Dairy Brain – Informing Decisions on Dairy Farms using Data Analytics

Michael C. Ferris
(based on collaboration with A. Christensen and S. Wangen)

Jacques-Louis Lions Chair and John P. Morgridge Professor of Computer Science
Computer Sciences Department and Wisconsin Institute for Discovery, University of Wisconsin, Madison

2020 Dairy Cattle Reproduction Council (DCRC) Annual Meeting

November 11, 2020

Decision Support Tools

- Aim to exploit data streams from farm and other economic, health and agricultural sources
- Descriptive: Use collected information for reporting (financial and medical apps)
- Predictive: Apply models to forecast future events (weather and air quality apps)
- Prescriptive: Increase sophistication of analysis to evaluate which decisions lead to desired outcomes (resilient electricity dispatch, traffic routing)
- At single cow or farm level

Successful data analytics: some features

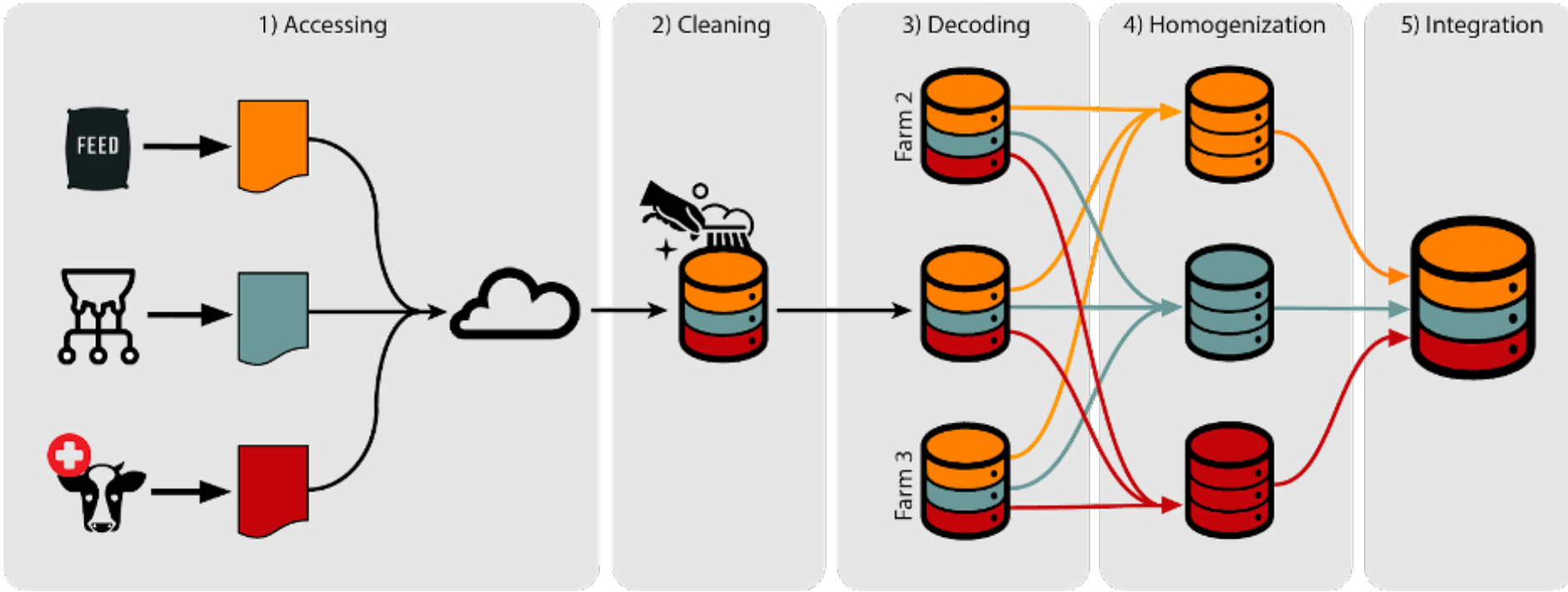
- Weather:
  - large scale, real time
  - open source/access
  - no private information (but apps that present information differently)
  - data provider is not the same as user
- Medical
  - shared/private information
  - multiple data types
  - recommender apps (diagnosis, treatments)
- Travel:
  - links different types of agents (drivers, riders, administrators)
  - real time, large scale
  - congestion pricing (public/summary information)
  - trips (private information)
  - required (user) inputs to generate specific user outputs
- Financial:
  - standards for interconnectivity (transfers)

All have reliable acquisition. Need to name things consistently.

