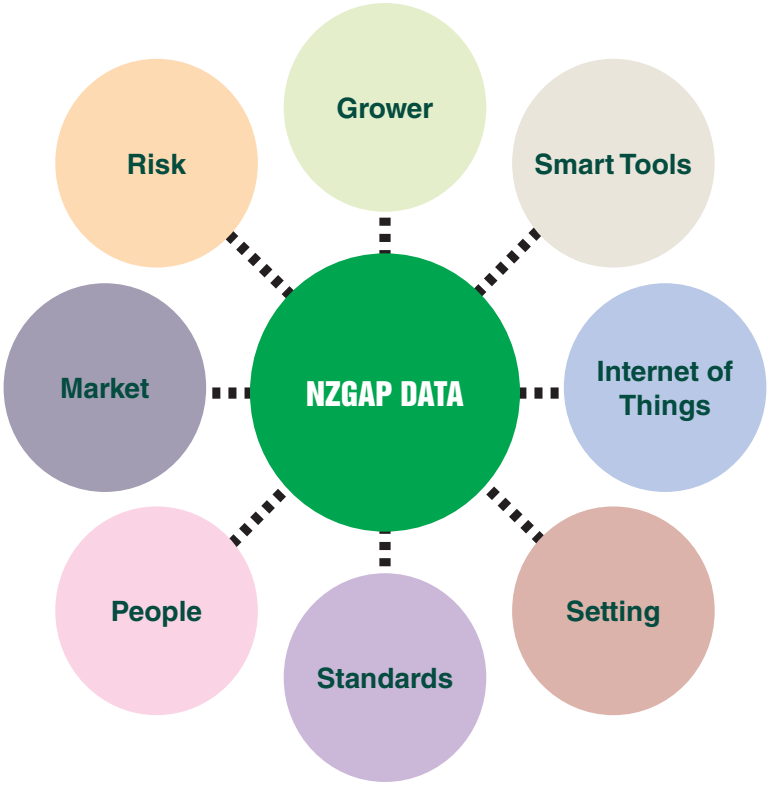

Figure 2: Key horticulture sector data management topics

3.1 The Grower

Growers need access to timely and relevant information about their operation and agricultural practices to assist with daily management tasks, to help inform important business decisions and generate records. There has been an explosion in the number and size of datasets that growers can access and use (Big Data).

Often datasets are held within the measuring device that generate them and are only accessible via the display panel – like a temperature monitor. Recently, these devices can use Wi-Fi to upload data to networked base stations so it can be accessed remotely (i.e. data stored in the cloud) .

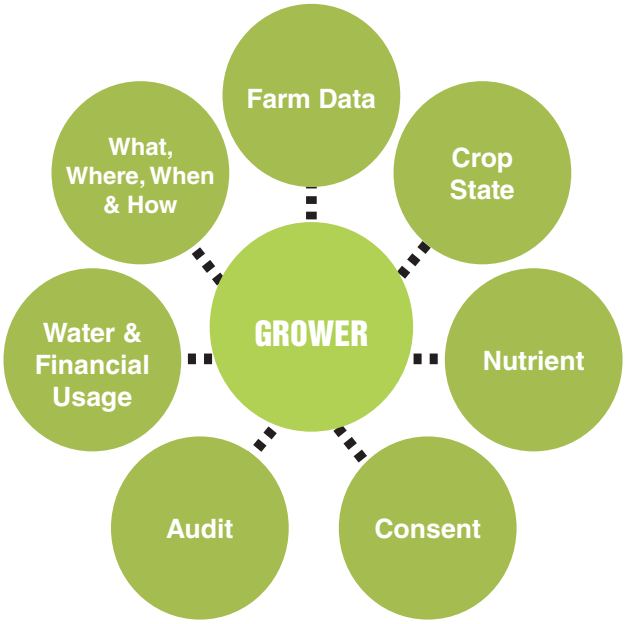

Figure 3: Data collected by a grower

Data is becoming more accessible to the grower.
Datasets collected include:

- day-to-day production activities
- crop state – tracking progress for planting, husbandry and harvest
- location – production areas, agrichemical or fertilizer applications, health & safety hazards
- farm data –, crop yield, spray shed inventory, irrigation, fertilizer use.
- service requests – e.g. requesting a residue test
- regulatory - water usage
- workers– income, salary payments, workhours
- audit and compliance data.

In many cases the result is made up of many disconnected datasets or data silos. A data silo is a dataset that is isolated from other datasets holding related information for the same operation.

Characteristics of data silos are:

- hard to get useful business information from the best data
- barriers to sharing data
- duplicated data
- incompatible and cannot be integrated with other systems
- difficult to establish security measures and to manage access.

There is clear evidence that horticulture businesses are actively collecting, using and managing large quantities of data, or "big data". What is less clear is whether growers understand what big data is and its potential value.

