DATA OWNERSHIP, ACCESSIBILITY AND STRUCTURE CONSIDERATIONS AFFECTING USABILITY BY RECORDING ORGANISATIONS

Steven Sievert

Technical Director, National DHIA & Manager, Quality Certification Services

Chair, ICAR Subcommittee for Measuring, Recording and Sampling Devices
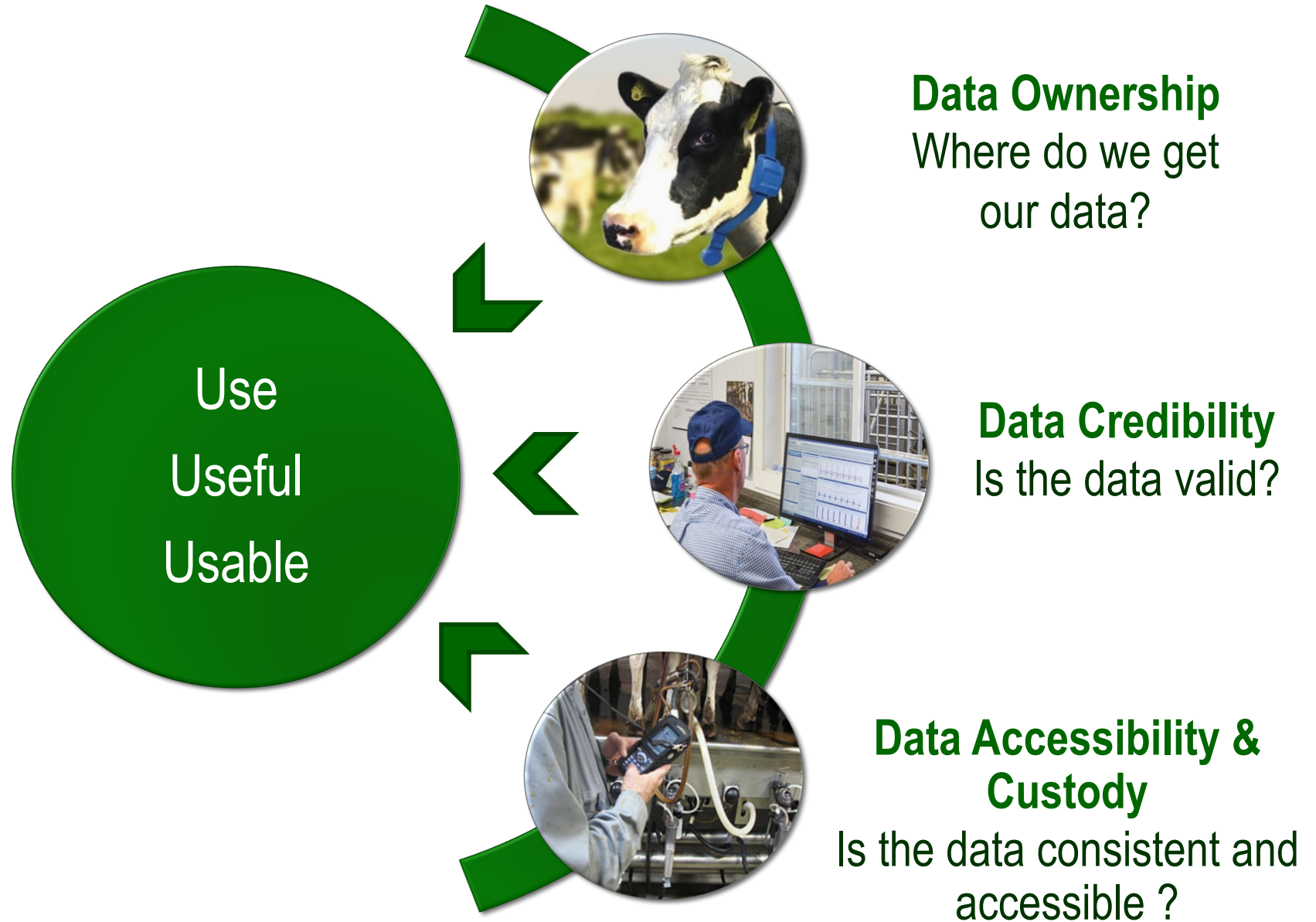
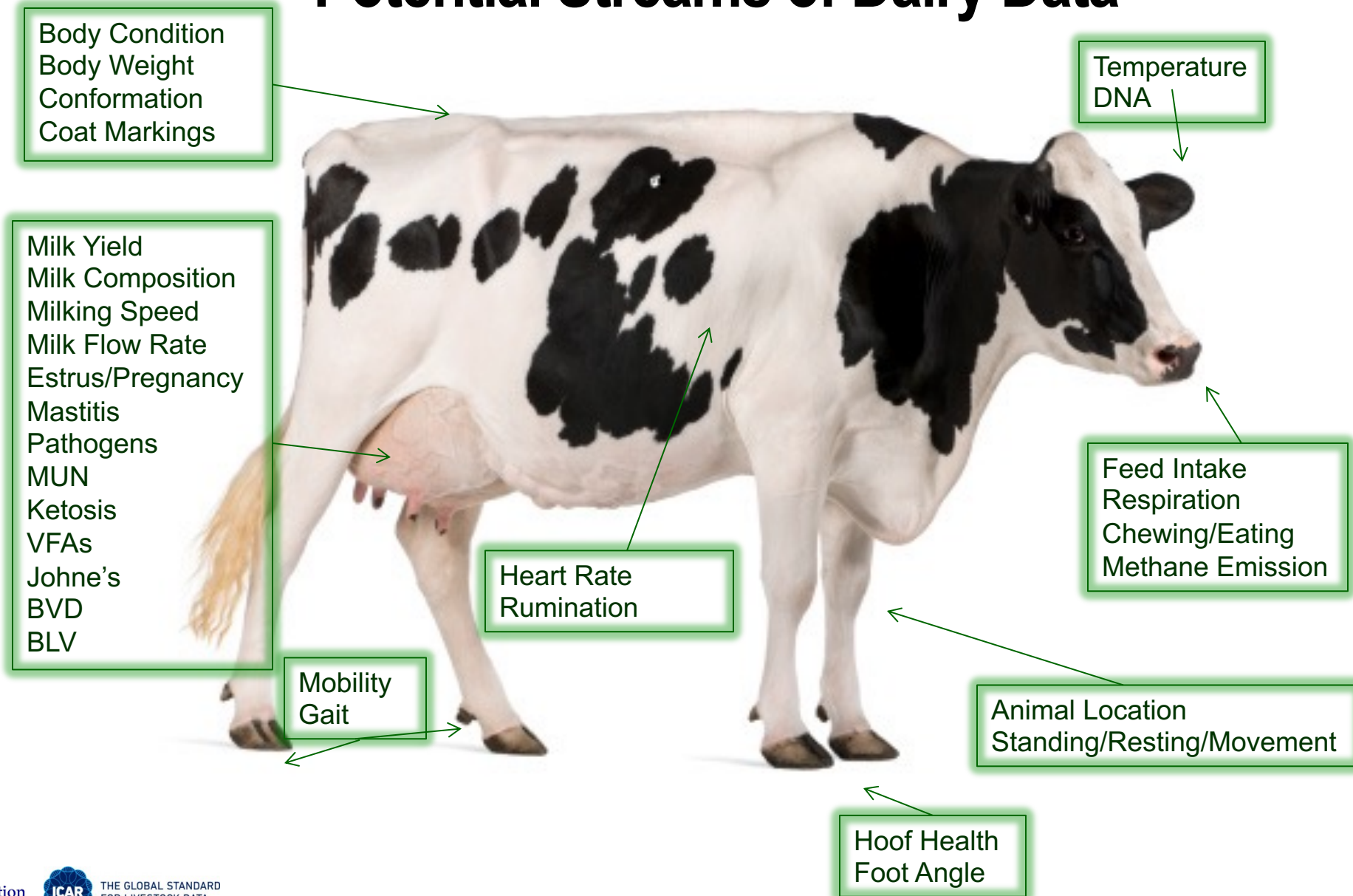*Board of Directors, American Guernsey Association*