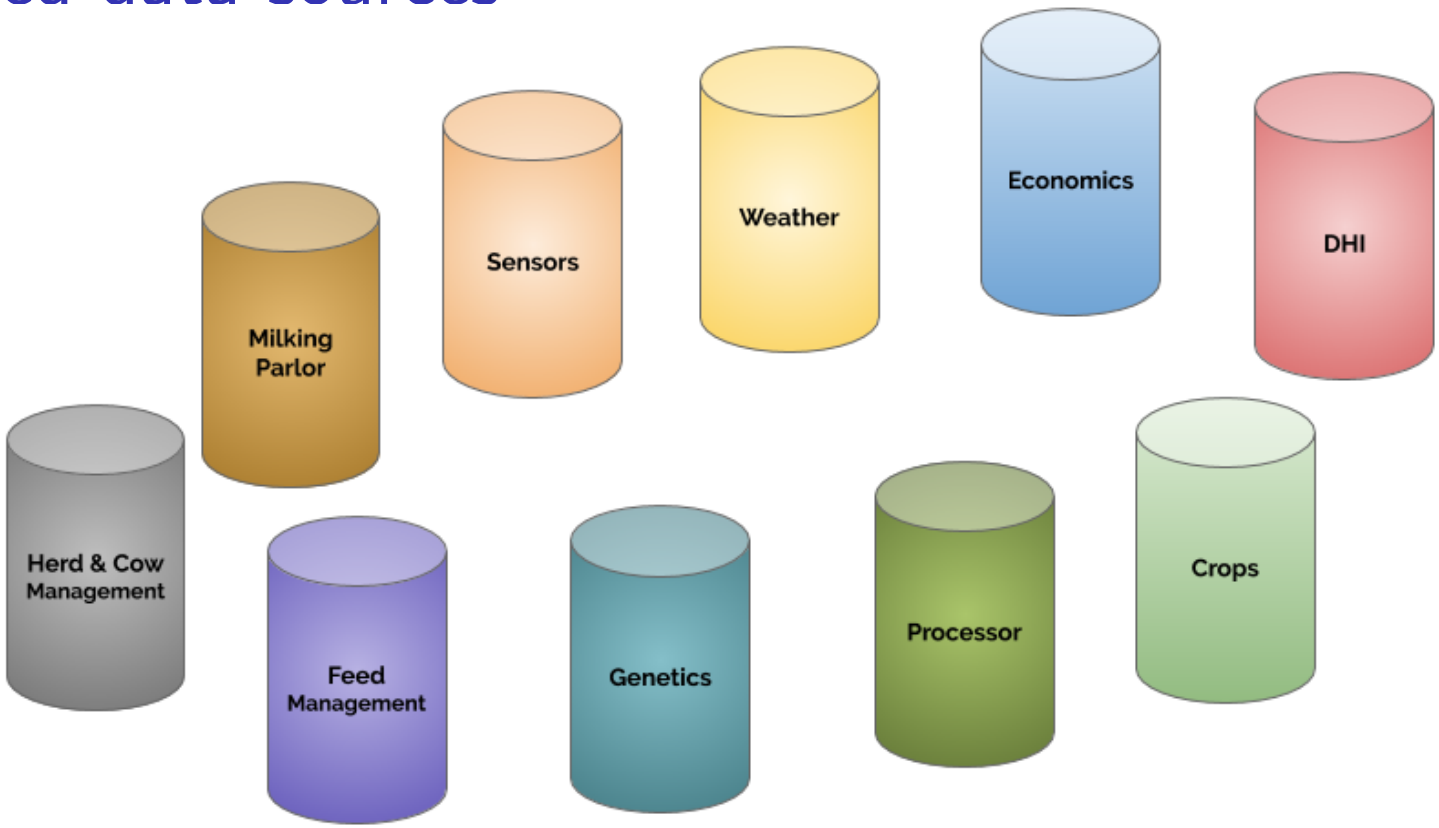
Issues regarding what to do for who?