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualisation, querying and information privacy.[2]

Characteristics of 'Big Data' are:

- Volume – the number of records and the size of the datasets. Traditional techniques for extracting and loading data becomes difficult and there is need to use specialist tools.
- Variety – there are many different types of big datasets.
- Velocity – the increasing speed at which data is created, and the increasing speed at which data can be processed, stored and analysed e.g. a real-time picture of water usage can be presented to a smart phone by processing data acquired by irrigation sprinkler heads.
- Veracity – big data is worthless if it is not accurate. Data needs to be clean and accurate. It needs to be reliable and predictable. This may require processing of logfiles to identify records of low value that need discarding. Duplicate values need to be eliminated.
- Visualisation – Big data poses challenges and opportunities to provide powerful tools to explore the dataset.
- Visualisation tools are often interactive and increasingly deployed to web browsers or mobile devices.
- Value – extracting business information and insights.

2 https://en.wikipedia.org/wiki/Big_data

3.2
Harvesting data - The Internet of Things (IoT)

The Internet of Things is the term that refers to a global network of connected objects. It is a convergence of multiple technologies that include wireless communication, embedded systems, control systems, home and organisation automation. These all contribute to enabling the Internet of Things.

The number of connected devices is rapidly increasing. This is driven by a combination of factors that include; reduction in prices for sensors, processors and networks, the widespread adoption of Wi-Fi to extend the network to where it is needed.

In 2016 there are more devices connected to the internet than the global population. This will increase exponentially in the future.

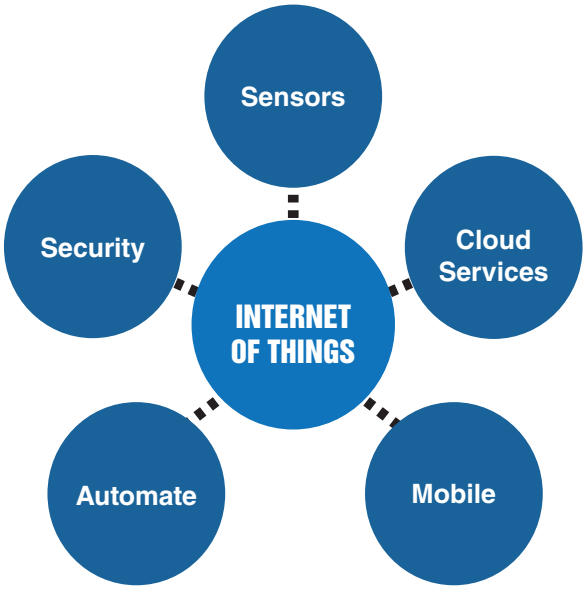

Figure 4: Harvesting data with the Internet of Things

There are two main uses of IoT; consumer grade (used in the home), and business grade. In terms of business, Gartner[3] consider there are two classes of connected things:

1. Generic or cross-industry devices used in multiple industries (e.g. light bulbs).
2. Vertical - found in particular industries. Vertical-specific devices are specialised and include tracking devices in containers, climate sensors, irrigation monitors and the like.

The use of IoT devices is common in precision horticulture. The use of IoT of interest to New Zealand GAP and growers include:

- Cloud services which aggregate data and provide services.
- Climate sensors record numerous climatic variables, e.g. when spraying, sending information not just to the grower but to the service provider. This allows a region-wide service where all growers benefit from data collected from various growers across the area.
- Equipment that can record where it has been, what spray or fertiliser was applied where and at what rate.
- IoT devices can be mobile (e.g. drones)
- Network security is a risk – the common example is hackers accessing web cams.

The table below is taken from Gartner[3] and shows the predicted increase in connected devices in the next few years.

Year	Number devices (Millions)	Spending ($US Billion)	Increase	
			Device Numbers	Total Spending
2015	4,902	1,183		
2016	6,392	1,414	30%	20%
2020	20,797	3,010	225%	112%

The number of IoT devices will grow from approximately 6.4 billion in 2016 to over 20 billion in 2020. The growth in IoT devices in horticulture is likely to follow a similar trend. What is less certain is the investment and skills required to manage the data. Managing data collected by IoT devices brings challenges, opportunities and some risk.

3 http://www.gartner.com/newsroom/id/3165317

3.2.1 Data Quality

All information products are vulnerable to the concept of "garbage in - garbage out". The importance of data quality standards cannot be understated.

Not calibrating sensors leads to datasets that hold less reliable data and lowers the overall quality of the dataset. There is risk that operational decisions are based on inadequate data leading to poor business outcomes.

Characteristics of data silos are:

- Who is at fault – the supplier of the sensor, or the operator who chose not to calibrate?
- How can you find out if information is reliable and can be trusted? Is it based on data acquired by well-maintained and calibrated instruments?
- Does maintenance and calibration history need to be accessible to users?
- What are the appropriate quality standards?
- Who is liable?

3.2.2 Data Management

IoT devices and sensors connected to the grower's farm network collect huge quantities of data. There are challenges in storing, managing and using these datasets. Overcoming these challenges require good technical data management skills and IT resources. Many datasets may be combined to form "data lakes", which can then be scrutinised for analysis and reporting.

3.2.3 Metadata

Metadata is data that holds information about data. Metadata may hold details of instrument calibration, or the date that a data record was last updated. Metadata often holds descriptive information and is useful when searching for data, or trying to determine if a dataset is fit for use. In some sectors (e.g. Spatial Sciences) there are standards[4] that provide guidance on what metadata to collect. New Zealand GAP is considering how organizations or individuals could effectively manage and oversee the metadata that is used to support assurance and verification in New Zealand horticulture.