Red Wing, Minnesota, USA

Challenges with Dairy Herd Data

The V's of Data

Volume
Velocity
Variety
Variability
Validity

Use
Useful
Usable

Data Ownership
Where do we get our data?

Data Credibility
Is the data valid?

Data Accessibility & Custody
Is the data consistent and accessible ?

Potential Streams of Dairy Data

Body Condition
Body Weight
Conformation
Coat Markings

Temperature
DNA

Milk Yield
Milk Composition
Milking Speed
Milk Flow Rate
Estrus/Pregnancy
Mastitis
Pathogens
MUN
Ketosis
VFAs
Johne's
BVD
BLV

Heart Rate
Rumination

Feed Intake
Respiration
Chewing/Eating
Methane Emission

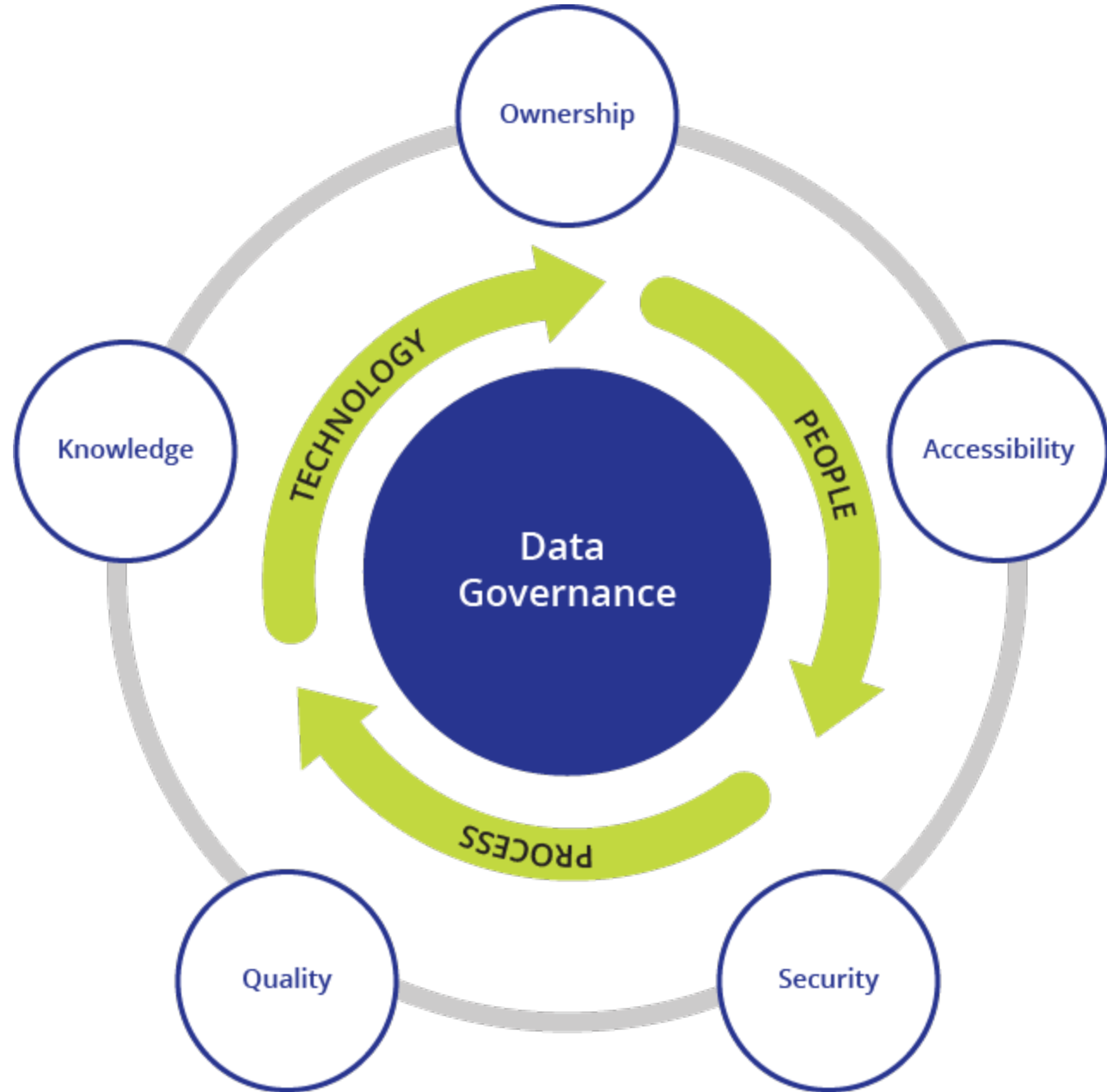Mobility
Gait

Animal Location
Standing/Resting/Movement

Hoof Health
Foot Angle

Data Governance
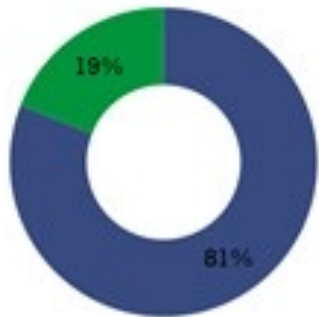
A Shared Producer - Industry Responsibility
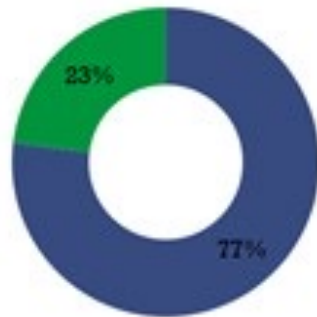
Ownership of Dairy Farm Data?

What do Dairy Producers Believe?

- Vast majority of dairy producers believe that they should own and control their data

81% think they own their data and 77% are worried about data security
(September 2014 AFBF Survey)

19% / 81% (Yes / No)

23% / 77% (Yes / No)

What do ATPs & Industry Organizations Believe?

- Agriculture Technology Providers (ATPs) may state the farmer owns the data but there are limits, use considerations and intellectual property involved.

- Many Industry Organizations believe the dairy producer owns the data and support the premise that the producer should control sharing of some or all of the dairy farm data.

Realities of Dairy Farm Data

- **Varied legal rulings or opinions (and ongoing cases) related to ownership of farm or agricultural data in many countries. The purpose of this discussion is related to understanding challenges with credibility and accessibility of dairy farm data, not an opinion on the ownership.**

- With modern data collection, ATPs provide framework (computations, aggregation, storage, and sharing) of dairy farm data. This framework may contain IP and have costs associated with data movement.