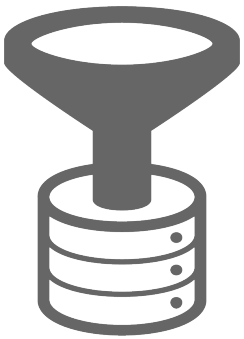
- Policy or individual farm?
- Operational (logistic, pen mgmt,) or strategic (capital expansion, pricing, culling)?
- When are decisions made: yearly, seasonal, daily, hourly?
- Inform human-in-the-loop decision making
- Ownership: whose data is it, after change/cleaning
- Privacy: who can see what and when
- Scale: the big data issue
- Missing data

The Data Setup



Monitored data sources

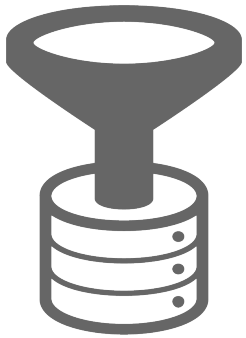




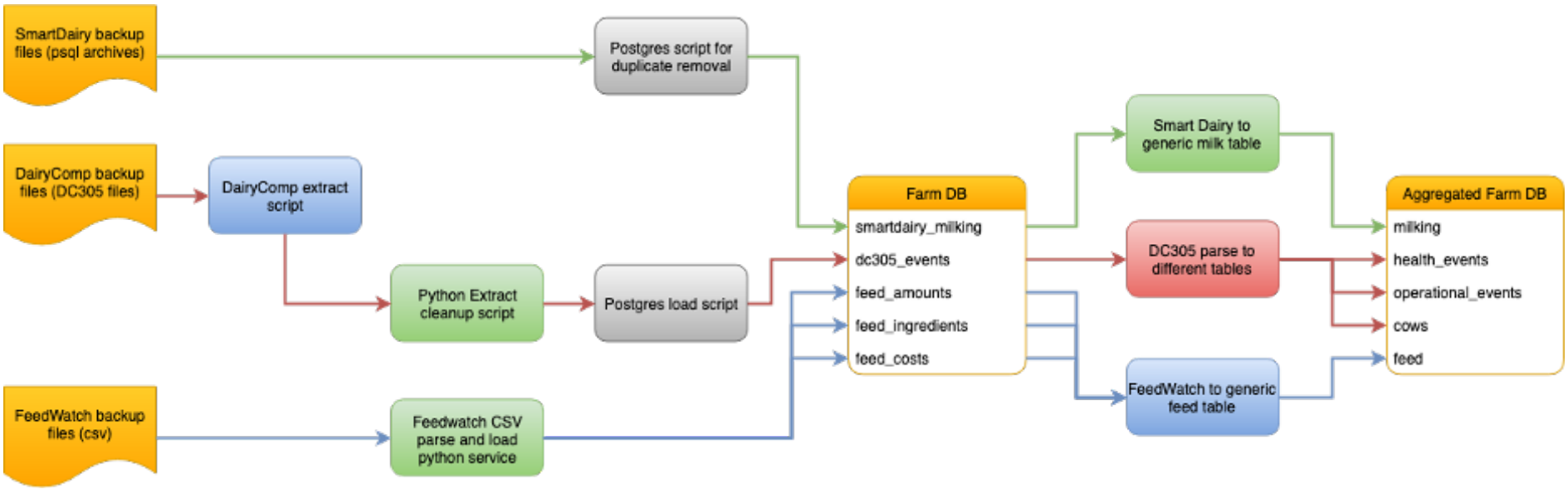
- Milk production
- Milk component analysis
- Milk spectral analysis
- Diet

