Workshop Series:
Reusable Research Data Made Shiny

Ontario Dairy Research Centre | Online
February 21st - 24th, 2023

University of Guelph
Improve Life.

Compute Ontario

ADC

General daily schedule

Time	Event
08:30 - 09:00	Morning Coffee
09:00 - 10:15	**Session 1**
10:15 - 10:30	Coffee Social
10:30 - 12:00	**Session 2**
12:00 - 13:00	Lunch
13:00 - 14:30	**Session 3**
14:30 - 14:45	Coffee Social
14:45 - 16:00	**Session 4**

Day 2 Overview

Session		Topic	Speaker (s)
Morning	1	FAIR Principles	Michelle Edwards
	2	FAIR Principles cont'd Data Reusability Mindset	Michelle Edwards Lucas Alcantara
Afternoon	3	Tour at the Ontario Dairy Research Centre (ODRC)	Lucas Alcantara Rebecca Dunn
	4	Discussion of Practical Data Collection and Challenges at the ODRC	Lucas Alcantara

QUESTIONS?

To make data “re-usable” there are things we need to consider and do:

1. Describe the data – think about telling the story about the Data – Life Cycle
2. Data Governance – is all about managing the data throughout the Life Cycle
3. Overarching to the data governance is WHO has the rights to share the data and allow it to be “re-used”.

All components of Research Data Management (RDM)

FAIR

Who has heard about the FAIR principles?

What do they mean to you?

How about OPEN data?

Is there a difference?

The FAIR Guiding Principles for scientific data management and stewardship

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

FAIR Principles

Let's take a look at the FAIR Principles in a little more detail:

https://www.go-fair.org/fair-principles/

FAIR Principles

Goal is to have all this occur as machine-actionable - Why?

- Imagine I have 30 years of data collected in binders – all handwritten numbers and notes
- How easy is it for me to search and find data collected on a specific date?
- Is it organized according to date?
- Is the handwriting legible? Was it in pencil?

How much data is collected today? Every hour?

FAIR Principles

How do we make data FAIR?

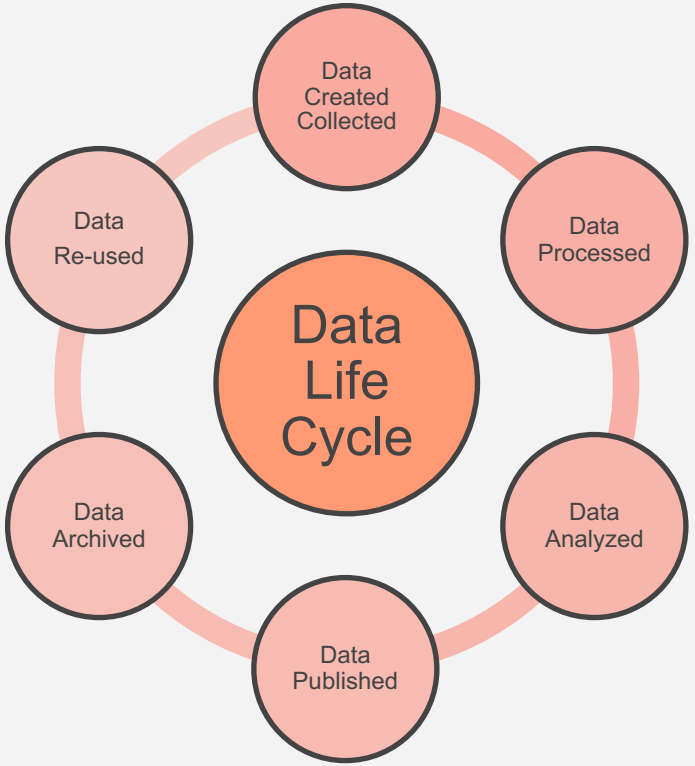
Think about our Data LifeCycle

F= Findable

A=Accessible

I=Interoperable

R=Reusable



FAIR Principles - Exercise

How do we make data FAIR?