- The definition of what data is owned by whom loses clarity as the data moves further away from the dairy.

- Data ownership could be defined as pre- and post-algorithm (native data vs. transformed data).

- EULAs with ATPs maybe be confusing, lack clarity, or misunderstood by dairy producers.

- Concerns from dairy farmers exist about the privacy, security, accessibility, sharing and accuracy of data as it moves off the farm operation.

- Concerns related to third-party use and/or monetisation of dairy producer data exist.

- Dairy Producers still have the primary responsibility for the accuracy and quality of data when sharing off the farm.

Data Accessibility

## Our Data Ingestion & Aggregation Paths Will Change and Vary

Challenges

- Agreements with manufacturers to access cloud-based data
- Agreements with data exchange hubs, where applicable
- Direct relationship with dairy producers to ensure access to data is granted
- Direct relationship with dairy producers to address data quality issues
- Data flow interruptions or presence of non-usable data in the data stream

Realities

- Each organisation or country has a traditional flow of data into herdbooks or databases
- Modification may be uncomfortable and perhaps costly but will be essential
- Data for different parameters may come from different sources

Collaboration is Key in U.S. Dairy Genetics
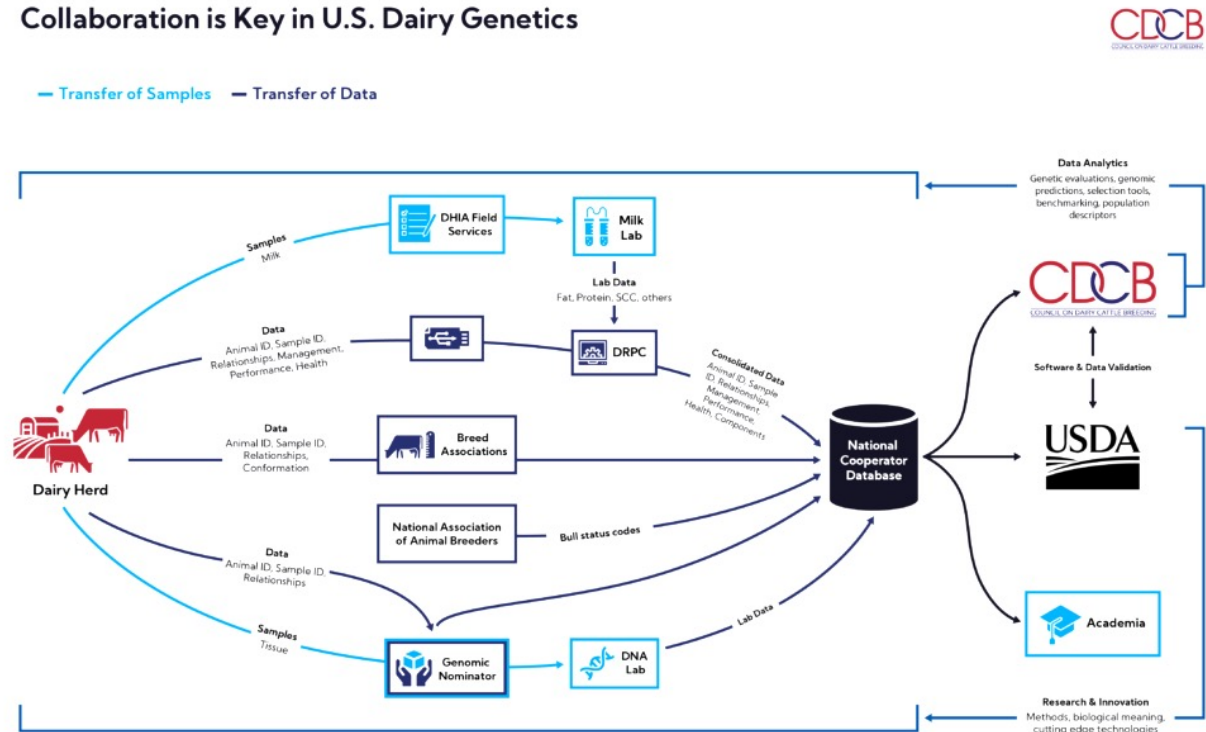
Uses of Data