3.2.4 Security

There are significant risks posed by IoT device security. Many generic IoT devices used by consumers have been reported to have hard-coded passwords, making it very easy for hackers to access the device. In some cases, these IoT devices introduce weak points that compromise an otherwise secure network.
IoT security is an important consideration. There have been numerous incidents and demonstrations of capability by hackers to remotely take control of everyday devices.

3.2.5 Network capacity

The size of datasets collected by IoT devices will use significant network capacity. Unlike other broadband internet users, growers are likely to push more data up than they download from the internet. The ability to provide sufficient upload data capacity (to send data) will be challenging for both growers and Internet Service Providers.

4 http://www.opengeospatial.org/standards/cat

3.3
Smart Tools

Smart Tools provide a wide range of capability – from field data collection through to analytical tools that may provide access to information for auditors and produce buyers. They are typically deployed on smart phones and tablets and follow the pattern of single-purpose easy to use apps downloaded from the app store, or play store.

Some Smart Tools are tightly coupled to datasets collected by a specific instrument, or service provider.

Smart Tools are not just for the grower. Increasingly, apps are targeted at buyers, consumers, consultants, auditors, and contractors.

Of major benefit to growers is the fact that highly relevant information is available anywhere, anytime. This includes:

- Alerts to ensure that decisions or actions can be made at the exact point in time when they are needed.
- Dashboards to provide overview information.

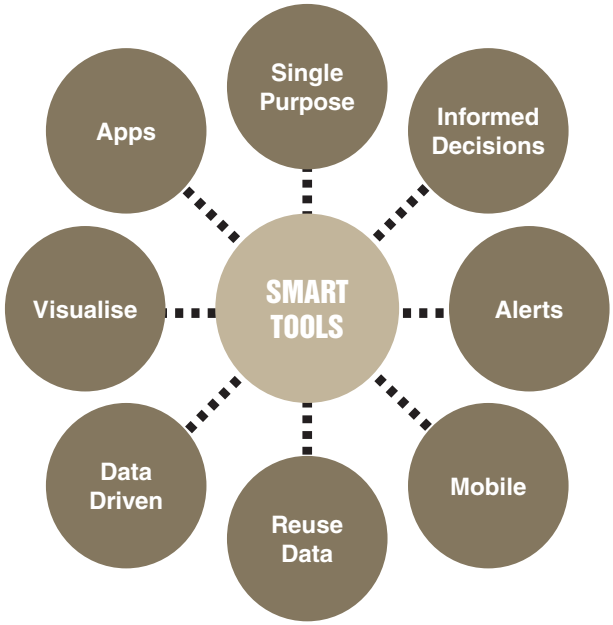

Figure 5: Smart Tools

A future goal for smart tools is to access quality data to provide a more detailed and complete picture. This requires accessing data from many unconnected datasets. There are challenges to aggregating data. In some cases, there will be value and business reasons to establish and maintain a single, aggregated dataset. In other cases, perhaps for privacy, the disparate datasets will never be aggregated.

Regulators are creating smart tools.
Environment Canterbury has published a web-based tool that calculates Nitrogen losses at Good Management practices as defined by industry. It relies on combining information form several datasets that include the Agribase FarmID, Legal description from Land Information NZ, S-map soil information, climate data, OVERSEER data and irrigation information.

https://farmportal.ecan.govt.nz/Home

Smart tools will be the key access point for businesses and growers to create well organised datasets. A significant amount of work is needed to establish data standards to enable data to be accessed and explored.

3.4
Guiding Data Management - Standards

Data management standards make it easier to create, share, and integrate data by making sure that there is a clear understanding of how the data are represented and that the data is in a form that is expected.[5]

Figure 6: Guiding Data Management

In general, data management standards provide:

- guidance to help the business owner
- technical specifications to allow specialist service providers deliver quality products and services
- efficiency – data is easy to use, reliable and can be reused.

From a technical viewpoint, the types of data management standards horticulture need to consider include:

- Data models – definition of the data structure
- Messaging – describes how data packets are formed
- Semantic Web – an extension of the Web through standards by the World Wide Web Consortium (W3C) to promote common data formats and exchange protocols on the Web.[6]
- Metadata – data about data.

Standards for managing datasets will provide robust and trusted datasets that are easy to access and use. Data management standards have been widely adopted globally by many sectors - including Spatial Sciences and Health sectors that also needed to resolve how to manage data effectively.

- The global standard that defines how medical records are shared, integrated, transferred and retrieved is called HL7[7]. The vision for HL7 is "a world in which everyone can securely access and use the right health data when and where they need it"
- The Open Geospatial Consortium (OGC) is an international not for profit organisation committed to making quality open standards for the global geospatial community[8]. OGC standards are widely used in many sectors, including agriculture and horticulture.

Location (place) is important to horticulture. The opportunity is for the New Zealand horticulture industry to adopt the appropriate Open Geospatial Consortium (OGC) standard at an industry level.

