Next Level Animal Sciences: Big Data in the livestock domain

Claudia Kamphuis & Annemarie Rebel
WUR - Animal Science Group – Wageningen Livestock Research

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Hi, my name is ....

Claudia Kamphuis
Married, mom of 2, and like to ride my R6, strolling through woods/mountains, true-crime podcasts/series

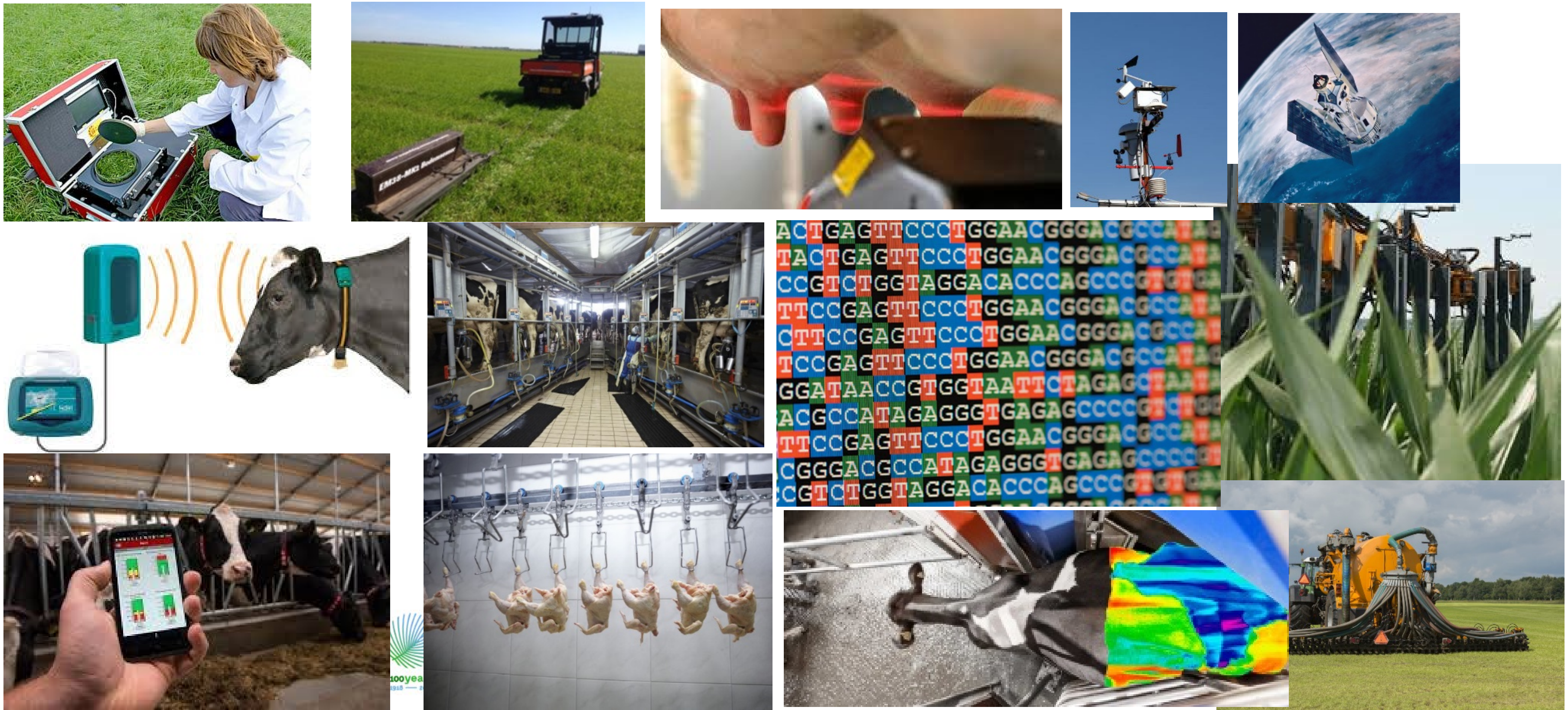
Animal Scientist (2004)
Preventive Animal Health & Welfare

PhD Utrecht University (2010)
Making sense of sensor data; using milk robot data to detect mastitis with machine learning approaches