With Different Needs for Accuracy & Precision

Management Data
- Yield
- Milking Speed
- Feed Efficiency
- SCC

Health Data
- Locomotion
- Reproduction
- Disease
- BCS/Weight

Welfare Data
- Activity
- Mobility
- Eating, Resting
- Heat Stress

Data for Genetic Evaluations & Herdbooks

Data Linked to Direct Farm Payments
- Yield
- Fat, Protein
- SCC

Alarm Data
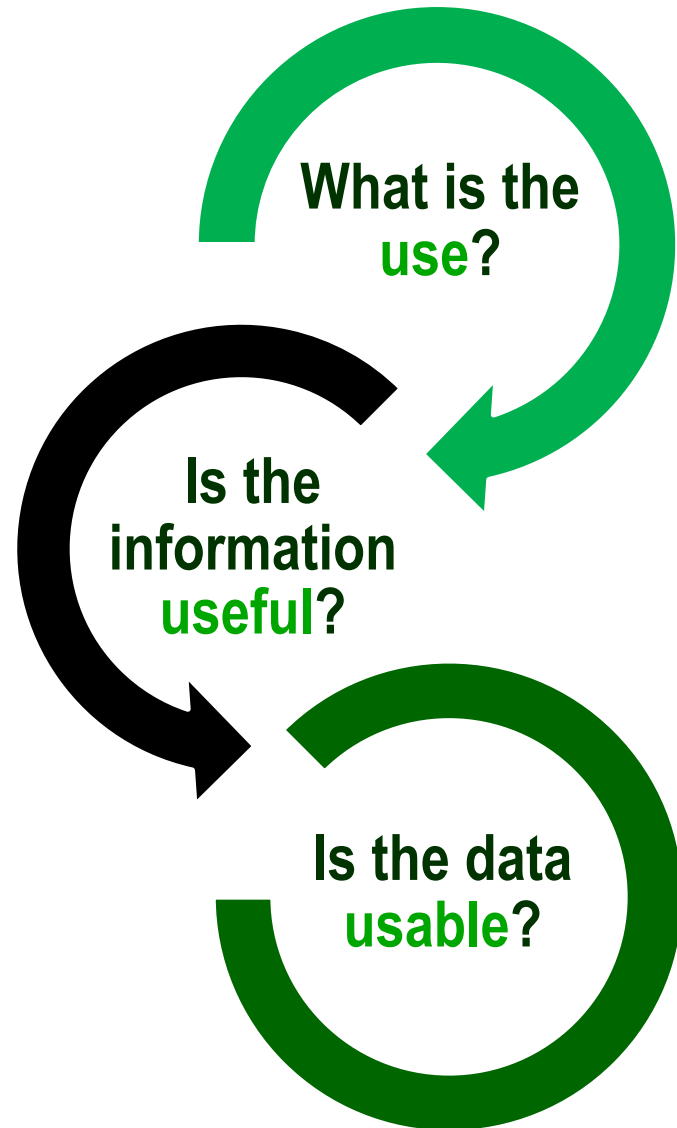- Heat Detection
- SCC
- Locomotion
- Location

Yes/No Data
- Pregnancy
- Disease

Trend Data
- BCS/Weight
- Milking Speed
- Feed Efficiency
- Activity

Decision Model for Data Usability

What is the **use**?

Is the information **useful**?

Is the data **usable**?

Usability may change based on the intended use and the credibility & quality of the dairy herd data

Data Credibility & Quality Affects Data Usability

Two Essential
Components of
Credible & Quality Data

Accuracy of the Identification System

Accuracy of the Measurement System

Animal ID is Not Equal

Animal ID is Not Easy

Simultaneous recording of animal ID and data measurement

- The 'official ID' of an animal most likely will not be the same as ID associated with measurements
- Animals may have multiple IDs for data flow over their lifetime
- Animals may have multiple IDs on their body or in the data flow/computer system(s) at once

Can we answer these questions?

- Do we have protocols for ID cross-referencing and validation?
- Do we have protocols for on-farm validation of the ID system & for data transfer/custody from farm management software to external users?

Accurate Animal ID is the primary responsibility of the producer and Accuracy of the ID data as it moves through the system is the shared responsibility of all industry stakeholders