5 http://www.usgs.gov/datamanagement/plan/datastandards.php
6 https://en.wikipedia.org/wiki/Semantic_Web
7 HL7 - http://www.hl7.org/
8 http://www.opengeospatial.org/

3.4.1 Data Governance

Data governance is a key consideration about data management. For horticulture, there is no single entity that 'owns' a dataset. This complicates data governance.

In general, data governance delivers[9]:

- Data policies guide what to do, and what not to do with data. Policies describe rules that control integrity, security, quality and use of data during its lifecycle. New Zealand GAP has data policies to enable it to use and manage basic data, but is aware that these need to be continually developed to manage complexity and changing risk in the future.
- Data standards are more detailed rules on managing data. Standards include modelling, architecture and messaging standards
- Issues – data governance controls the process of resolving data management problems including quality, data naming, conflicts in business rules, security and service-level issues
- Coordinate projects and services

Organisations like New Zealand GAP (HortNZ), GLOBALG.A.P. and horticulture industry organisations are well placed to be able to fulfil some data governance roles and provide some of the deliverables required for governance.

3.4.2 Compliance

An increasingly important dataset for horticulture is compliance. Compliance data is based on standards that cover certification of operators and equipment. To meet compliance queries, the grower needs to provide information on how they conformed to regulations and accepted practices for food safety, environment and worker welfare. These are topics that may be covered by developing industry accepted codes of practice.

9 http://www.oracle.com/technetwork/articles/entarch/oea-best-practices-data-gov-400760.pdf

3.5
The Operational Setting – Contextual data

Measuring and reporting on environmental conditions is becoming increasingly important for customer and regulatory compliance. There are datasets that provide detail on the unique and specific soil, climatic and other physical characteristics. These can be thought of as contextual as like the physical environment they provide the backdrop to the orchard or farm. Some of these datasets are maintained at a regional scale; others are collected and managed on a national scale.

Datasets that help by providing horticulture with contextual information include:

Physical data - topographic maps, aerial imagery, 3Dlidar, elevation models
Climate data - sunshine, rainfall, wind direction, temperature
Regulatory commitments - total catchment water usage, minimum water quality
Planning - District plans and permitted activities
Farm location

Figure 7: Contextual datasets

3.5.1 Fit for purpose

As with any dataset, the purpose for which it was created and maintained needs to be known and understood so that it can be used appropriately. There are risks with using datasets that are not fit-for-use. Often datasets are widely adopted and used because they appear to provide suitable information. However, they may have been developed for just one land use, or for use at a particular scale (e.g. for pastoral farming) that is not detailed enough for horticulture.

A good example is the use of Overseer (nutrient budget system) in a regulatory context. There may be more appropriate tools available such as SPASMO (Soil Plant Atmosphere System Model) developed by HortResearch. This tool is integrated into the NIWA CLUES nutrient modelling platform although is yet to be incorporated into the widely adopted OVERSEER model. The underlying concept of Overseer may be valid, but there is need for research to understand nutrient usage in the horticultural context. With this understanding, datasets and models can be revised accordingly and then used confidently by horticulture.

3.5.2 Strategic datasets

There are a number of reference datasets that are used across the industry to calculate limits or thresholds for growing activities. It is important that these datasets are appropriate for their use, are based on current science and are overseen and governed by competent organisations.

The challenge for horticulture is the leadership conversation around the use and maintenance of critical datasets created for one purpose being used to influence planning or regulating horticulture. This may be difficult when the datasets are owned by other sectors and there is need for investment (e.g. certification bodies, primary industries, Regional Councils, Central Government).

3.6
The Market

The market will be looking for simple to use, single purpose tools that provide the information they need. They need to discover what data is available. Over time, their queries will be complex and sophisticated as they seek information about where a crop was grown and what inputs were used. They may want information on the environmental status of the local environment and management of the workforce. In return, growers may be able to access data about the issues that consumers care about.

The market is already awash in data:

- Increasingly, consumers want to know the provenance of food
- There are sales datasets that record the habits of consumers, broken down by demographics, location and what they purchased.
- Data used to determine the most efficient way of getting products to market, and tracking them while in transit.
- Economists tracking prices

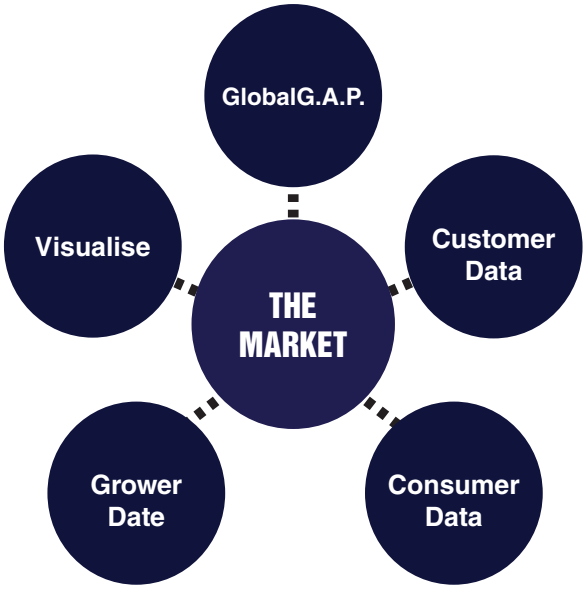

Figure 8: Data Managements themes in the context of the Market

3.7
People and Capability

To meet the demand for changes in how the horticulture sector collects, manages and analyses data, there is need for skilled people. Ideally the sector should be seen as attractive to work in because it has interesting and challenging data problems to resolve. These challenges should be attractive to both those of a technical nature and those more interested in strategic leadership challenges.

