Use of big data in cattle practice

Christopher David Hudson; School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Sutton Bonington. Leicestershire LE12 5RD. chris.hudson@nottingham.ac.uk (corresponding author)

Jasmeet Kaler; School of Veterinary Medicine and Science, University of Nottingham

Peter Down; School of Veterinary Medicine and Science, University of Nottingham

Word count: 2,983 (excluding title page, abstract, tables, boxes, references and figure legends)

Abstract

The concept of big data, associated data sources and analytics is becoming increasingly talked about both in society as a whole and within the livestock industry. This article provides a clinician-focused review of what big data means, how it is already influencing farm and veterinary businesses, and where this may lead in the future.

Introduction

The first problem faced by the practitioner wishing to embrace the big data revolution is in understanding what the term "big data" really means. It will be clear to readers that it has become a highly prevalent buzz-phrase over the past 10 years. As is often the case, this has spawned an explosion of related technical terms and jargon, many of which have crossed over into common usage, in some cases despite being poorly understood outside of the technology sector. Box 1 provides a glossary of some of these terms. This article aims to provide the cattle practitioner with an accessible overview of how big data and associated ideas and technology may influence farm businesses and veterinary practice. For a more technically focused review, readers are referred to Wolfert and others (2017).

Box 1 – Big data jargon

- **Artificial intelligence**: A general term for the branch of computer science which deals with simulating or using mechanisms from the way humans think in order to solve problems.
- **Machine learning**: A set of tools (usually considered to be a subset of artificial intelligence), most often used for classification, prediction or pattern recognition problems. Machine learning techniques are often defined by their ability to "learn" (usually in the sense of modifying an algorithm) from data.
- **Algorithm**: A sequentially defined set of operations which convert one or more inputs into one or more outputs. Many machine learning (qv) techniques result in algorithms designed to solve a particular problem (e.g. classifying the raw output from multiple on-cow sensors into a yes/no output representing whether the cow is likely to be in oestrus). Algorithms can often be represented as flow charts.
- **Internet of (agri) things**: Generally refers to the connection of devices to the internet, including devices not primarily used for internet access (such as computers, tablets and smartphones). In the wider world, this pertains to objects as simple as light switches, or as complex as cars. Within agriculture, objects could be tractors and other machinery or on-animal sensors.
- **Precision agriculture**: The concept that farming outcomes can be improved by making measurements and decisions at a more granular (i.e. precise) level of detail than has traditionally been the case. The phrase originated mostly in relation to arable farming, typically implying use of sensor technology to modify an activity (e.g. using sensors mounted on farm machinery or a drone to adjust chemical application rates for small sub-areas within fields, rather than at whole field level). More recently, terms such as "precision livestock farming" and "precision dairying" have become much more commonly used. In many ways, technologies such as robotic milking (which pre-dates most arable applications by some years) are good examples of precision agriculture, but future applications could include, for example, scraping systems driven by image analysis.
- **Disruptive technology**: An innovation – often either a physical product or a new data processing or analytical approach – which forces other market competitors to change their offering; this can be as a result of superior performance, reduced price or other differentiating factors.

A number of different definitions of “big data” are provided by authoritative sources (commercial technology organisations such as Microsoft and Google are very commonly cited sources of definition), but a common theme amongst these is the “Vs” of big data. Originally, the “four Vs” were often referred to, but over time the list has grown, and some sources refer to as many as ten Vs. The original four Vs are:

- **Volume:** Perhaps the most self-evident: most definitions of big data start with the fact that it needs to be “big”! Key problems here are that bigness is inherently a subjective and relative concept (so what would be considered big data for the beef industry may be extremely small in absolute size compared to, say, data generated by social media interactions), and also that it tends to change over time as the capacity to collect and store data expands and costs fall.
- **Veracity:** Data quality is a key consideration wherever data is analysed, but as the volume of data grows and it is increasingly used to inform high-stakes decisions, it becomes even more critical. This can be particularly problematic where data is being used for purposes other than that for which it was originally recorded, and where multiple sources of data are integrated for analysis. Data quality is discussed in Box 2.
- **Velocity:** This can either refer to the speed with which information is accumulated, or the speed with which it can be interrogated and analysed.
- **Variety:** The ability to integrate multiple sources of data to provide additional insight is one of the key benefits of applying big data approaches.