- Activity
- Rumination
- Disease occurrence
- Management actions

Agricultural Data Hub (AgDH)



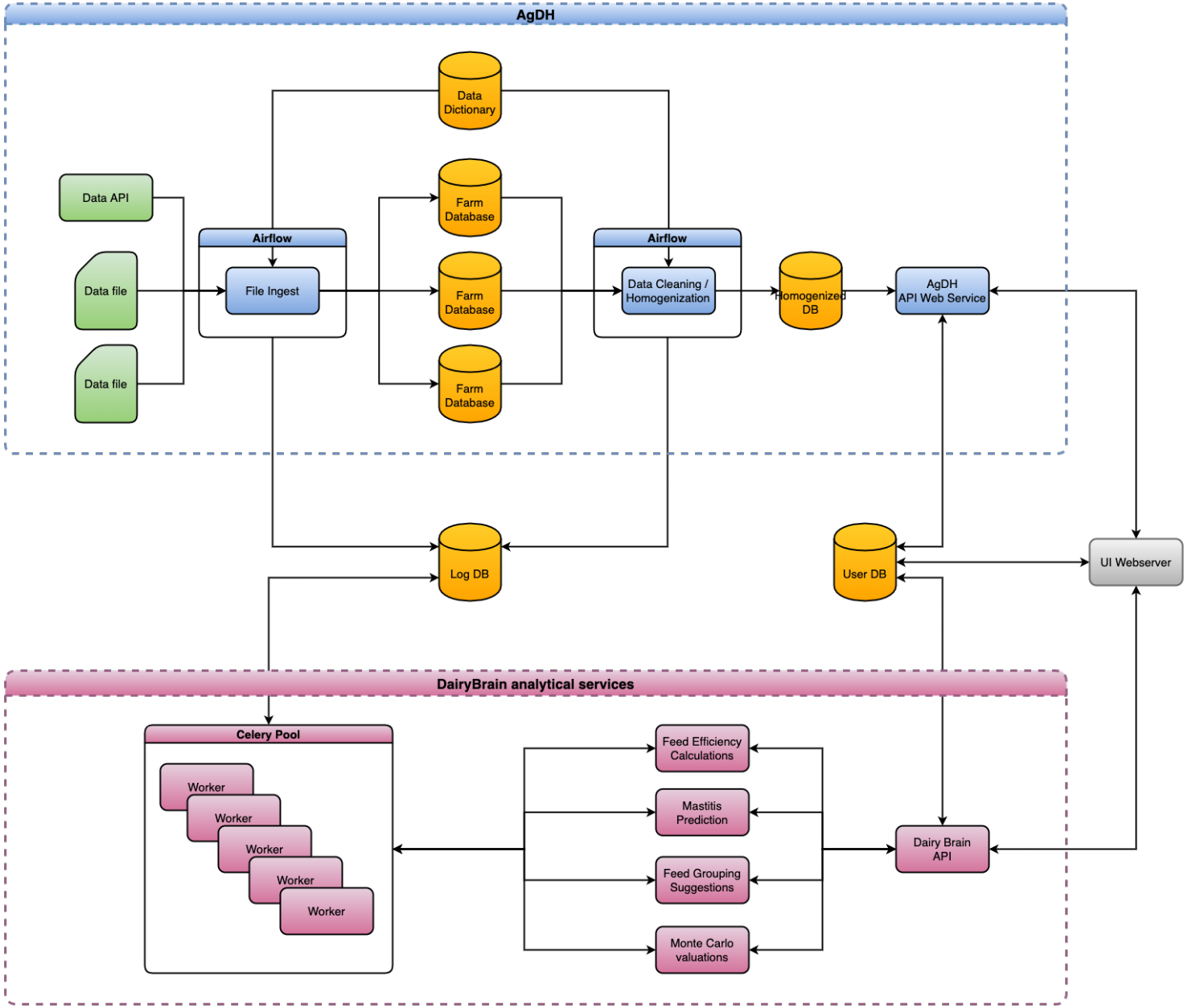
- Aggregate data from different sources
- Homogenize vendor-specific reporting
- Make data available for retrieval and analysis
- Entity matching (even for Cow ID)