Sources of Identification Errors in Dairy Data

ID Association
- Wrong live animal ID with data source ID
- Wrong animal ID with service sire or embryo ID
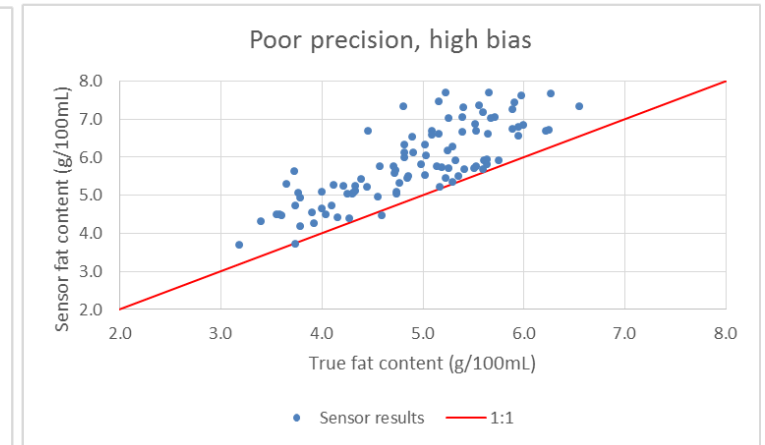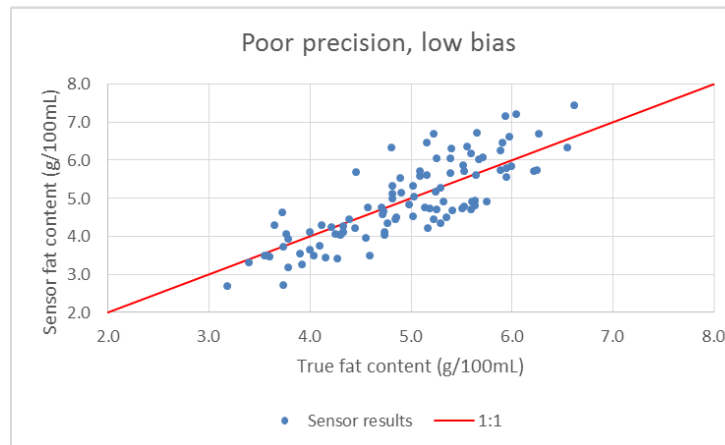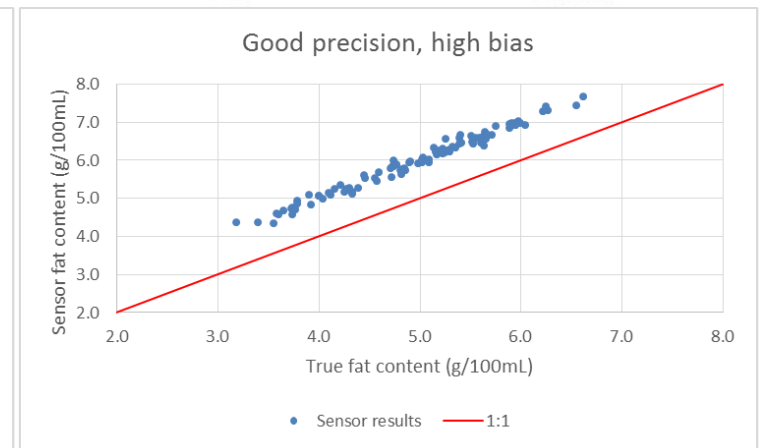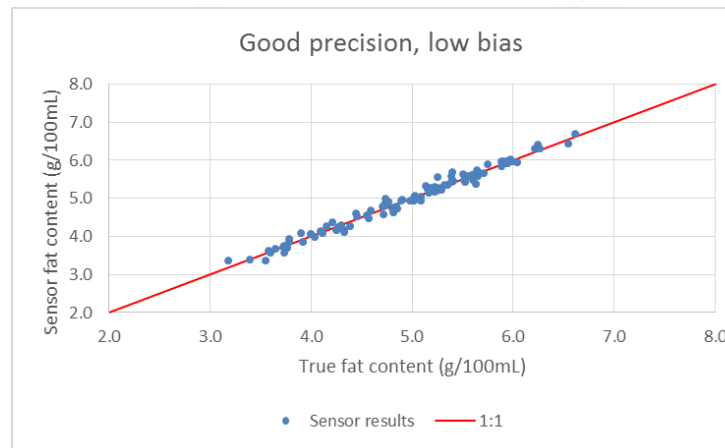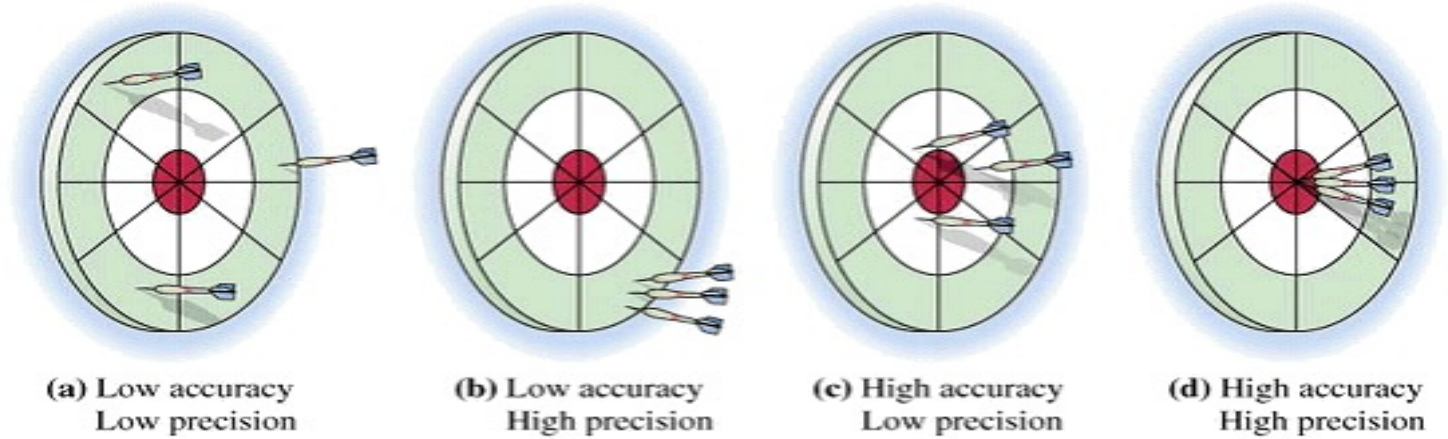- Override of ID corrections by other software systems

ID Linkage or Cross-Reference
- Wrong sensor ID with official animal ID
- Animal ID with DNA or Milk Sample ID
- Replacement ID with Original ID (tag replacement)

ID Positional Errors
- Missing or unreadable ID creating shifts in data
- Cows out of order after ID reading
- Samples out of order in DNA or milk laboratory

DHIA

Quality Certification Services Inc.