Visualisation is also commonly mentioned, referring to the need for effective graphical ways to visualise large datasets in order for end users to be able to make use of them quickly and easily (see Hermans and others (2018) for a summary of some aspects relevant to dairy data). Value refers to the requirement for any use of big data to add value to a process in order to be useful. Other less commonly mentioned concepts include volatility (usually meaning the time for which data is considered useful before becoming obsolete), variability (the concept that, as well as coming from multiple sources, the nature of the data may change), vulnerability (the organisational risks associated with storing big data, especially where it includes personal data) and value. Examples of how some of these concepts of big data apply in dairy farming are shown in Figure 1. Interestingly, more recent attempts to define big data have become wider, and the term is now often used to refer to any situation in where data are used to inform decision making. Much current activity in herd health and production management in dairy herds falls under this definition.

Box 2 - Data quality

For as long as clinicians have been attempting to use herd data to monitor performance and health, it has been widely recognised that a certain level of data quality is required for meaningful analysis: “garbage in, garbage out” is a common maxim. In the context of cattle farming, data quality most commonly refers to how accurately a given set of data reflects the events which have occurred in real life. A classic example would be the accuracy of recording insemination events in a dairy herd. Where such events are under-recorded (i.e. not all inseminations result in a record), the herd’s submission rate (proportion of eligible cows inseminated every 21 days) will tend to be under-estimated, whilst conception rate (proportion of serves leading to a pregnancy) is over-estimated. Where the degree of under-recording is high, this can produce results which appear unlikely to the clinician (for example, a conception rate over 60% is unlikely in most circumstances). When monitoring performance in an individual herd, this is critical to bear in mind, but this becomes even more important where "big data" principles are used to calculate performance metrics across a large number of datasets for benchmarking or automated reporting. In such contexts, efforts to develop and apply methods for measuring data quality without human input are useful. There are a number of statistical techniques which can help to detect data which is missing (Hudson 2015; Hermans and others 2017); currently there are some applications of these implemented in software, but it is likely that this process will become more sophisticated and accurate in future.

Big data is pervading most aspects of industry and society, and the dairy and beef sectors are no different. The volume of data available on dairy farms has increased particularly rapidly with the relatively widespread adoption of on-animal sensor technology (most notably activity monitoring) and the advent of milking systems which can collect and store much more detail on the milking process (more obviously in robot systems, but also in conventional parlours). The beef industry is interesting in that there is generally much less data-driven decision making on typical beef enterprises compared to dairy farms; however, there are a number of ways in which big data concepts are particularly applicable to beef farming, and the next decade may see more engagement of the beef sector with recording and using data. For example, using statutory data recording (for example by combining registration and movement data from the online British Cattle Movement Service database with medicines use data) can provide highly useful insight even where record keeping is relatively minimal (Hewitt and others 2018).

It is important to remember that the increasing quantities of data being generated on many farms is worthless from a decision-making perspective unless it is analysed or processed in time to yield critical information which can then be employed to make informed decisions. As in other sectors, agricultural big data will have no real value without appropriate analytics (ZhongFu and others 2013). Historically, this process has been hampered by a failure of analytical techniques and computing power to keep pace with the scale of data collection, leaving decision makers "data rich, information poor" (a phenomenon often referred to as the DRIP conundrum). This has resolved to a large extent in recent years, unlocking value from information in a manner that can support informed decisions (Tien 2013). Machine learning has played a key role in this process, and has become a common term over the past decade. In many ways, machine learning is similar to existing statistical methods, in that it is a way of looking for patterns in data. A key characteristic of a machine learning approach is that its performance tends to improve when it is exposed to more data; some methods are based on understanding of the way human brains process information. Very broadly, machine learning processes usually aim to make a prediction of an outcome based on many inputs (where the system is provided with cases of known outcome to "learn" from; this is "supervised" machine learning) or to identify data patterns or clusters where there is no outcome ("unsupervised" machine learning). Conventional statistical approaches (such as regression modelling, which has been widely applied in the field of dairy science for many decades) are sometimes considered as types of supervised machine learning.

Where can big data add value?