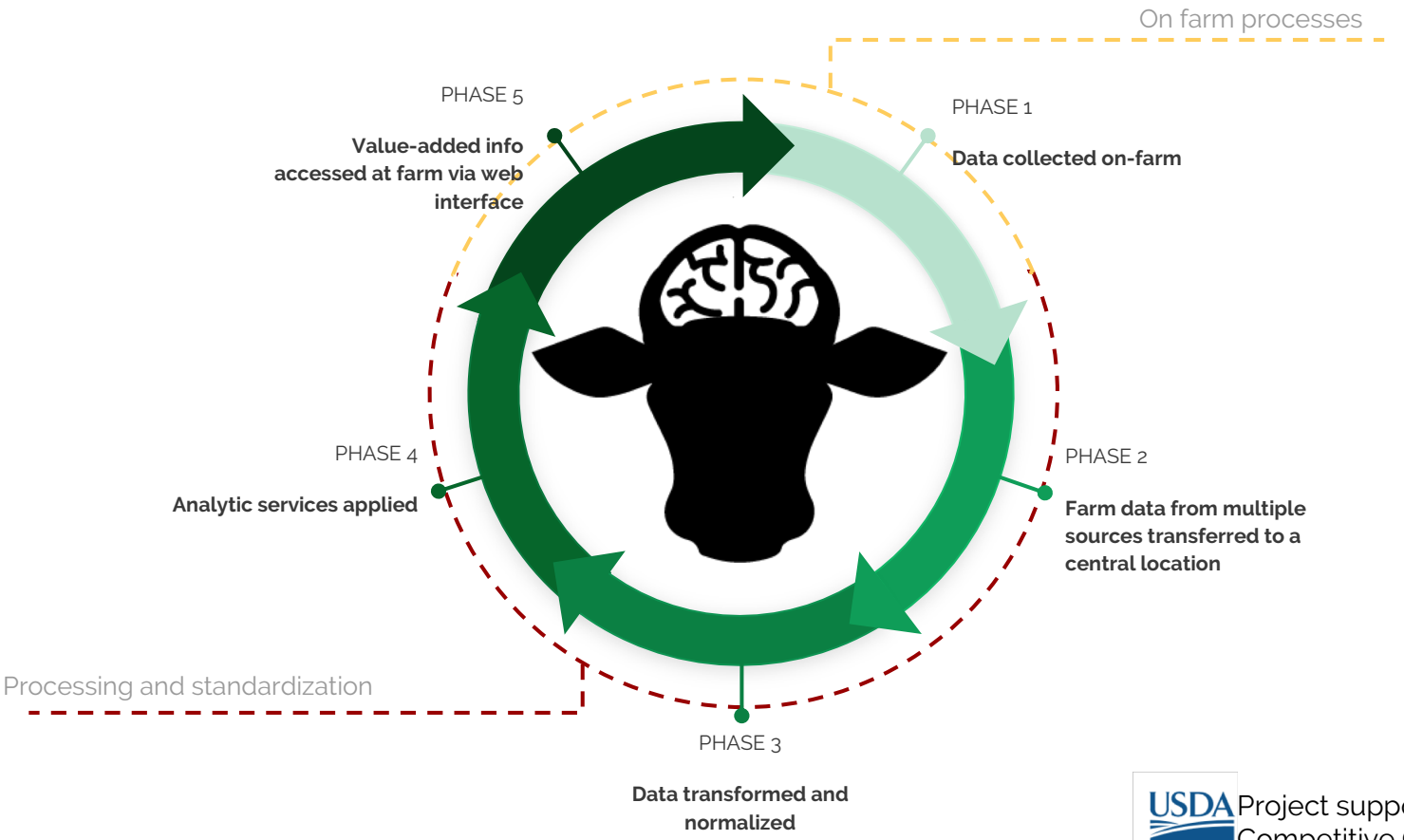
The model is only as good as the data

- Missing data: the norm rather than the exception (in farm data and many other settings)
- Unit-level (fail to take survey, handled via appropriate weighting) vs item non-response
- Listwise deletion (exclude whole item) or pairwise deletion (exclude only if missing in needed entry), mean imputation is ad-hoc, likely to be biassed and/or inefficient
- Newer and principled methods: multiple-imputation, full information maximum likelihood, expectation maximization, matrix completion (consider conditions under which missing data occurred, combine information with statistical assumptions)
- Database approach (dirty data): data cleaning via probabilistic inference, automatic repairs

Application Programming Interface (API) design



Dairy Brain - a continuous decision aiding engine

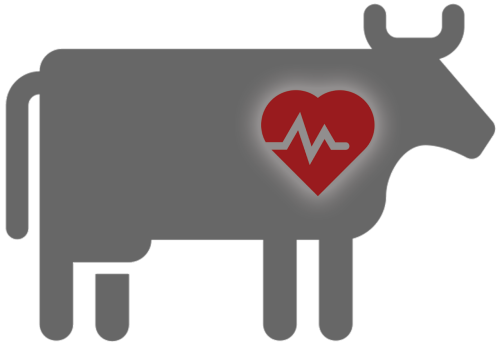


USDA Project supported by NIFA-AFRI-FACT Competitive **Grant No. 2018-09256**

- Translate research outcomes to practical applications
- Provide access to analytical services to enhance operations

Cow health

- Early ketosis identification
- Monitoring the Risk of CM for 1st Lactation Heifers
- Early Prediction of Clinical Mastitis