ICAR THE GLOBAL STANDARD FOR LIVESTOCK DATA

Accuracy of the Measurement System

(a) Low accuracy Low precision

(b) Low accuracy High precision

(c) High accuracy Low precision

(d) High accuracy High precision

Good precision, low bias

Good precision, high bias

Poor precision, low bias

Poor precision, high bias

Sensor fat content (g/100mL) vs True fat content (g/100mL)

Sensor results — 1:1

Accuracy of the Measurement System

Testing & Certification or Validation

Certification of Recording Devices

- Standards and Guidelines for testing
- Direct relationships with manufacturers
- Assurance of device measurement accuracy

Validation of Sensor Systems

- System-based approach that covers eight major areas including parameter measurement, data handling, ID systems
- Understanding of the system from both operational and usability perspective

Routine Calibration and Performance Procedures

- Assurance of ongoing data quality as new data is sent to industry stakeholders

Example of New Technologies - Conformation Estimates

Opportunities for new data for the industry

- ICAR has guidelines and processes for system testing and validation
- Cooperative process that needs input and support from all stakeholders
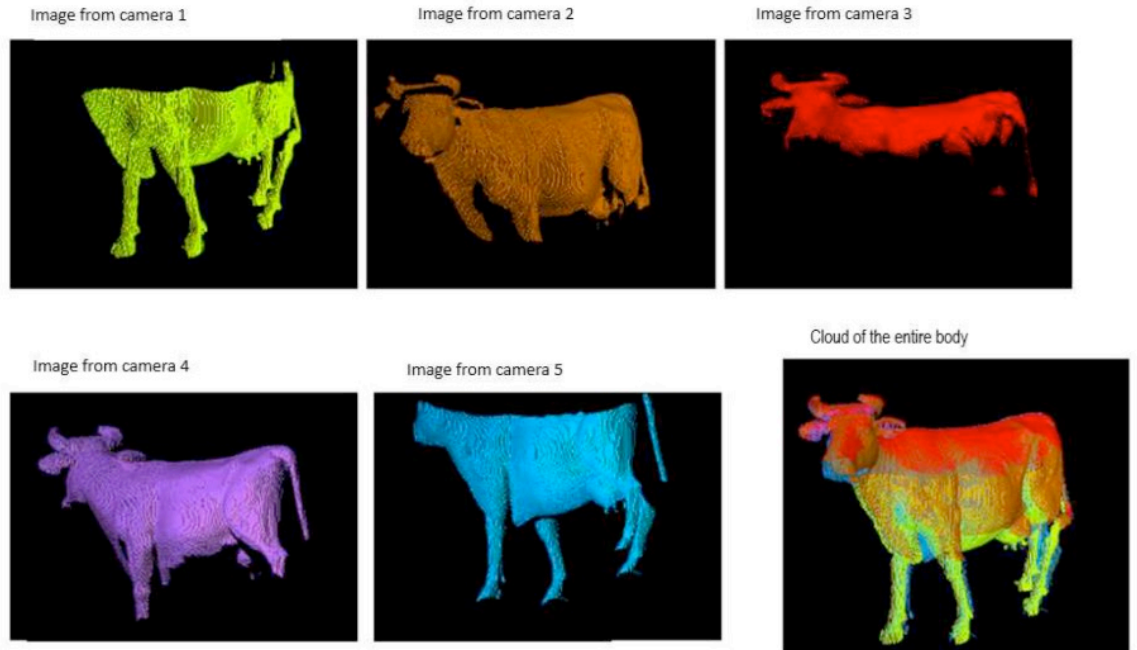
What are We Measuring?

Measured Parameter is often Different than the Reported Parameter

Multiple Indicators of Mastitis or Milk Quality

- Automated CMT/WMT
- Electrical conductivity
- L-lactate dehydrogenase
- N-acetyl-beta-D-glucosaminidase
- ATP luminescence
- Thermal imaging
- Visible, NIR, MIR spectroscopy

SCC (x1000)