There are many ways in which big data is already adding value to farm businesses, either through improving the performance or efficiency of a system, or by automating processes to reduce labour cost and improve consistency. Use of activity monitoring for oestrus detection is a clear example here, whereby a number of big-data related concepts (on-animal sensor technology, amalgamation of data from multiple sources, data pre-processing and machine learning) come together in a product that can improve performance whilst also reducing labour costs (see Figure 2). Whilst the cost-effectiveness of automatic oestrus detection systems has been reported in the veterinary literature (van Asseldonk and others 1999), there are also examples where investment in sensor systems has failed to result in any tangible improvements in animal health, productivity or farm profitability. One possible explanation of this is that farmers may not fully utilise the potential of sensor systems (Steeneveld and Hogeveen 2015), but it is important to see the potential benefits in the context of the wider farm system. For example, improvement in submission rates as a result of activity monitoring are likely to be much smaller where there is limited space for cows to express heat.

For farmers, the decision to invest in sensor technology to support decision-making will depend largely on the perceived cost-benefit of the system (Lima and others 2018). This is an area where the veterinary practitioner can play a critical role, acting as an independent advisor who can review the evidence and help decide if a particular system is likely to be suitable for the purpose intended. It should always be remembered that new technology is always "competing" for a limited farm investment budget, so money spent in this area cannot be used for other improvements, which in some cases would be expected to give better returns. This can be challenging, as there is often limited good-quality evidence on which to base expectations of a system (see Box 3). In part, this is because improvements will vary a great deal between different farms, so a study comparing (for example) different heat detection systems would need to measure a very large number of herds over a prolonged period of time, making it expensive and difficult to carry out. The nature of machine learning also makes it hard to gather evidence on what to expect from a particular technology, as the algorithms are inherently updateable, and work best when "trained" on additional new data to improve accuracy. In many cases, the expected performance of a system will therefore tend to increase over time. Vets also have a role in helping farmers to get the best value from any systems which they have invested in, for example by helping set up and use alert lists or thresholds within the system or by advising on environmental or management changes to allow technology to work better.

Box 3 - Sensor Technology

Sensors are a major source of big data in cattle farming and represent the area where these techniques are currently most widely applied. Most sensors currently used to measure physiological or behavioural parameters on cattle farms are focussed on the detection/monitoring of mastitis and fertility (oestrus), with a growing number also being marketed for the detection of lameness and metabolic conditions.

- **Mastitis:** Electrical conductivity is the most commonly reported sensor followed by sensors to detect milk colour and certain enzymes such as haptoglobin, l-lactate dehydrogenase and N-acetyl-β-d-glucosaminidase.
- **Oestrus:** The most commonly used measure is cow activity using pedometers, activity meters or 3-dimensional accelerometers. Other fertility-related sensors include milk progesterone, mounting behaviour and body temperature.
- **Lameness:** Sensors include pedometers and activity meters, 3D-accelerometers, force-plates and video cameras (with output analysed by computer vision). For these systems to genuinely add value, they must be able to detect the milder presentations of lameness which are the ones most likely to be missed by farmers (Leach and others 2010). The ability for any of the locomotion sensors

currently available to detect subtle manifestations of lameness remains uncertain (Van Nuffel and others 2015).

- **Metabolic conditions:** This includes sensors that measure the pH of rumen fluid, rumen/ear canal temperature, milk butterfat or betahydroxybutryate concentrations, rumination frequency and body condition score. The use of these sensors is less prevalent commercially perhaps due to the complex nature of many metabolic disorders and less validation of their clinical application within the veterinary literature.

It is difficult to compare the performance of the various sensor types because of the large variation in reported performance, gold standards, test scales, and algorithms used. Of the more commonly evaluated sensors, reported sensitivities and specificities have ranged between 70-91% and 87-98% respectively for conductivity sensors, and 80-90% sensitivity with >90% specificity for heat detection via pedometers/accelerometers (Rutten and others 2013). Whilst this is useful information, specificity in particular can be difficult to interpret in practice where the prevalence of the event is low; and studies reporting a positive predictive value are highly valuable (Holman and others 2011). A growing number of products are being marketed that combine multiple sensors e.g. accelerometers combined with real-time location and rumination time. Such combinations potentially provide a very powerful aid to decision making, but as with any investment it is critical to appraise potential value added against the investment required.