In groups of 4-5 (Online attendees we will create ROOMS for you), discuss how you would take one of your projects (past or present) and make it more FAIR. Document one way for each of the following:

F= Findable

A=Accessible

I=Interoperable

R=Reusable

Identify a reporter in your group to share your results with the larger group after Coffee Break

Coffee Break!

FAIR Principles - Exercise

Exercise DISCUSSION

FAIR Principles – Reuse

How we are working to make "older" data reusable

OAC Historical Research Data and Reproducibility Project

https://borealisdata.ca/dataverse/ugardr

How to showcase your results to improve data reusability?

Use an R-Shiny App

J. Dairy Sci. 104:9002–9015
https://doi.org/10.3168/jds.2021-20227

Estimated genetic parameters for all genetically evaluated traits in Canadian Holsteins

G. A. Oliveira Junior,[1]* F. S. Schenkel,[1] L. Alcantara,[1] K. Houlahan,[1] C. Lynch,[1] and C. F. Baes[1,2]
[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, N1G 2W1, Canada
[2]Institute of Genetics, Department of Clinical Research and Veterinary Public Health, University of Bern, Bern, 3001, Switzerland

ABSTRACT

Genetic improvement is a crucial tool to deal with the increasing demand for high quality, sustainably produced dairy. Breeding programs are based on genetic parameters, such as heritability and genetic correlations, for economically important traits in a population. In this study, we estimated population genetic parameters and genetic trends for 67 traits evaluated on heifers and first-lactation Canadian Holstein cows. The data consisted of approximately 500,000 records with pedigree information collected from 1980 to 2019. Genetic parameters were estimated using bivariate linear animal models under a Bayesian approach. Analyses for the 67 traits resulted in 2,211 bivariate combinations, from which the estimated genetic parameters are reported here. The most highly heritable traits were fat percent (0.66) and protein percent (0.69), followed by stature (0.47). Lowest heritabilities (0.01) were observed for disease-related traits, such as lameness and toe ulcer, and calf survival. The genetic correlations between gestation length, calf size, and calving ease measured on both heifer and cows were close to unity. On the other hand, traits such as body condition score and pin width, cystic ovaries and sole ulcer, rear teat placement, and toe ulcer were genetically unrelated. This study reports genetic parameters that have not been previously published for Canadian Holstein cows, and provides updates of those previously estimated. These estimates are useful for building new indexes, updating existing selection indexes, and for predicting correlated responses due to inclusion of novel traits in the breeding programs.

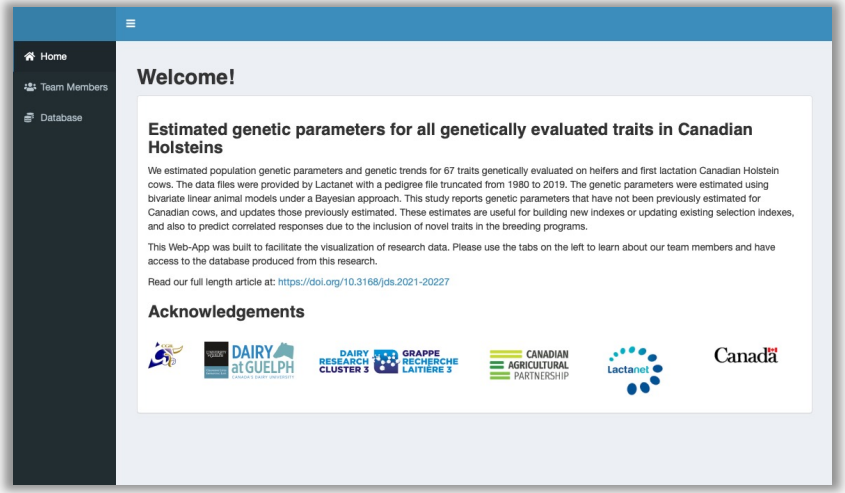
Key words: heritability, genetic correlation, genetic trends