Researcher WLR - ABG (2016)
Project leader of several Big Data projects

What is Big Data

Data diarrhoea due to increasingly tech-saffy livestock domain

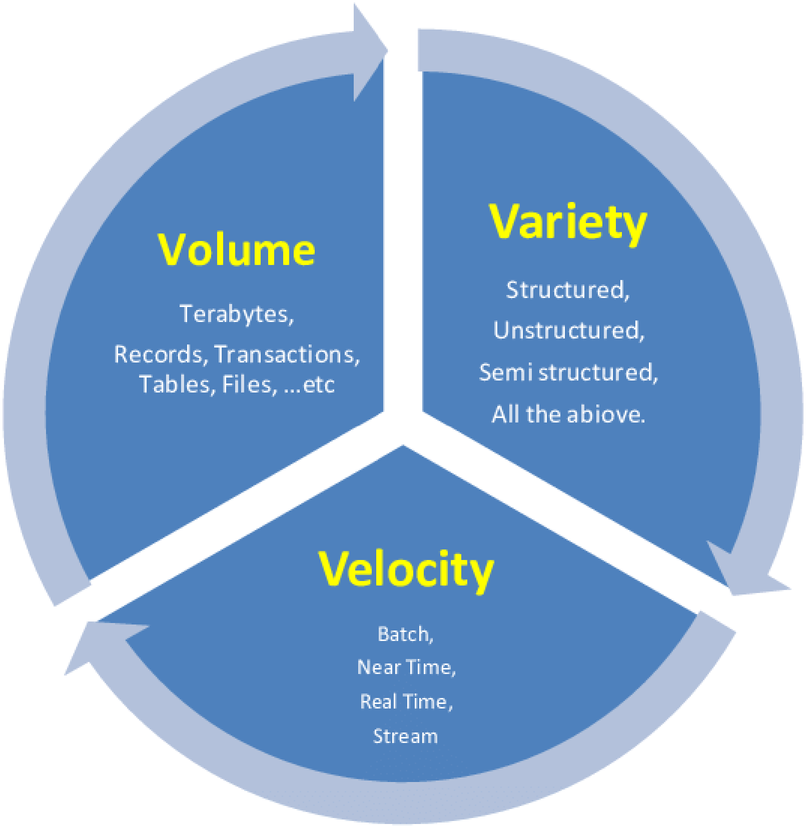


The V’s of Big Data

Data diarrhea due to increasingly tech-saffy livestock domain

The V’s because of it

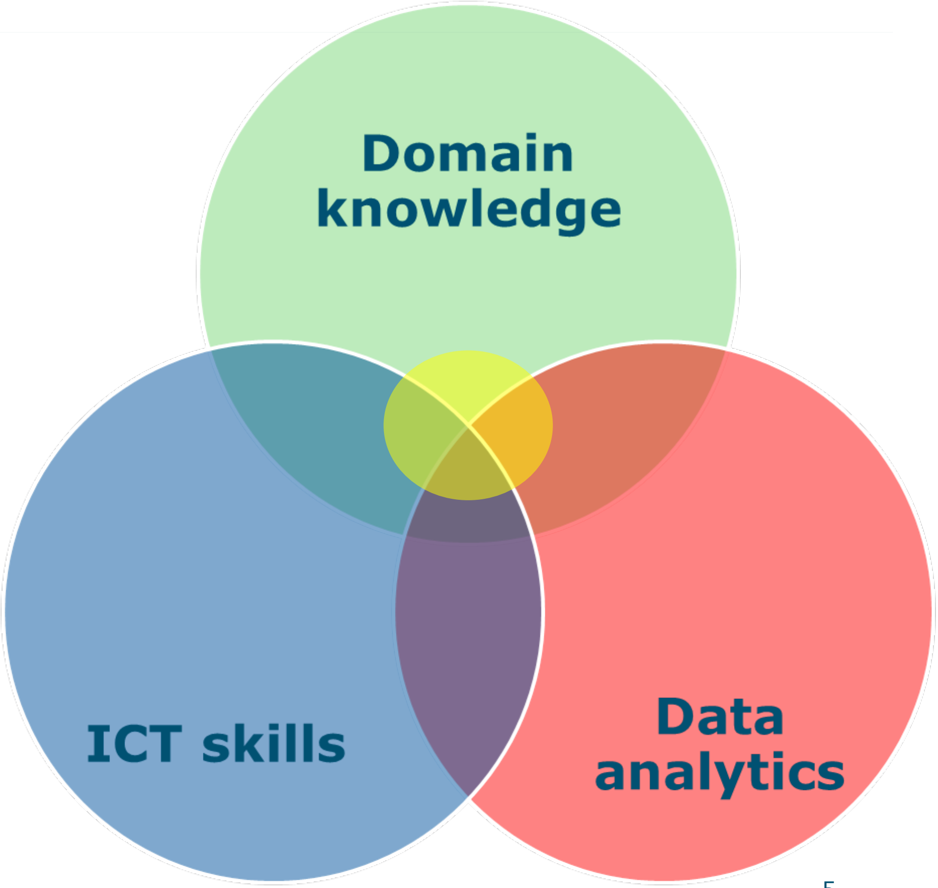
Not defining Big Data but **challenges** that we need to tackle so we can ‘do’ Big Data



Big Data @ WUR

Connect these three areas to face the challenges...

to start doing things that we could not do previously



1 Big Data in the broiler chain

2 Development of an automated disease scoring tool

3 Predicting resilience using sensor data & ML

4 Flexible data architectures in the cloud

Big Data in the Dutch broiler chain

Data are collected throughout the chain

a lot of data are collected at different stages, but limited in sharing and certainly not connected routinely throughout the chain

Broiler breeder
Egg production
Egg weight
Growth performance
Mortality
Antibiotic use