Central Lab
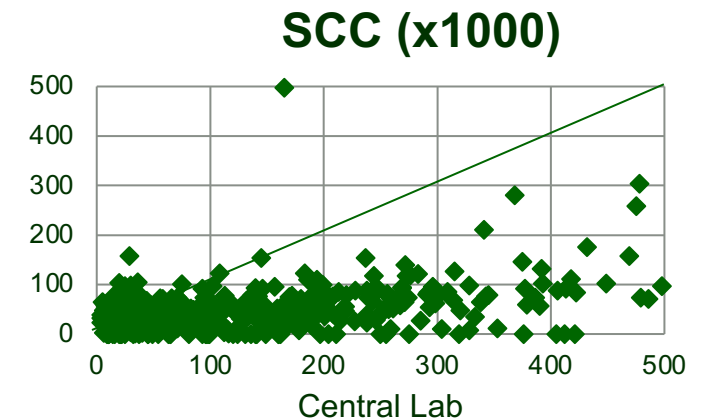
All of these system measurements report SCC results, each with different accuracy and precision

The Need to Understand the Parameter Measured
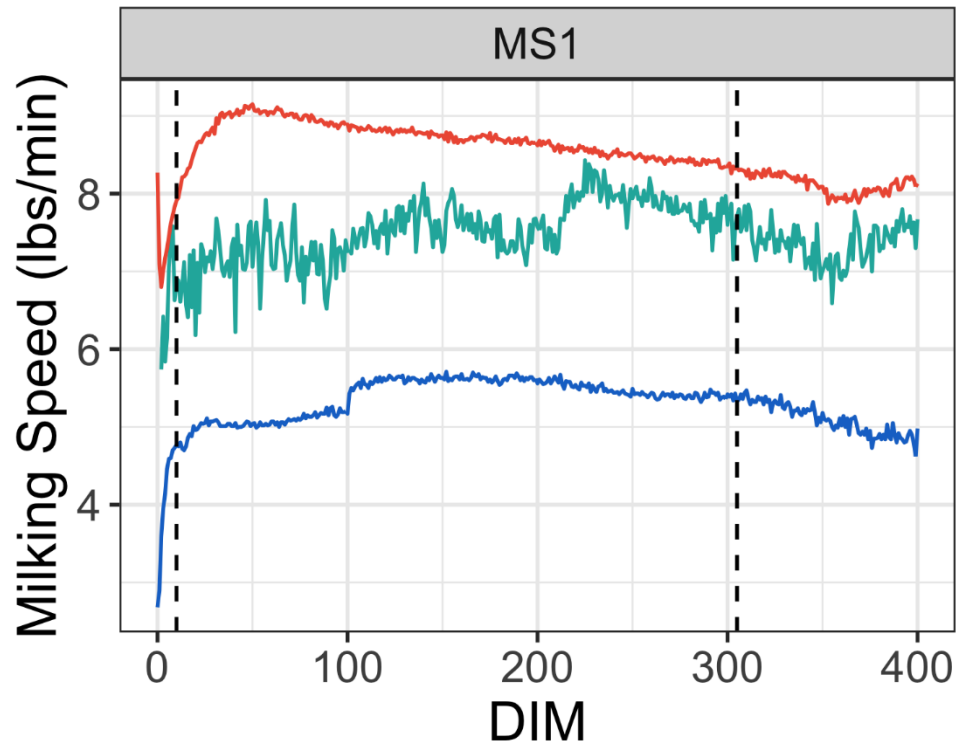
Milking Speed by OEM Manufacturer (Holsteins only)

Example of Different Data Definitions

- Milking speed, as reported in data output, is a different measurement for various manufacturers.

- This does not affect the usability of the data, rather demonstrates the need to understand data available

- Standardisation is possible and data usability increases.

- Complete Industry Stakeholder Initiative in USA

- ICAR Joint Group and Animal Data Exchange (ADE-WG) is working towards standardising parameter definitions

Data Handling and Validation Questions Worth Asking

Data Handling

- **Handling of missing data points**
  - Estimated data included?
  - Means of actual data only?
- **Data smoothing**
  - Rolling means
  - Other data manipulations
- **Data packaging**
  - Single observations
  - Multi-day means
  - Projections

Data Validation

- **Range of accurate measurement for sensor**

- **Decision rules for handling and/or exclusion of outliers**

- **Range(s) of biological measurements**

- **Evaluation of algorithm or machine learning principles involved in calculations**

Data Quality and Usability

A Balanced Approach

- System Quality Validation and Recording the Source of Data Measurements
- Focus on Accessibility and Ingestion of Data from Farm Management Software instead of Exclusion
- Develop Routine Quality Monitoring Tools to Assure Consistent and Credible Data
- Focus on Data Use and Usability for Specific Needs Instead of All or Nothing Approach
- Removal of Incomplete Data Instead of Editing the Data to Meet Standards or Improve Quality

Points to Ponder

- Attention to all components of **Data Governance** is essential – Ownership, Accessibility, Security, Quality & Knowledge

- **Accurate Animal ID** is the primary responsibility of the producer and Accuracy of the ID data as it moves through the system is the shared responsibility of all industry stakeholders

- **Opportunity to merge dairy herd data** from various sources together when using validated and quality system approach

- **Data credibility** continues to be the driver of all industry stakeholders, only the opportunities and challenges have changed.