As traditional income streams continue to be eroded in farm animal practice, there is a growing need for vets to be delivering active herd health programmes to their clients, and a change of emphasis from being a reactive ambulatory practitioner to a proactive advice-oriented consultant (Down and others 2012). Big data and associated analytics have the potential to aid with this transition by helping the practitioner to make use of current data to predict different outcomes that would be expected in light of different decisions. Examples of 'predictive biology' already exist in the literature and will rapidly become the norm with advances in on-farm technologies, scientific approaches and data handling capabilities (Green and others 2016). These predictive models are especially powerful when they provide probabilistic outputs, which allow decisions made to reflect the attitude to risk of the decision maker.

Despite the rapid growth in the availability of biosensors and our ever-improving ability to apply big data analytics to the resulting data, the decision making and corresponding interventions remain largely the responsibility of the veterinarian and farm team as the automated treatment of cows remains largely unfeasible at present. This is another important area requiring engagement from the veterinary practitioner who is well placed to provide insights into what the results of data analysis mean, as well as providing evidence-based advice with respect possible interventions. It is likely that decision support systems will play an increasing role in this process as they become integrated into sensor systems, but the vet will always have an important role to play in terms of quality control and herd health advice.

Current applications of big data in cattle farming

Robotic voluntary milking systems provide a good example of existing big data in action on dairy farms. Robotic milking setups are increasingly common, and often include a number of on-animal and inline sensors as part of the system. These typically capture real-time data at a very high level of detail (for example, raw activity or rumination data in small time units recorded by on-cow sensors; or multiple measurements of milk flow and concentrate intake during each milking session). This data is then pre-processed (for example, by smoothing activity data to average over a longer time-period and reduce the "noise" in the signal). This data is often accessible to the user, allowing evaluation of relatively detailed information usually through graphs or other visualisations. In many cases, multiple sources of data are then aggregated and a machine-learning derived algorithm applied to detect deviation from expectation (for example, where rumination, activity, yield and temperature data are combined to generate an alert list of cows which should be examined for signs of ill health; see Figure 2). Much of this system is equally applicable outside of a voluntary milking context, and it seems likely that uptake of this approach in herds milked through conventional parlours will increase in future.

Sensors are not the only area where big data is influencing dairy management: there are an increasing number of examples where other sources of data are amalgamated and analysed (see Table 1 for some examples of data sources). There are a number of commercially available web-based products where multiple sources of data for a herd are amalgamated, processed and visualised. For example, data from a milk recording organisation may be combined with BCMS data and/or event data entered by farm staff to generate charts showing how a particular performance metric changes over time, and how this benchmarks against other similar herds. This process is often dependent on different agencies allowing access to their data via an automated process – this is increasingly common in the dairy industry. This has also been important in arable farming, where a large number of commercial and open-source platforms exist to facilitate data exchange. This "babel" of competing alternative data structures and data exchange systems can sometime be problematic, and a common data schema (for example, a consistent set of event definitions) could facilitate data exchange whilst minimising loss of information.

TABLE 1: SOURCES OF DATA IN CATTLE FARMING

Data source	Dairy/beef /both	Accessibility	Notes
British Cattle Movement Service	Both	Simple, web interface for data download	
Milk recording organisation	Dairy	Simple, often both online analysis and data download	Commonly store basic data, fertility and some health events (either via farm software or transcribed from paper) as well as milk records
Dedicated herd management software	Both	Variable across products	Often provides some performance analysis features, but highly variable
Milking plant/robot	Dairy	Access to "dashboard" info often simple	Data export for separate analysis often more difficult
Animal weights	Both	Variable according to data capture/ storage	Usually recorded either in herd management software or on paper

Activity monitors	Dairy	Access to "dashboard" info often simple	Raw data often hard to access, some systems communicate with other farm software
Other on-animal sensors	Both	Variable across products	
Environmental sensors	Both	Mostly simple	e.g. temperature/humidity monitors
Additional paper/ electronic records	Both	Simple to access but may require digitising for use	e.g. written meds records, scoring data
Processor	Both	Variable between processors	i.e. abattoir for beef finishing enterprises, dairy plant for dairies