INTRODUCTION

The Canadian dairy industry has approximately 968,700 dairy cows on 10,371 farms. Among these animals, around 64.2% are enrolled in milk-recording programs, of which 39.8% are on supervised systems and 24.4% are on unsupervised systems. The Canadian dairy cattle population consists of 7 main dairy breeds, with 93% of cows registered as Holstein (Canadian Dairy Information Center, 2019). Demand for dairy products is likely to increase over the coming decades as global population and wealth increases. Although the genetic improvement of livestock has been very successful in increasing productivity, there is still pressure to improve the sustainability of the dairy industry and the traits contributing to it.

In Canada, genetic selection in Holstein cattle has been done through the Lifetime Performance Index, formerly known as Lifetime Profit Index (**LPI**), since it was introduced in 1991 and, more recently, through an alternative index called Pro$. In the last 20 yr, 3 changes were made to LPI's main components; that is, production: durability: health and fertility, shifting their respective emphasis from 57:38:5 (2001) to 54:36:10 (2005), 51:34:15 (2008), and 40:40:20 (2016). The estimation of genetic parameters for economically important traits is a crucial component of animal breeding programs. Miglior et al. (2017) reviewed the development of phenotypes used in the Canadian dairy cattle selection program over time. Genetic parameters related to the Canadian Holstein population have been estimated for production (e.g., Miglior et al., 2007; Loker et al., 2012), conformation (e.g., Van Dorp et al., 2004), fertility (e.g., Jamrozik et al., 2005; Sewalem et al., 2010, Alves et al., 2020), and relatively more novel traits, such as metabolic diseases (e.g., Guarini et al., 2019), hoof health (e.g., Malchiodi et al., 2017) and milk properties (e.g., Nayeri et al., 2020). Martin et al. (2019) looked at the genetic correlation among a subset of 28 traits on Canadian Holstein cows from all trait groups, except hoof health. In addition to the importance of re-estimating these parameters using current data, a study involving all current routinely recorded

Received January 27, 2021.
Accepted March 25, 2021.
*Corresponding author: gensou@uoguelph.ca



https://cgil.shinyapps.io/correlations

Oliveira Junior, G. A., Schenkel, F. S., Alcantara, L., Houlahan, K., Lynch, C., & Baes, C. F. (2021). Estimated genetic parameters for all genetically evaluated traits in Canadian Holsteins. Journal of Dairy Science, 104(8), 9002-9015. https://doi.org/10.3168/jds.2020-19830

How to showcase your results to improve data reusability?

Use an R-Shiny App

Canadian Journal of Animal Science

OPEN ACCESS | Article

Conformation traits of Holstein cows and their association with a Canadian economic selection index

Lucas M. Alcantara*, Christine F. Baes*,†, Gerson A. de Oliveira Junior*, and Flavio S. Schenkel*

*Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada; †Institute of Genetics, Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland

Corresponding author: Flavio S. Schenkel (email: schenkel@uoguelph.ca)

Abstract

Pro$ is a Canadian economic selection index aimed to maximize profit by increasing production yields, while maintaining conformation and functional traits. Currently, there is an interest in understanding the individual contributions of conformation traits recorded in Canada to the overall economic value of a cow and whether they are equally important. We used multiple polynomial regression and principal component analysis to assess the association of 26 conformation traits with Pro$ using relative breeding values (RBVs) from 9351 proven bulls. The best reduced regression model explained 72.5% of the variance in Pro$, with heel depth and body depth having the highest and lowest effect on Pro$ values, respectively. Four traits classified as intermediate optimum traits, e.g., teat length, showed significant linear association with Pro$ instead of quadratic, whereas bone quality was not significantly associated with the index. Principal component analysis indicated that highly profitable bulls share similar RBV, with a subclustering of bulls of daughters with better mammary system versus better dairy strength and feet and legs. These results provide understanding of the individual contributions of conformation traits to Pro$ and give information to guide the Canadian dairy industry on how to best consider these traits in recording and genetic evaluation programs.