Hatchery
Hatchability
Egg storage time
Antibiotic use

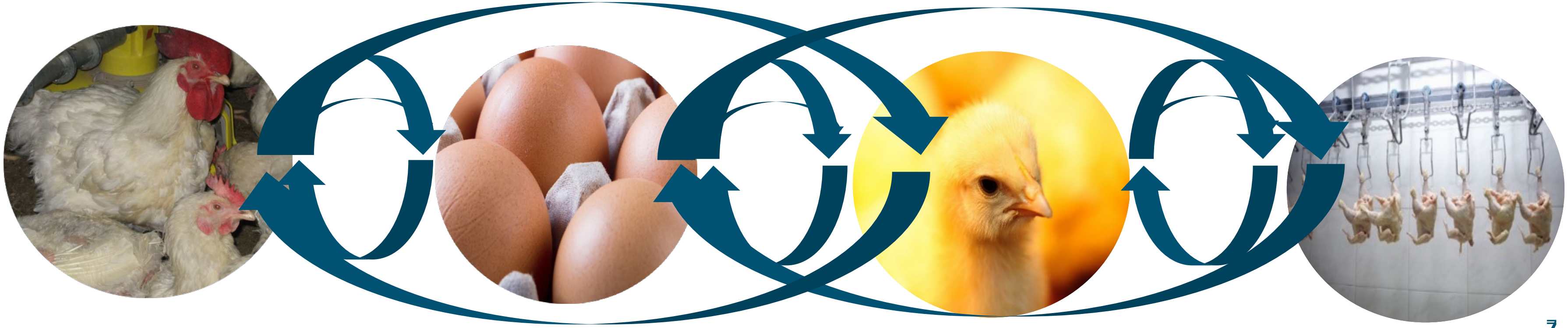
Broiler farm
Growth
Mortality
Antibiotic use

Meat processor
Carcass condemnation
Carcass uniformity
Footpad dermatitis scores

Big Data in the Dutch broiler chain

If so, systemic (repeatable) effects of one part of the chain on performance & health down the stream can be assessed

And improve chain performance as a whole



Big Data in the Dutch broiler chain

Identify existence of systemic effects of broiler breeder flock/farm on production/health performance of broiler farm/flock based on data throughout the chain

Fast-growing broiler breed, 6 years of data (2011-2016)

2 nutrition companies
1 breeding company
209 breeder flocks; 88 farms
National database

2 hatcheries

2174 broiler flocks;
74 farms

1 processing plant
2 locations

Big Data in the Dutch broiler chain

Idea was great and we were lucky with a good starting position

Collaboration already existed, partners owned data, all data were digitalized, and willingness to share it (anonymised)

But there were still some huge challenges

- Existing databases are not made for combining
- Data quality: lot of manual recording. Considerable manual effort to ‘solve’ impossible data
- Limited ‘additional’ data availability (and lack of energy in collecting this)

Biggest challenges for future work in Dutch situation is to overcome **fear for claims**

Main result

No systematic effects of breeder farm & only small systematic effects of breeder flock on performance and health indicators of the progeny

- Relatively small effect of breeder flock on rejection% (7%) and uniformity% (5%)
- Either the following chain phases have a relatively large effect, that overrule the effects of breeder flock or farm, or the effects are relatively short-lasting (disease?)

Relative contribution of production chain phases to health and performance of broiler chickens: a field study

Ingrid C. de Jong[1] and Johan W. van Riel

Wageningen Livestock Research, Wageningen University and Research, PO Box 338, 6700 AH, Wageningen, The Netherlands

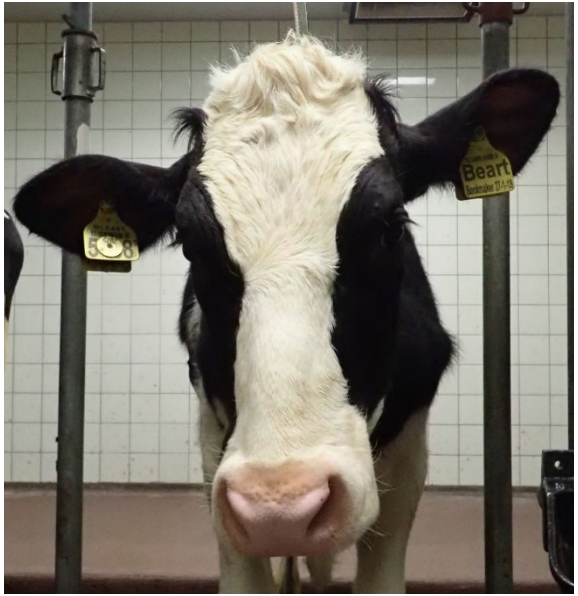
Poultry Science, 2019

Ingrid.dejong@wur.nl

Development of an automated disease scoring tool

Why an image-based scoring system?

easy to acquire
possibilities for 24/7 automated scoring
potential for early warning system
wide applicability (research/commercial)



A face-based disease scoring system using 7 facial features based on literature (Students HAS)

Ocular discharge; Corneal opacity; Eye position;
Ear position; Nasal discharge; Drool; Wounds

Feature values were combined into a single “Final disease score”

With a multiplier to add weight to symptoms associated with infectious diseases
Potential range 0-32