The future of big data: Challenges and opportunities

A number of existing projects and near-market or early-life products are likely to influence the use of big data in cattle farming in the near future. A number of industry initiatives aiming to amalgamate data from multiple sources (for example, the Livestock Industry Data Exchange Hub pilot project) are already running. Such projects have potential to make it massively easier for practitioners to access data relating to the animals under their care, although a number of obstacles (including data protection issues) may hinder this. At an on-farm level, the emergence of open-source (i.e. freely available) platforms allowing different farm software products to communicate with each other is also likely. These have the potential to make integration of data from multiple sources (for example, from herd management software, milking plant software and a milk recording organisation) more straightforward, in turn making it simpler for the clinician to analyse data and add value to a farm business. Figure 3 shows an example of data transfer between on-farm systems. Both of these concepts (the centralised "data hub", and the increased exchange of information between on-farm systems) also have the potential to open new doors for research based on routinely collected data.

Availability of centralised data hubs also creates opportunities for syndromic surveillance. This process uses real-time data to assist in early detection of diseases by looking for clusters of events or measurements that deviate from the expected norm. A classic example is the use of frequency of search engine queries for early detection of human influenza outbreaks (Ginsberg and others 2009). As syndromic surveillance represents a relatively low-cost method compared to many conventional approaches, this has attracted some interest at national level, for example in detection of exotic disease incursion (Marceau and others 2014). However, this approach could also be taken at veterinary practice level, for example by identifying farms where measured outcomes (such as conception rate, or milk yield) deviate from what would be expected based on previous data both from that individual herd and from others in the practice. Big data also has the potential to transform research into health, welfare and production on dairy herds. This is true both at individual herd level (for example, by using multiple sensing systems to add value to research on behaviour and disease in facilities such as the Centre for Dairy Science Innovation at the University of Nottingham) and across multiple herds, where data amalgamation techniques and machine learning are unlocking increasing value from routinely recorded data.

As with most disruptive innovations, the advent of big data and associated technologies presents both an opportunity and a threat to cattle practitioners. A clear potential threat to practice income is the improvement in technologies associated with reproductive management in dairy herds. Routine fertility visits have been a core source of fee income from dairy herds for at least the last two decades, but as technology to improve oestrus detection becomes more effective and less costly it is likely that the quantity of veterinary time required will be reduced. This could result from improvements to existing technology (such as the ongoing improvements in activity monitor systems), appearance of new products (such as inline milk progesterone monitoring), or the increase in accessibility of existing products (through improved ease of use and falling cost). This, along with a number of other trends discussed in more detail by Statham and others (2013), places increased emphasis on practice business models generating income from delivery of herd-level advice and consultancy. Big data offers opportunities to enhance and streamline this type of service (see Table 2). The increased availability of algorithmic decision support tools (both at individual animal and at herd level) is also likely to be a positive for clinicians, at least in the short- to medium-term, where they will augment an individual's clinical decision making and allow practitioners to give more evidence-based (and possibly probabilistic) advice. Further into the future, it is possible that such systems could come to represent a threat to veterinary income by partially supplanting the clinician's role.

TABLE 2: EXAMPLES OF HOW A CATTLE CLINICIAN COULD BE INVOLVED WITH BIG DATA

Area	Useful skills	Notes
Helping clients get best value from systems on farm	Understanding both what a sensor is measuring and the underlying biology	
Using big data analytics to improve herd-level decision making	Conventional herd health skills; understanding of basic statistical/epidemiology concepts; marketing skills and business model for consultancy work	Skills in data handling and manipulation less relevant here as this is often done by a product (e.g. a website)
Using big data principles to derive value from practice data	Basic data handling/ visualization skills	e.g. amalgamating practice management software and BCMS data to benchmark medicines use
Advising farmers on potential value from tech investments	Understanding wider farm context; evaluating evidence to assess potential value of this versus competing investments	An outside view on whether an investment is really likely to be useful can be very helpful!

Conclusions

The big data revolution is already having an influence on the dairy industry, as it has in other sectors. The timing is currently ideal for the veterinary profession to ensure that they are at the forefront of these developments, taking the opportunities created by big data and helping farmers maximise value from investment in technology. It is important to remember that this does not generally require a clinician to possess "big data" skills as such; but a skillset including understanding of the underlying biology alongside knowledge of epidemiology gives cattle vets the potential to make a big difference to the amount of value unlocked by these changes.