Key words: conformation, Holstein cattle, Pro$, multiple polynomial regression, principal component

Résumé

Pro$ est un indice de sélection économique canadien qui vise à maximiser les profits en augmentant les rendements de production, tout en maintenant les caractéristiques de conformation et de fonction. Actuellement, il y a un intérêt pour la compréhension des contributions individuelles des caractéristiques de conformation enregistrées au Canada à la valeur économique générale d'un bovin et pour savoir s'ils sont d'importance égale. Nous avons utilisé la régression polynomiale multiple et l'analyse des composantes principales afin d'évaluer l'association de 26 caractéristiques de conformation à l'aide de Pro$ en utilisant les valeurs d'élevage relative (REV — «Relative Breeding Values») provenant de 9351 taureaux éprouvés. La meilleure régression réduite expliquait 72,5 % de la variance dans Pro$, avec la profondeur du talon et la profondeur corporelle ayant les plus grands et plus petits effets sur les valeurs Pro$, respectivement. Quatre caractéristiques classées comme caractéristiques optimales intermédiaires, p. ex. longueur de trayon, ont montré une association linéaire significative avec Pro$ au lieu de quadratique, tandis que la qualité d'os n'était pas associée de façon significative à l'indice. L'analyse de composantes principales a indiqué que les taureaux fortement profitables partagent des RBV semblables, avec un sous-regroupement de taureaux provenant de filles ayant un meilleur système mammaire contre une meilleure force laitière et meilleurs pieds et jambes. Ces résultats offrent une compréhension des contributions individuelles des caractéristiques de conformation à Pro$ et offre l'information pour guider l'industrie laitière canadienne sur comment mieux considérer ces caractéristiques pour enregistrement et programmes d'évaluation génétique. [Traduit par la Rédaction]

Mots-clés : conformation, bovins holsteins, Pro$, régression polynomiale multiple, composante principale

1 Introduction

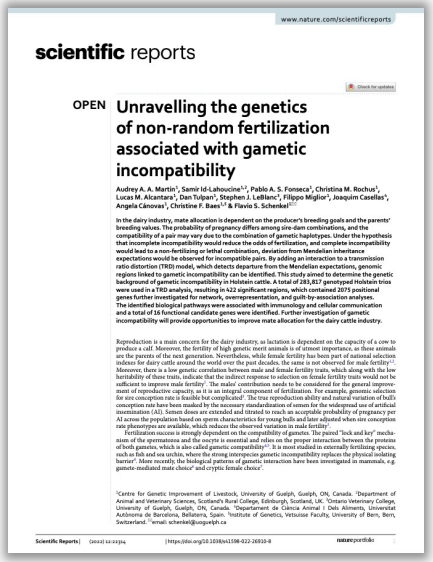
Breeding programs across the globe are shifting their selection goals to more balanced indices that include traits of direct and indirect economic significance, such as conformation, fertility, and health (Cole et al. 2021). Selective breeding in Holstein cattle in Canada has utilized the Lifetime Performance Index (LPI), formerly known as Lifetime Profit Index, since 1991. LPI has been a key index for producers who want a herd with strong conformation and high solid yields, while maintaining functional traits. In

https://alcantara.shinyapps.io/prodollar

Alcantara, L.M., Baes, C.F., Oliveira Junior, G.A., and Schenkel, F.S. (2022). Conformation traits of Holstein cows and their association with a Canadian economic selection index. Canadian Journal of Animal Science. 102(3): 490-500. https://doi.org/10.1139/cjas-2022-0013

How to showcase your results to improve data reusability?