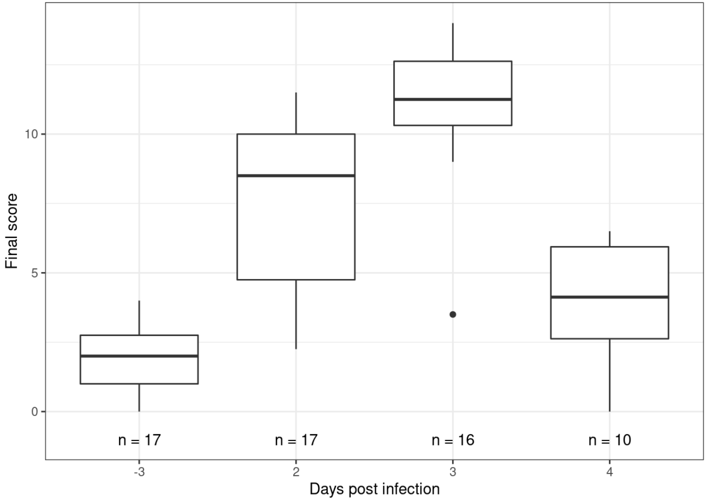
Development of an automated disease scoring tool

Scoring system was applied manually during a FMD vaccination trial at WBVR

- 17 cows received full, ½, ¼ , or $^{1}/_{8}$ dose of vaccine
- Images collected on a daily basis and manually scored
- Drool and nasal discharge received highest scores
- Highest scores on day 3 post infection

Individual feature scores into one final score

- Highest final score on day 3 post infection
- 6 animals removed on day 3



Development of an automated disease scoring tool

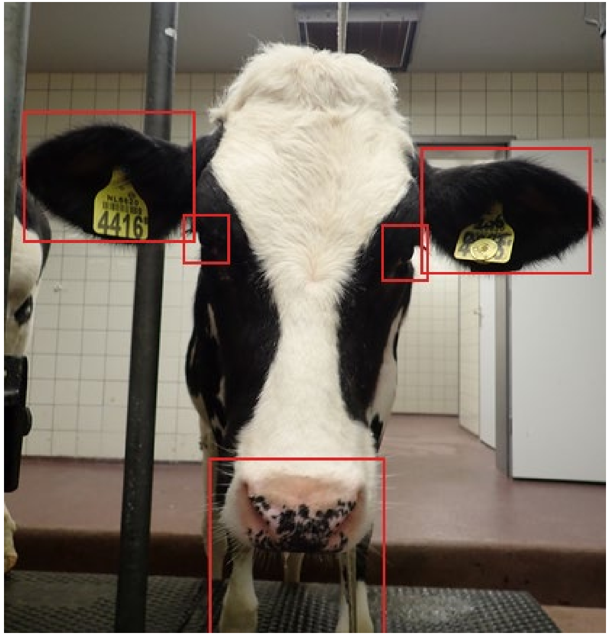
Automation done by students of HAS & AVANS "*Data Science in Agrifood*" minor

- Proof of principle
- Using 4 most prevalent facial features
- ear position, ocular discharge, nasal discharge & drool

Using Microsoft CustomVision

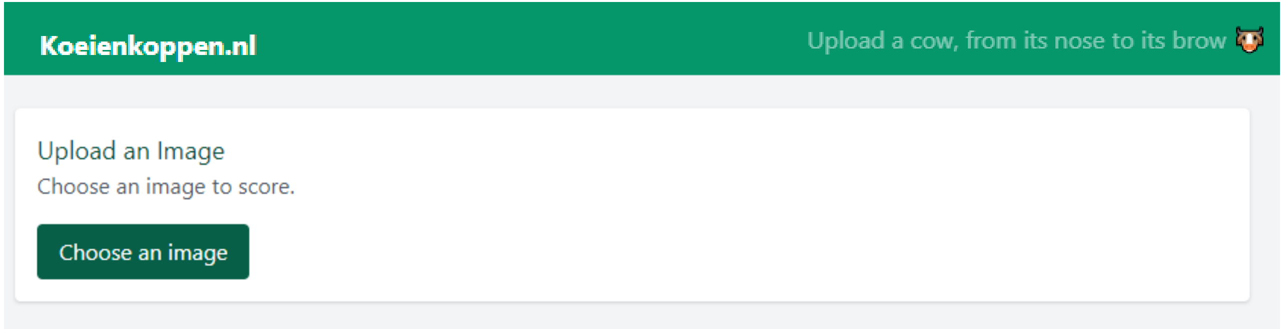
- Model development

API for uploading images, automatically finding the 4 facial features, scoring them, and combining them into a final disease score



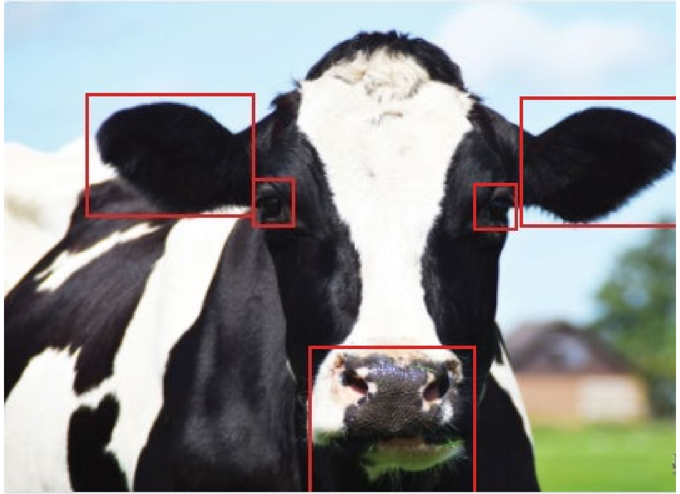
Development of an automated disease scoring tool

https://delightful-sky-028358b03.azurestaticapps.net/#/score/score



Development of an automated disease scoring tool

https://delightful-sky-028358b03.azurestaticapps.net/#/score/score



Koeienkoppen.nl

Cow ID Cow9aN1000.jpg

See all the data that was retrieved from the different models, just like magic!