In terms of people and capability there may be datasets to acquire and manage.
Growers record activities for health and safety management, such as accidents, incidents, and near misses.

Human rights information includes the legality of workers, employment agreements and payment of wages, etc. There may be a need to certify input provides, such as labour contractors, so that datasets created and maintained by second parties can be assessed as art of an audit process.

Figure 9: Data topics and sector capability

3.8
Risk Management

There are several different ways to view risk in the context of horticulture data.

Some datasets help identify potential risk. Risk may be adverted or minimised by the use of:

- Climate and weather data to support decision making
- Records of the activities of a contractor while on the property
- Obtaining evidence of the employment practices of a labour contractor
- GROWSAFE certification records and experience of a spray operator

There are data specific topics that can pose significant risk with serious consequences.

- Collecting and relying on bad data
- Combining two datasets that use different units of measure or to different levels of precision and accuracy.
- Poor security on data and network assets

Figure 10: Data topics that influence Risk

Understanding how to mitigate risk is often the underlying problem that many smart tools are trying to address. The ability for apps on phones to trigger alerts and escalate those alerts to different people has been adopted widely. Another example is a wind speed alarm to manage spray drift during an agrichemical application.

4. Discussion

This paper uses the term data to mean the raw, discrete data collected by a device, sensor or person. This contrasts with information, which has context and has been summarised or processed. This is in line with the data, information, knowledge, wisdom spectrum as illustrated in the diagram below.

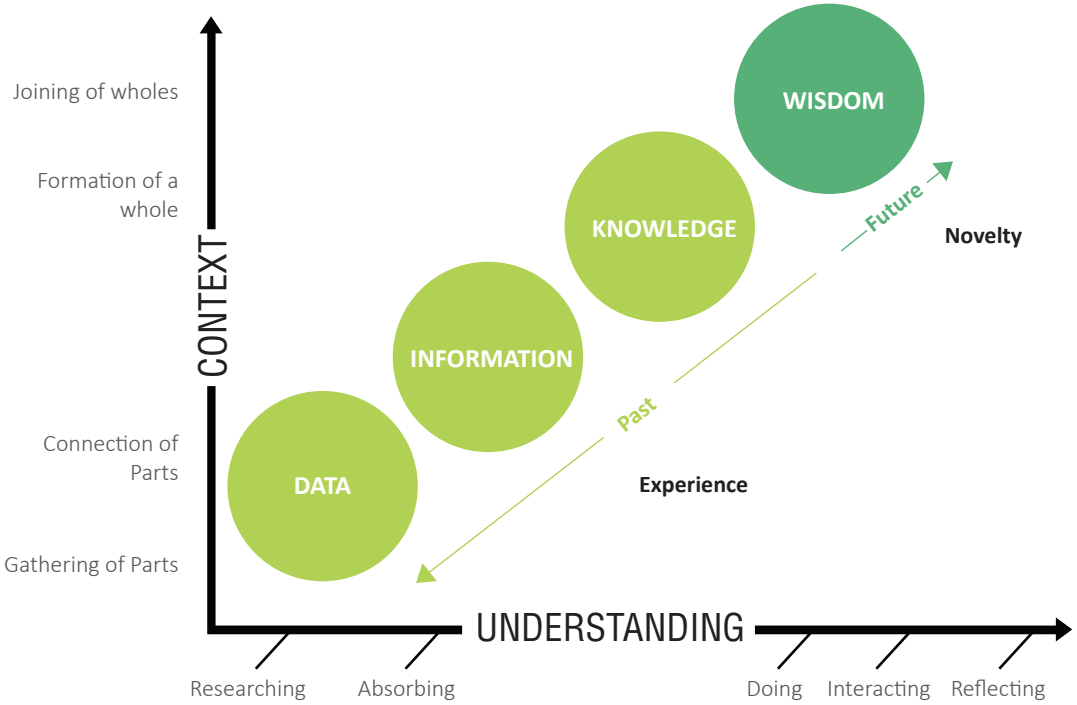

Figure 11: Data, information, knowledge, wisdom spectrum

Information should be distilled clearly – Water usage presented as "litres per kg" of product doesn't tell a consumer anything useful. However, a certification that, based on farming techniques and whether it is grown in a water stressed area, confirms the produce is grown in a "water-friendly" way could be useful.[10]

Information can be considered a message communicated from a sender to a receiver. There are patterns within information and these become knowledge when a person is able to realise and understand the patterns and their implications[11]. In the context of data management for horticulture, Figure 11 provides a guide on how to proceed.

4.1 Challenges and Opportunities

4.1.1 Big Data

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualisation, querying and information privacy.[12] Along with the Internet of Things, Big Data is one of the biggest challenges facing data management in horticulture.