Use an R-Shiny App



www.nature.com/scientificreports

scientific reports

OPEN

Unravelling the genetics of non-random fertilization associated with gametic incompatibility

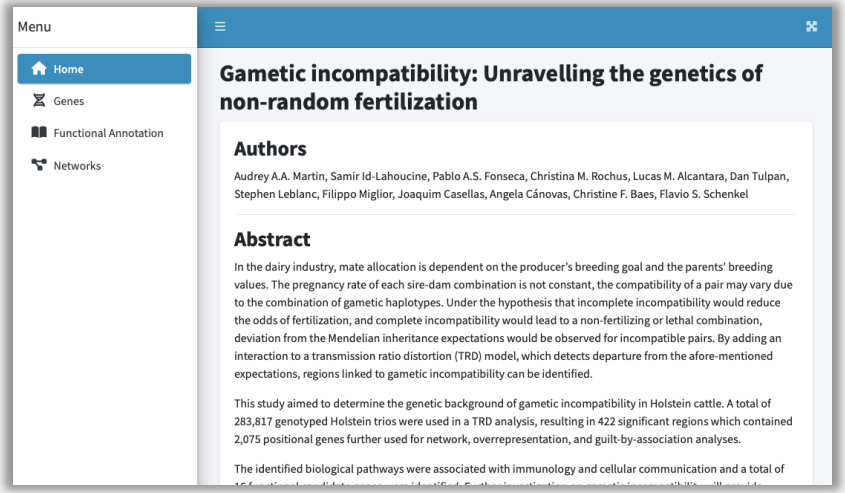
Audrey A. A. Martin[1], Samir Id-Lahoucine[1,2], Pablo A. S. Fonseca[1], Christina M. Rochus[1], Lucas M. Alcantara[1], Dan Tulpan[1], Stephen J. LeBlanc[3], Filippo Miglior[3], Joaquim Casellas[4], Angela Cánovas[1], Christine F. Baes[1,3] & Flavio S. Schenkel[1]

In the dairy industry, mate allocation is dependent on the producer's breeding goals and the parents' breeding values. The probability of pregnancy differs among sire-dam combinations, and the compatibility of a pair may vary due to the combination of gametic haplotypes. Under the hypothesis that incomplete incompatibility would reduce the odds of fertilization, and complete incompatibility would lead to a non-fertilizing or lethal combination, deviation from Mendelian inheritance expectations would be observed for incompatible pairs. By adding an interaction to a transmission ratio distortion (TRD) model, which detects departure from the Mendelian expectations, genomic regions linked to gametic incompatibility can be identified. This study aimed to determine the genetic background of gametic incompatibility in Holstein cattle. A total of 283,817 genotyped Holstein trios were used in a TRD analysis, resulting in 422 significant regions, which contained 2075 positional genes further investigated for network, overrepresentation, and guilt-by-association analyses. The identified biological pathways were associated with immunology and cellular communication and a total of 16 functional candidate genes were identified. Further investigation of gametic incompatibility will provide opportunities to improve mate allocation for the dairy cattle industry.

Reproduction is a main concern for the dairy industry, as lactation is dependent on the capacity of a cow to produce a calf. Moreover, the fertility of high genetic merit animals is of utmost importance, as these animals are the parents of the next generation. Nevertheless, while female fertility has been part of national selection indexes for dairy cattle around the world over the past decades, the same is not observed for male fertility[1,2]. Moreover, there is a low genetic correlation between male and female fertility traits, which along with the low heritability of these traits, indicate that the indirect response to selection on female fertility traits would not be sufficient to improve male fertility[3]. The males' contribution needs to be considered for the general improvement of reproductive capacity, as it is an integral component of fertilization. For example, genomic selection for sire conception rate is feasible but complicated[4]. The true reproduction ability and natural variation of bull's conception rate have been masked by the necessary standardization of semen for the widespread use of artificial insemination (AI). Semen doses are extended and titrated to reach an acceptable probability of pregnancy per AI across the population based on sperm characteristics for young bulls and later adjusted when sire conception rate phenotypes are available, which reduces the observed variation in male fertility[5].