Score: 3

This score was calculated with a certainty of 98.57%. The score without multipliers is 1

Ear Position: 0

Certainty: 99.253% / Points: 0 / Multiplier: 2

Ocular Discharge: 3

Certainty: 99.994% / Points: 1 / Multiplier: 3

Nasal Discharge: 0

Certainty: 99.995% / Points: 0 / Multiplier: 1

Drool: 0

Certainty: 95.039% / Points: 0 / Multiplier: 3

Extra data

This data was used to calculate the score, and may be usefull for further analysis.

Left Ear

Ear Position: Axial

Certainty: 99.99983300000001

Right Ear

Ear Position: Axial

Certainty: 98.50625

Ocular Discharge

Score: 0

Certainty: 61.91219

Ocular Discharge

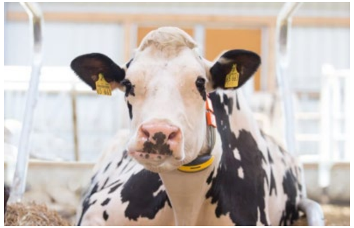
Score: 1

Certainty: 99.99405

WBVR & HAS den Bosch

Ronald.Petie@wur.nl

Predicting resilience using sensor data and ML



Multiple lactations, good (re)productive performance, no/few health problems that are overcome easily, efficient and consistent in milk production
(Adriaens et al., 2020)

Predicting resilience using sensor data and ML

Lifetime Resilience Scoring system

(Adriaens et al., 2020)

A summation of scores for
- number of lactations
- age at first calving/calving interval
- number of inseminations
- number of curative treatment days
- When culled in lactation

1,800 cows scored
Average 1,518 (31 – 6,031)

Divided into 3 evenly distributed classes (H,M,L)

Cows with data from 4 sensors in first parity
N = 370 (109H, 141M, 120L)

WAGENINGEN UNIVERSITY & RESEARCH

100years 1918 — 2018

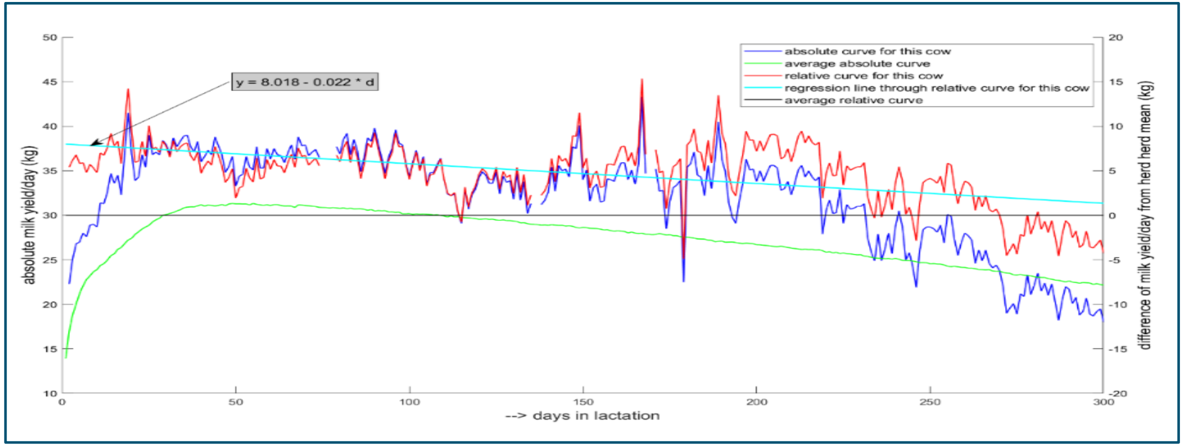
Predicting resilience using sensor data and ML

Activity, Rumination, Weight and Milk 1st lactation to predict LRS

sensor data aggregated to daily values

for each cow, for each sensor, 14 sensor features

mean, minimum, maximum, 25th 50th and 75th percentile, std, skewness, kurtosis, autocorrelation (lag1)
slope, intercept, residual standard deviation
correlation relative curve values - fitted curve values



Predicting resilience using sensor data and ML

Activity, Rumination, Weight and Milk yield to predict LRS

sensor data aggregated to daily values

for each lactation, for each sensor, 14 sensor features per sensor (56 total)

absolute daily values and their lactation averages (1,204 features)

Ordinal logistic regression
56 features
Stepwise selection (p ≤ 0.2)
6 features selected