References

DOWN, P.M., KERBY, M., HALL, J., STATHAM, J.M.E., GREEN, M.J., BREEN, J.E. and HUDSON, C.D. (2012) Providing herd health management in practice - How does it work on farm? *Cattle Practice* **20**, 112–119.

GINSBERG, J., MOHEBBI, M.H., PATEL, R.S., BRAMMER, L., SMOLINSKI, M.S. and BRILLIANT, L. (2009) Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1014.

GREEN, M.J., ARCHER, S.C., BREEN, J.E., DAVIES, P.L., DOWN, P.M., EMES, R.D., GREEN, L.E., HUDSON, C.D., HUXLEY, J.N., LEIGH, J.L. and BRADLEY, A.J. (2016) Predictive biology: The future for mastitis control? In Proceedings of the World Buiatrics Congress 2016. pp 69–71.

HERMANS, K., OPSOMER, G., WAGEMAN, W., MOERMAN, S., DE KOSTER, J., VAN EETEVELDE, M., VAN RANST, B. and HOSTENS, M. (2018) Interpretation and visualization of data in dairy herds. *In Practice* **In Press**.

HERMANS, K., WAEGEMAN, W., OPSOMER, G., RANST, B.V., KOSTER, J.D., EETVELDE, M.V. and HOSTENS, M. (2017) Novel approaches to assess the quality of fertility data stored in dairy herd management software. *Journal of Dairy Science* **100**, 4078–4089.

HEWITT, S., GREEN, M.J. and HUDSON, C.D. (2018) Evaluation of key performance indicators to monitor performance in beef herds. *Livestock* **In Press**.

HOLMAN, A., THOMPSON, J., ROUTLY, J.E., CAMERON, J., JONES, D.N., GROVE-WHITE, D., SMITH, R.F. and DOBSON, H. (2011) Comparison of oestrus detection methods in dairy cattle. *Veterinary Record* **169**, 47–52.

HUDSON, C. (2015) Big data and the dairy cow: Factors affecting fertility in UK herds. http://eprints.nottingham.ac.uk/28896/1/Hudson%20PhD%202015.pdf.

LEACH, K.A., WHAY, H.R., MAGGS, C.M., BARKER, Z.E., PAUL, E.S., BELL, A.K. and MAIN, D.C.J. (2010) Working towards a reduction in cattle lameness: 1. Understanding barriers to lameness control on dairy farms. *Research in Veterinary Science* **89**, 311–317.

LIMA, E., HOPKINS, T., GURNEY, E., SHORTALL, O., LOVATT, F., DAVIES, P., WILLIAMSON, G. and KALER, J. (2018) Drivers for precision livestock technology adoption: A study of factors associated with adoption of electronic identification technology by commercial sheep farmers in England and Wales. *PLOS ONE* **13**, e0190489.

MARCEAU, A., MADOUASSE, A., LEHÉBEL, A., SCHAIK, G. van, VELDHUIS, A., STEDE, Y.V. der and FOURICHON, C. (2014) Can routinely recorded reproductive events be used as indicators of disease emergence in dairy cattle? An evaluation of 5 indicators during the emergence of bluetongue virus in France in 2007 and 2008. *Journal of Dairy Science* **97**, 6135–6150.

RUTTEN, C.J., VELTHUIS, A.G.J., STEENEVELD, W. and HOGEVEEN, H. (2013) Invited review: sensors to support health management on dairy farms. *Journal of Dairy Science* **96**, 1928–1952.

STATHAM, J.M.E., ARCHER, S.C., BIGGS, A.M., BRADLEY, A.J., BREEN, J.E., BURNELL, M., COOPER, R.L., DAVIES, P., DOWN, P.M., GREEN, M.J., HAYTON, A., HUDSON, C.D., HUSBAND, J., HUXLEY, J.N., KERBY, M., MAY, W., MAXWELL, O., RANDALL, L., READER, J., REMNANT, J.G., THORNE, M. and WAPENAAR, W. (2013) Future veterinary business models. *Cattle Practice* **21**, 78–87.