Fertilization success is strongly dependent on the compatibility of gametes. The paired "lock and key" mechanism of the spermatozoa and the oocyte is essential and relies on the proper interaction between the proteins of both gametes, which is also called gametic compatibility[6,7]. It is most studied in externally fertilizing species, such as fish and sea urchin, where the strong interspecies gametic incompatibility replaces the physical isolating barrier[7]. More recently, the biological patterns of gametic interaction have been investigated in mammals, e.g. gamete-mediated mate choice[8] and cryptic female choice[9].

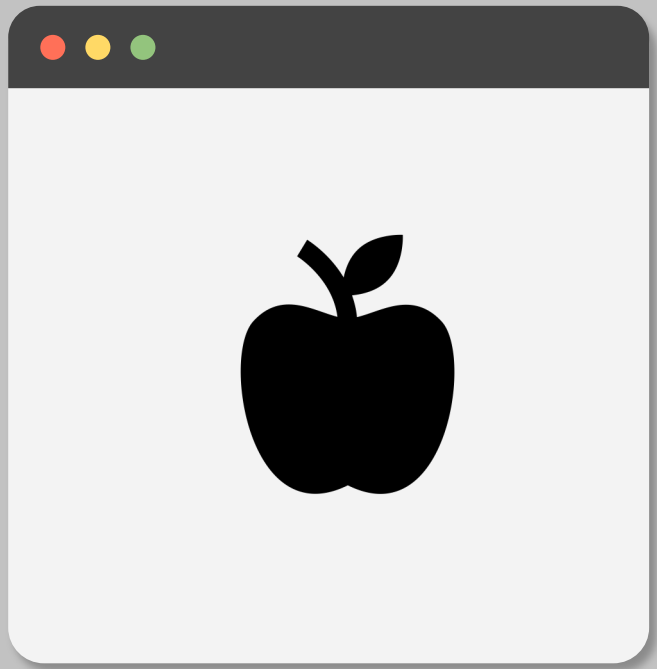
[1]Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada. [2]Department of Animal and Veterinary Sciences, Scotland's Rural College, Edinburgh, Scotland, UK. [3]Ontario Veterinary College, University of Guelph, Guelph, ON, Canada. [4]Departament de Ciència Animal i Dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain. [5]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland. ✉email: schenkel@uoguelph.ca



https://aaamartin.shinyapps.io/netview

Martin, A.A.A., Id-Lahoucine, S., Fonseca, P.A.S., Rochus C.M., Alcantara, L.M., et al. (2022). Unravelling the genetics of non-random fertilization associated with gametic incompatibility. Scientific Reports 12, 22314. https://doi.org/10.1038/s41598-022-26910-8