3 Random forests
6 significant sensor features
56 sensor features

1,204 daily values as features

all models: 10-fold cross validation

Predicting resilience using sensor data and ML

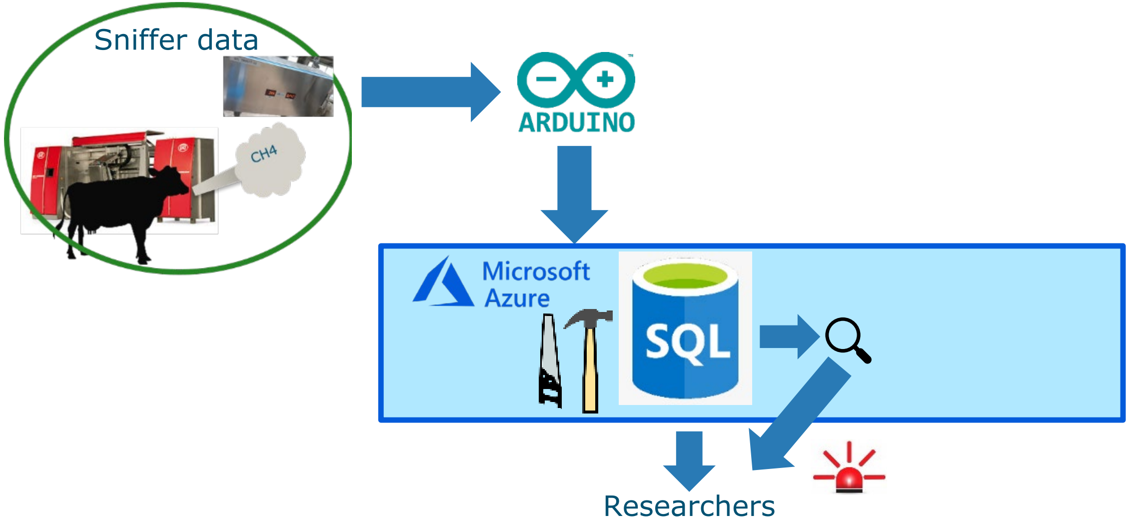
Performance — Accuracy (ACC)
Critical misclassified (CritMis)

	Predicted Resilience class		
True Resilience class	L	M	H
L	ACC		CM
M		ACC	
H	CM		ACC

Submitted to peer-reviewed journal

Model	ACC (%)	CritMis (%)
Ordinal Logistic Regression	45.1 ± 8.1	10.8
Random Forest 6F	45.7 ± 8.4	16.0
Random Forest 56F	51.2 ± 10.9	8.7
Random Forest 1204F	50.5 ± 6.3	8.4

A flexible data architecture in the cloud; the birth of a methane data lake

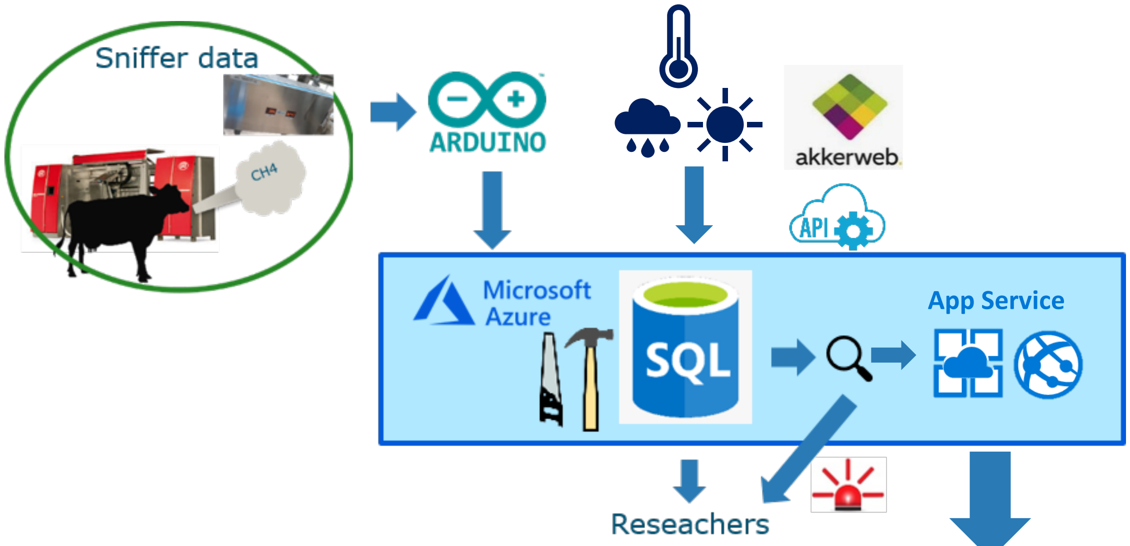


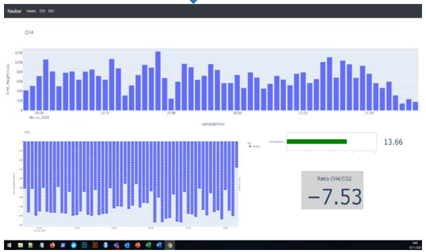
15 Farms with sniffers in Climate Envelop and in PPS: **>100** in 2021

Can we develop a flexible data architecture to collect methane data (2018)

1. (near) real-time
   every 3min new batch of data
2. Scalable to more farms and data
   cheap and WIFI independent
3. That allows for a quality check as soon as data comes in

A flexible data architecture in the cloud





Prototype visualisation tool

Now we have it (near) real-time, can we ***DO*** things in the cloud?