STEENEVELD, W. and HOGEVEEN, H. (2015) Characterization of Dutch dairy farms using sensor systems for cow management. *Journal of Dairy Science* **98**, 709–717.

TIEN, J.M. (2013) Big Data: Unleashing information. *Journal of Systems Science and Systems Engineering* **22**, 127–151.

VAN ASSELDONK, M.A.P.M., HUIRNE, R.B.M., DIJKHUIZEN, A.A. and BEULENS, A.J.M. (1999) Dynamic programming to determine optimum investments in information technology on dairy farms. *Agricultural Systems* **62**, 17–28.

VAN NUFFEL, A., ZWERTVAEGHER, I., VAN WEYENBERG, S., PASTELL, M., THORUP, V.M., BAHR, C., SONCK, B. and SAEYS, W. (2015) Lameness Detection in Dairy Cows: Part 2. Use of Sensors to Automatically Register Changes in Locomotion or Behavior. *Animals* **5**, 861–885.

WOLFERT, S., GE, L., VERDOUW, C. and BOGAARDT, M.-J. (2017) Big Data in Smart Farming – A review. *Agricultural Systems* **153**, 69–80.

ZHONGFU, S., KEMING, D., FEIXIANG, Z. and SHOUYI, Y. (2013) Perspectives of research and application of big data on smart agriculture. *Journal of Agricultural Science and Technology (Beijing)* **15**, 63–71.

Figures

Figure 1 The "Vs" of big data

Examples of application of the four "Vs" (a commonly used concept to define big data) to a dairy herd. [1] Hudson (2015); [2]MRO: milk recording organisation; [3]BCMS: British Cattle Movement Scheme