10 Marc Postle, Environment Economist, Jacobs UK.
11 http://www.nwlink.com/~donclark/performance/understanding.html
12 https://en.wikipedia.org/wiki/Big_data

4.1.2 The data iceberg

Much of the data collected and maintained by any organisation is not actively used. Only part of the total dataset is "above water". This is not a new analogy, but big data contributes to the size and number of "data icebergs" This diagram summarises some of the data management tasks required to be able to efficiently use Big Data.

The underlying effort required to ensure big data delivers includes:

- Discover what is being collected, where it is held and how to access it
- Analyse the content and metadata (e.g. logfiles) to release value that is present in the data.
- Cleanse data – remove data that is not needed or is ROT (Redundant, Obsolete or Trivial)
- Transform – Standardise and reduce content thus creating information ready for use. Add, modify, aggregate and create metadata to make it easier to discover information.
- Migrate – Upload to a structured repository.

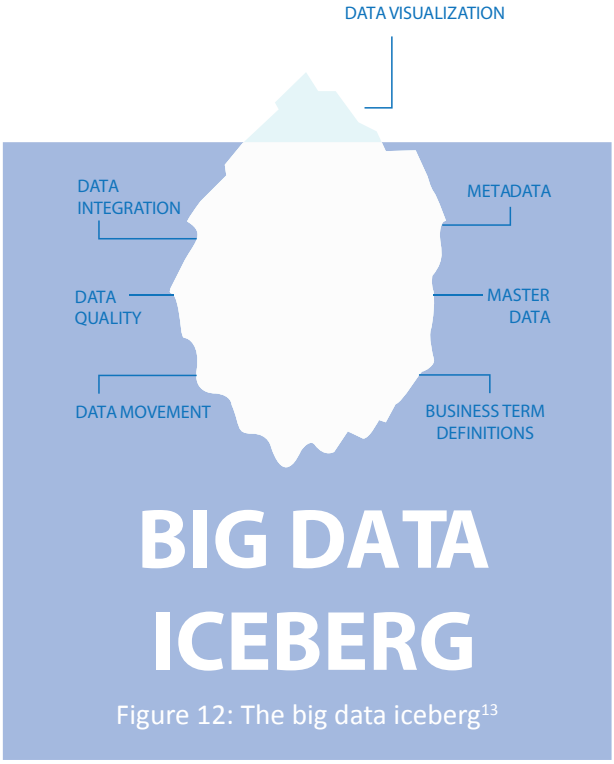

Figure 12: The big data iceberg[13]

The benefits of processing the underlying data include separation of the raw data from the analytical dataset that supports high value business decisions. This also allows better access control and reduced risk of unauthorised access to sensitive raw data.

4.1.3 Data Quality

Data Quality is essential to establish high value and easy to access and use datasets.

Topics to consider include;

- Defining quality standards
- Establishing and maintaining quality data collection.
  - Using technology correctly – within specifications and for the purpose for which it was intended
  - Maintaining technology – sensor and instrument calibration. Replacing equipment when instrument tests fail
  - Actively finding and pruning bad records
  - Recording technology maintenance and data management tasks
  - Creating keys that make it easy to find a record or a dataset
- IT challenges posed by data management, including security and access permissions
- Metadata – data that describes data (e.g. creation date, sensor model that created the data record).

13 http://timoelliott.com/blog/2015/06/the-big-data-iceberg.html

4.1.4 Master Data Management

Master Data Management (MDM) comprises the processes, governance, policies, standards and tools that consistently define and manage the critical data of an organisation to provide a single point of reference.

The data that is mastered may include:

- Reference data
- Analytical data – supports decision making

Implementing MDM eliminates duplicate records, standardises data and defines rules that prevent incorrect or data that is not required from entering the system. The result is a clean, authoritative dataset of Master Data that can be used across all systems.

4.1.5 Data is an asset

Data is an asset – it is valuable and it is useful. Treating data as an asset means that the data owner has responsibilities in terms of quality, usability and reliability. Some datasets are mission critical and any error, or omission can have consequences in terms of safety, or reputational risk.

The challenge is to treat every dataset as an asset. This may need investment in specialist skills. Challenges include:

- Format variability.
- Big Data is often unstructured and classical data management techniques do not work so well. Video, 3D data, Lidar, remote sensing, and photographs are all examples of big, unstructured datasets that need different data management strategies.
- Spreadsheet Sprawl. It is so easy to open a dataset in Excel and create a spreadsheet. And then do it again. The net result is a massive number of spreadsheets and no easy way to know which ones have the latest and most reliable data.
- Data locked into a Smart App or a sensor.

The benefit that comes from treating data as an asset is that because there are real costs to acquire, manage and store the data, the data owner should ask the question "is the investment in the data asset useful"?. As with other assets, this focuses expenditure on what is absolutely needed and not what is nice to have. The opportunity here is to outsource the acquisition and management of data.

4.1.6 The Supply Chain of Data

Supply chains ensure continuity of products so they meet standards (e.g. Coke tastes the same worldwide). The challenge is to do the same with data – to create a supply chain of data so the quality is always the same and so users can trust what they get. Supply chains are expensive to establish. Building a supply chain of data is an ambitious task which will take time, skill and money.