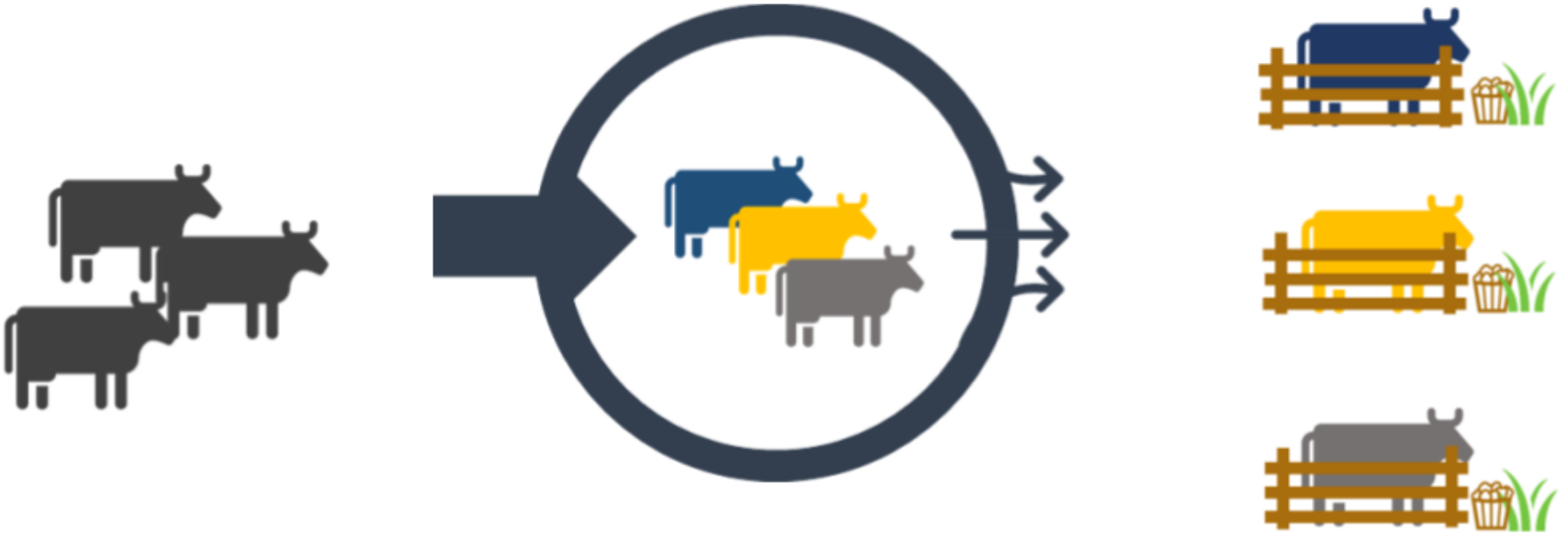


Nutritional grouping

Group of cows

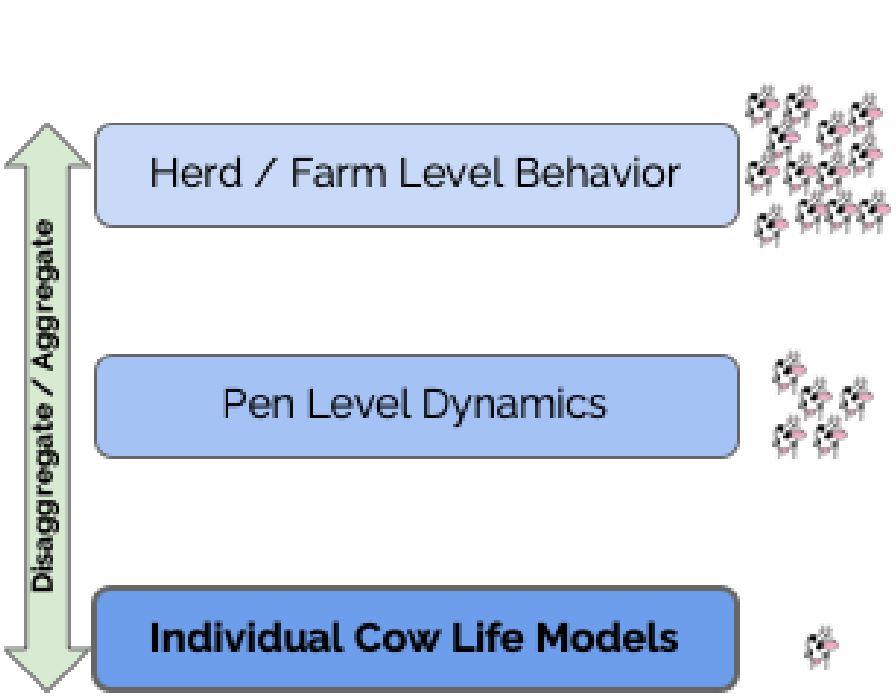
Cluster cows

Differentiated diet



Multiple applications at different scales

- Separately developed research products (as above)
- Connect via data and visualization APIs
- Utilize standard DS tools: feature selection, clustering, tensorflow, python, R, SQL, hadoop, deep learning, etc
- Conduit to translate research into (commercial) products



Scalable Decision Products

- Inter-farm performance comparisons
- Farm business performance evaluation
- Identify operational inefficiencies
- Cow group comparisons / optimization
- Sub-pen level cow movements / groups
- Individual cow valuation / herd expansion

Take home messages: Hoard's Dairyman (Feb-May 2020)

- Dairy Brain: Multiple data sources, multiple models (specialize to the question at hand)
- British proverb: Horses for courses (racehorse analogy)
- Establish a Coordinated Innovation Network (CIN)
- Ownership and security: Must determine data sharing policy/procedures
- Collection and communication: Must plan for missing data
- Farmer adoption: Must focus products/questions: be specific
- Business API's Must have API (standards) - think of this as bank transfers