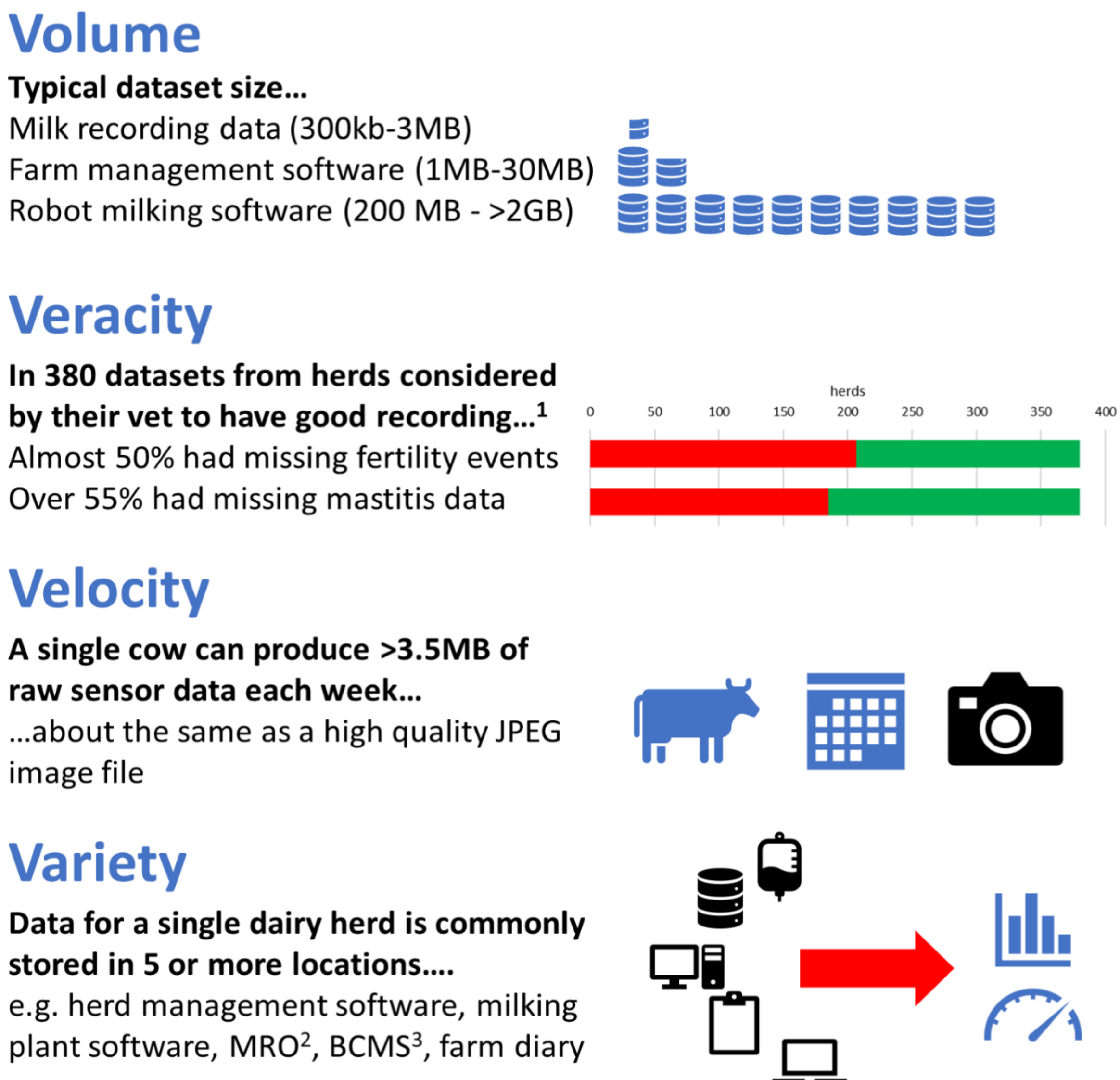

Figure 2 A data analysis pipeline

An example of the steps in data processing which occur between a sensor and the end user, based on an activity monitoring system used to detect oestrus.

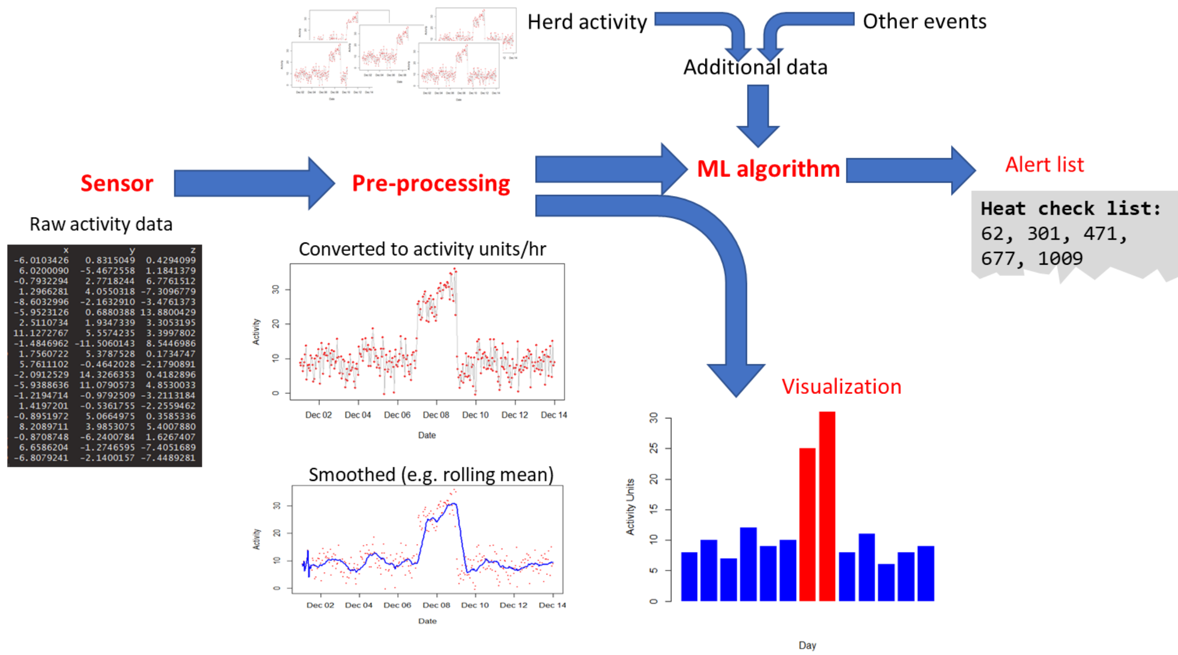

Figure 3 Example of data flow on a dairy herd

An example of the flow of data between different stores and sources of data on a typical dairy herd.