4.1.7 Data Silos

Data management in horticulture is not part of a "single-system". The result is that data and information products are held in silos. The challenge is to leverage these to share information and to create virtual system that has access to data from all stages. The opportunity is to define standards to help ensure data silos can be easily accessed and high-quality information extracted.

4.1.8 Dataset Version

Different versions of the same dataset are a problem. Often the later version has only a few records more than the older version. They are also very common – particularly in the use of excel spreadsheets.
Good data management practices can eliminate multiple versions of data, or with good metadata, make it easy to discover information about each version.

The issue is when downstream informational products (e.g. models, graphics and maps) have been derived from a particular version of data. The challenge is for the data owner to manage how their data is versioned and provide simple ways for users to discover versioning information.

4.1.9 Concept of a Grower's Data Locker

Imagine if all the datasets collected by a grower were available from a single source. And that source could be shared – so authorised users could readily access data without having to ask the grower for their latest water usage data, or for crop status, or spray history and current residue levels. This is the concept of a data locker.

There are big challenges to establishing data lockers. One challenge is security. Another is standards that define the methods for discovering data and then accessing it. But, there are some big and real benefits too.

4.1.10 Disruptive Technologies

New technology and new ideas will prove disruptive. Peer-to-peer sharing is an approach to data sharing is proving to be a useful approach for assurance systems. This technology is similar to that used in Linkedin, where individuals agree to share, or un-share information by mutual agreement. The peer-to-peer approach has real benefits in relation to sharing assurance data as it enables the owner of the data to control its distribution.

One example of peer-to-peer data sharing in agriculture is the US-based Farmers business network. It allows farmers to discover what seed other farmers use and the resultant crop yields. This provides a real-world and disruptive view of seed performance as it supersedes the expected yields based on trial data traditionally provided by seed distributors and used to help sell seed.
Reference: https://www.farmersbusinessnetwork.com/.

4.2 Future Data Flows in New Zealand GAP

The future flow of assurance data and information may look very different in a business intelligence architecture (Figure 13). This structure allows multiple devices (apps) and databases to interact with a core system and pool data into an enterprise "data lake". A data lake is like a large storage warehouse that holds vast amounts of raw data where it is held until it is needed for analysis or sharing. The main benefit of a data lake is the centralisation of fragmented data sources (i.e. data silos), from which big data searches, analytics and reporting can be easily performed using modern tools and computing power. This is impossible with data silos.

As sensitive and proprietary information is often held in an organisations database, security and data sharing policies will be required to grant access to certain information. The establishment of data standards, governance and the resourcing of the system are key challenges which need to be overcome. Collaboration between New Zealand GAP and its stakeholders will also be fundamental to success.