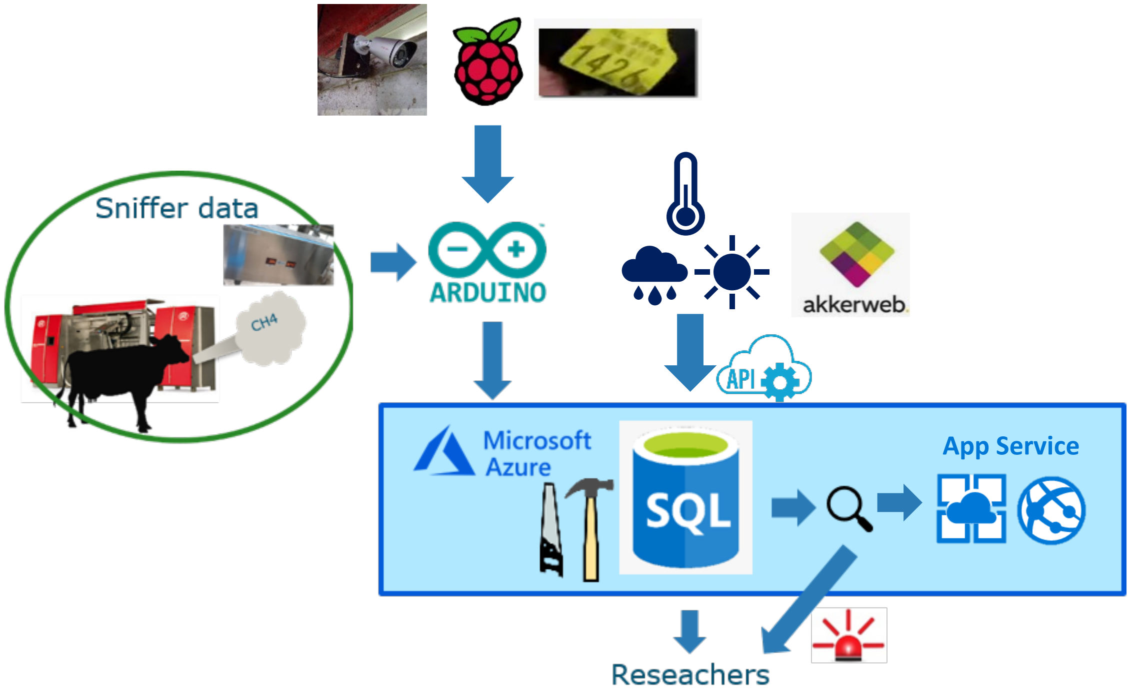
Connect to other data platforms?
Akkerweb

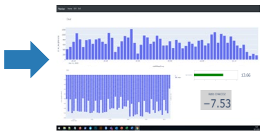
- Weather information
- Past 30yrs, current, future
- 50x50 mtr grid