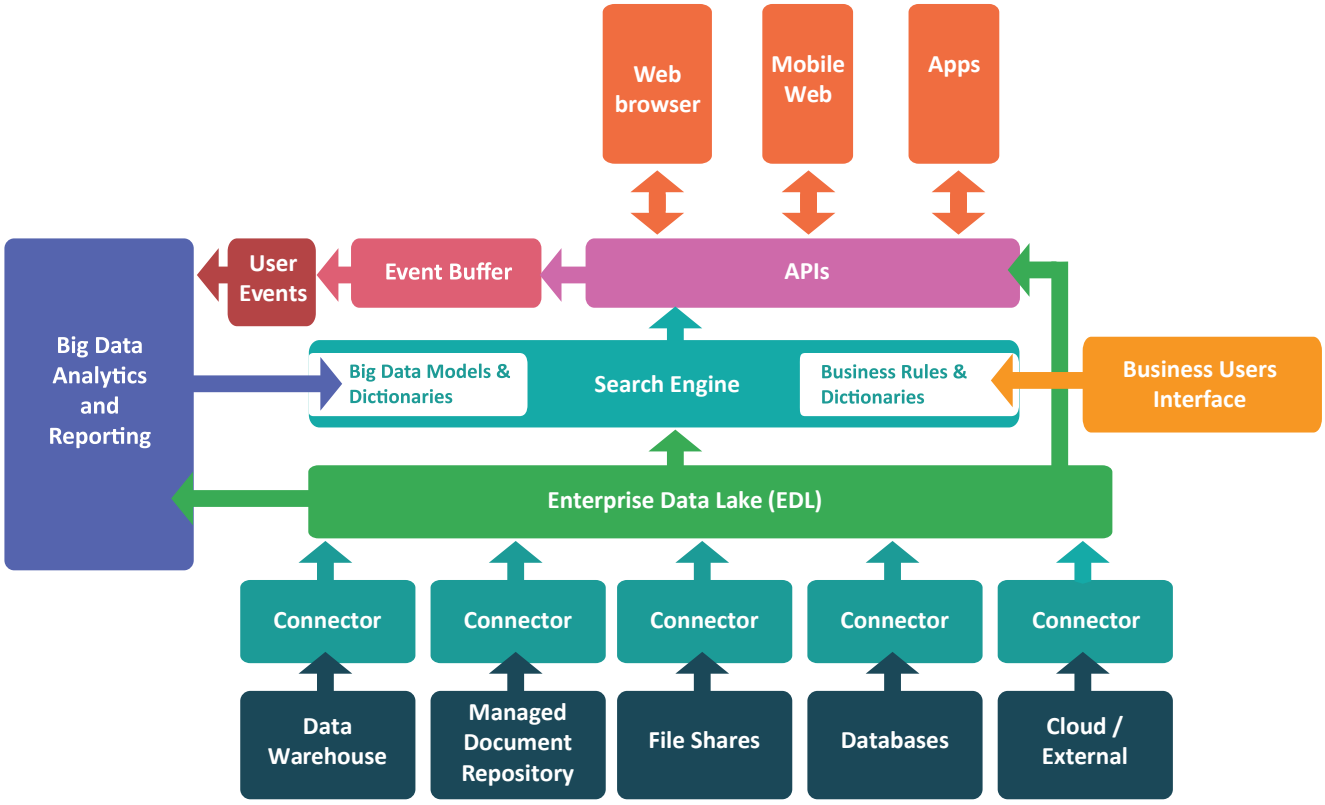

Figure 13: Enterprise Data Lake (EDL) High-Level Architecture
Source: https://dzone.com/articles/a-data-lake-architecture-with-hadoop-and-open-sour

Relevant apps and databases may be linked to the data lake using an Application Programming Interface (API), which will continually update the data in the central system, allowing for live data streams to provide crucial information and reports for informing day to day decision making and actions.
An example application of this system would be the sharing of tailored audit reports for a market or regulator, to reduce administrative costs on growers and the assurance system.

For example, the Food Act (2014) is currently being implemented for horticulture, and as there is a current lack of a centralised database of certified businesses, so growers may need to individually register with their local council rather than using the "data lake" approach, which would allow MPI to see a business's compliance centrally (e.g. the New Zealand GAP register). Once the system is established it will reduce complexity, administration costs, and the cost of registration for growers by meeting multiple standards for a range of customers with one core system.

5. Conclusion

New Zealand GAP will increasingly encounter data management issues more commonly found in the Information Technology (IT) sector. In the supply chain side, there have already been significant IT investments (e.g. Zespri). Growers need to be able to confidently participate in IT conversations that include maintaining data quality, master data management, collecting and maintaining metadata, backup strategies and being able to reliably and efficiently export and import large volumes of data. Big data and the Internet of Things will rapidly introduce change and provide both challenges and opportunities to horticulture data management. Innovative Smart Tools will continue to be released onto mobile devices and smart phones.

There are opportunities for New Zealand GAP to help develop a central system to influence how assurance data in horticulture is used, visualised and shared, ensuring smart tools are intuitive to use and are appropriate for the audience. The growth of new networking technology like the 5G mobile network and Wi-Fi will make it easier to deploy applications and to acquire even more data from networked devices in an orchard, or field crop. There are opportunities for data-savvy growers to create joined-up datasets that provide a picture of their entire operation and for brave growers to share parts of this information with others for mutual benefit and to add value.

Development and adoption of standards are essential to realise these opportunities, as there are many data management issues facing horticulture. The volume of data being uploaded and shared by growers will continue to increase – and this may cause problems for some Internet Service Providers. It is not clear if infrastructure initiatives like the Rural Broadband Initiative provide the network capability and capacity that will be required.

Using this big data system, growers may have a feedback loop which adds value to their business by providing them with useful information on:

- Performance during an audit
- Resource use compared to their peers (e.g. industry/regional average)
- Decision support for spraying, irrigation, fertiliser, harvest, reducing costs
- Identifying risks and issues
- Continual improvement

There is a need for leadership and collaboration within the horticulture sector, including taking an active part in the development of data management standards and to ensure information used is fit for purpose. Leadership is also required to influence other sectors outside horticulture. There is a need to raise with the telecommunications sector the issue of network traffic volumes and the concerns regarding upload capacity.

The risk of doing nothing is that the opportunity to influence, guide and lead the development of standardised datasets and information products is lost. By taking a lead, New Zealand GAP can realise the goals of providing high quality and reliable data and reports, related to assurance, to those who need it.

In summary, this report recommends the investment of time and effort to:

1. Develop a technology and data strategy
2. Lead the development of global, open sector wide data management standards
3. Treat data as an asset
4. Actively manage data, discard what is not needed and providing methods to discover data
5. Establish a supply chain of data. Create information from the raw data to help improve understanding
6. Participate in leadership conversations at a national level on topics that include using datasets that fit-for-purpose and network infrastructure
7. Grow work force capability so there are people who have both technical skills and useful horticulture knowledge.

Acknowledgements

New Zealand GAP has prepared this report with the assistance of Jacobs Engineering Group, an international organization with expertise in construction, natural resources and engineering, including technology and large scale geospatial data systems.

The research for this report included interviewing a number of businesses and organizations in horticulture industry that are developing and using the types technology described in the report. We would like to acknowledge and thank these organisations; Landwise Inc, Agfirst Ltd, Heinz Watties, Hortplus Ltd.

Contact us
Horticulture New Zealand,
Freepost 729, PO Box 10232,
The Terrace, Wellington, 6143

Phone: **+64 4 472 3795**
Free phone: **0508 467 869**

Website: **www.nzgap.org.nz**