Visualise data using MA tools

- PowerBI (nope)
- Web-based app
- Farmers sees own data only

A flexible data architecture in the cloud





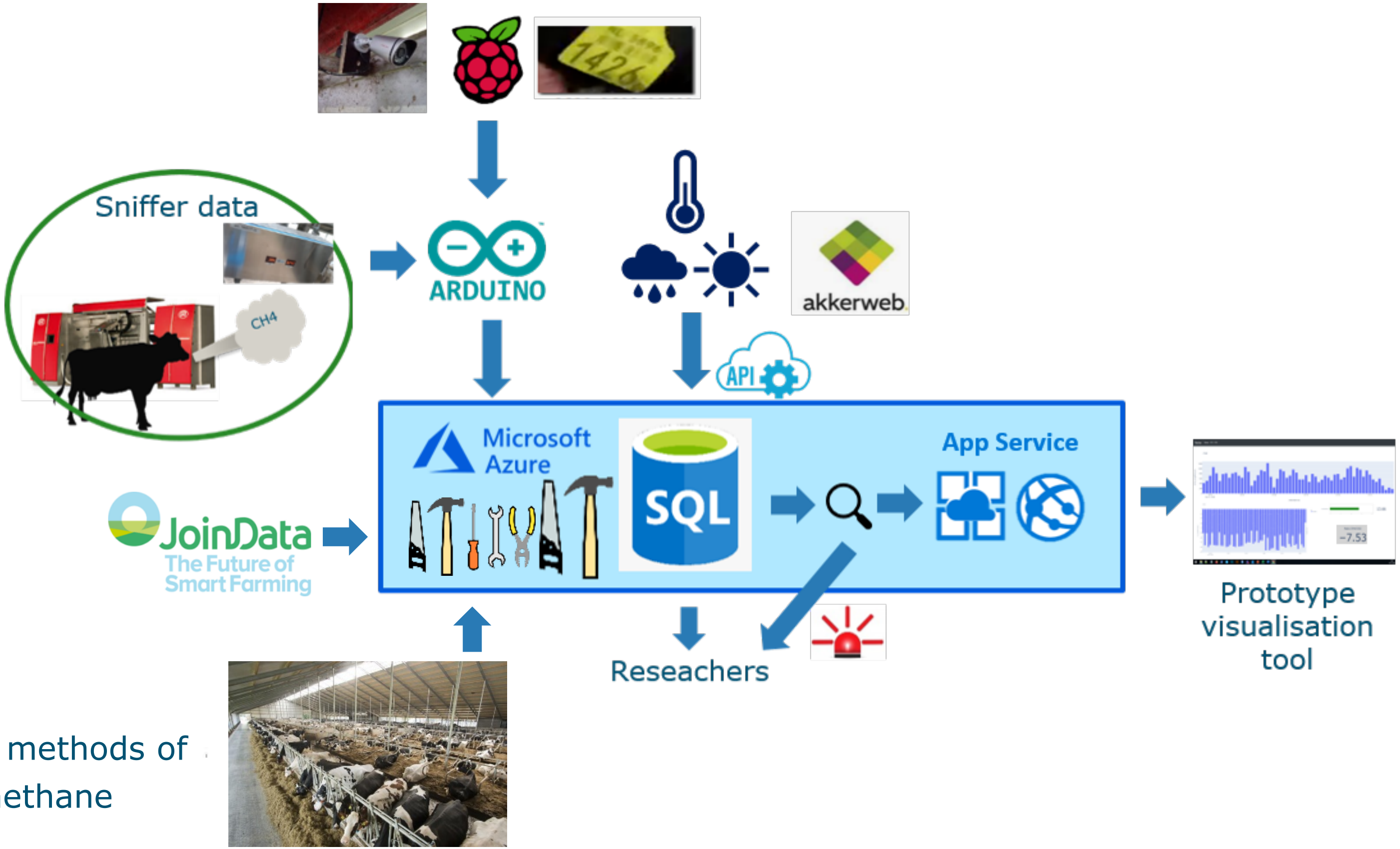
Prototype visualisation tool

Can we also get cow identification (near) real-time?

Infrastructure involves

- Live-stream video recording
- Deep learning to detect yellow blobs
- Transfer learning to ‘read’ digits
- Push of digits to Arduino and cloud
- >80% accuracy for 4 digits
- Now testing on-farm

A flexible data architecture in the cloud



Collect other methods of measuring methane

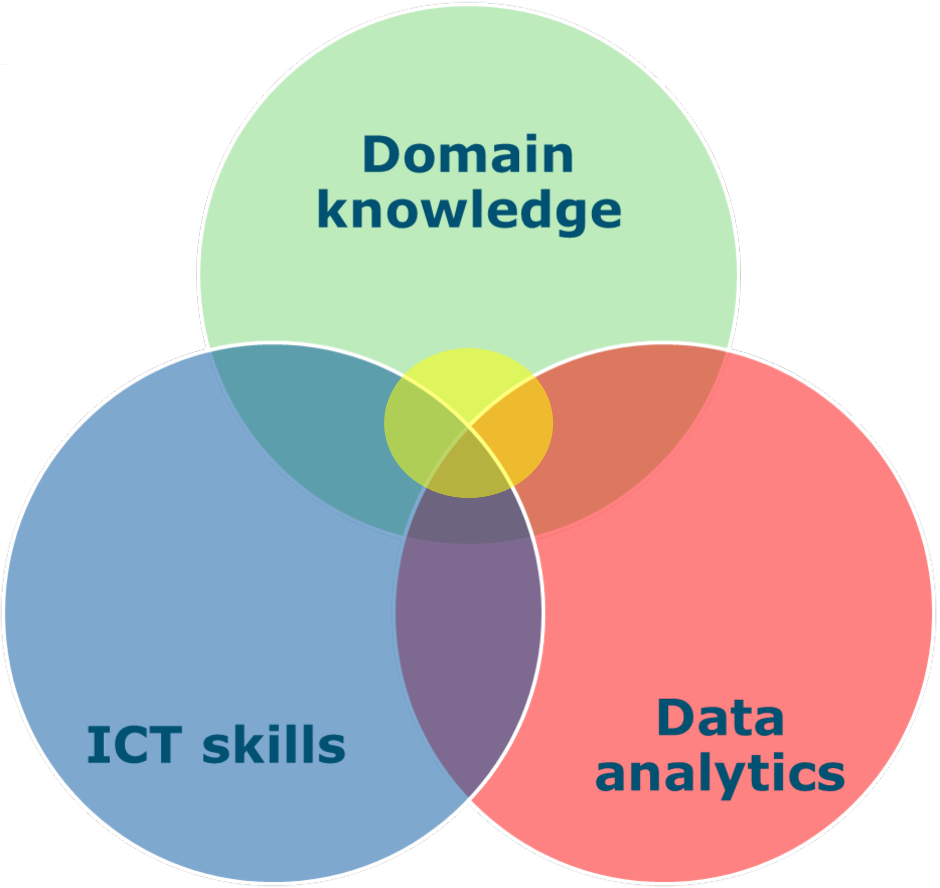
Thank you

Acknowledge

Annemarie Rebel, Ingrid de Jong, Ronald Petie

KB Data Driven & High Tech
GenTORE
Breed4Food

Claudia.Kamphuis@wur.nl



WAGENINGEN UNIVERSITY & RESEARCH

100years 1918 — 2018