Lunch



Tour at ODRC

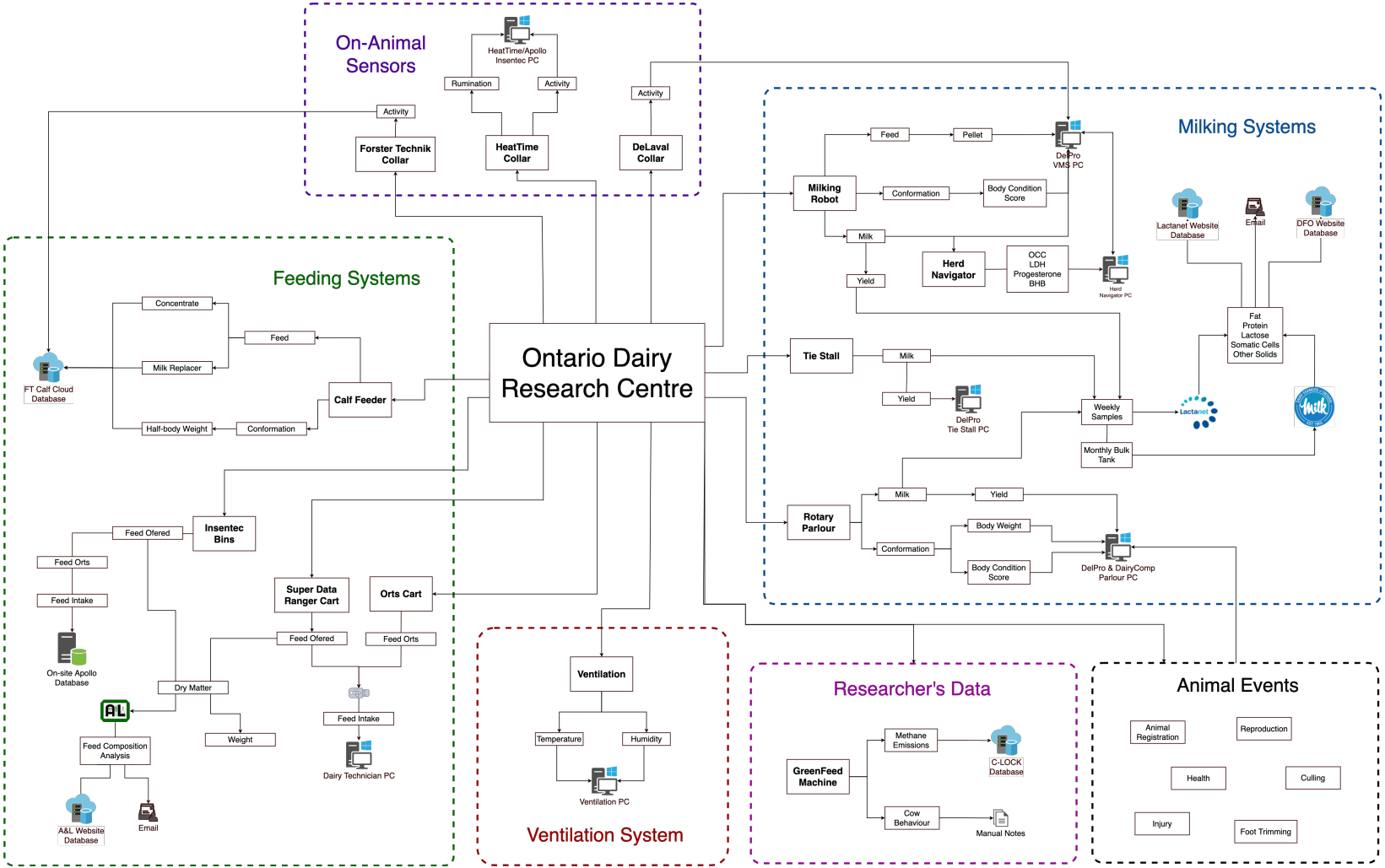
Online attendees:

- 360º Tours at the Ontario Agri-Food and Innovation Alliance YouTube channel: @AgInnovationON

- Maternity and Special Needs: https://youtu.be/OiMbjdyZXBw
- Robotic Milking System: https://youtu.be/2DYhSSAFMIc
- Rotary Parlour: https://youtu.be/VuIIFAv0jNE
- Calf Nursery: https://youtu.be/HdMAsASf43Q
- Lactating Cow Housing: https://youtu.be/lJncYUAeSRY

Coffee Break!

DISCUSSION


On-Animal Sensors
HeatTime/Apollo Insentec PC
Rumination
Activity
Activity
Activity
Forster Technik Collar
HeatTime Collar
DeLaval Collar
Milking Systems
Feed
Pellet
DelPro VMS PC
Milking Robot
Conformation
Body Condition Score
Milk
Yield
Herd Navigator
OCC LDH Progesterone BHB
Herd Navigator PC
Lactanet Website Database
Email
DFO Website Database
Fat Protein Lactose Somatic Cells Other Solids
Feeding Systems
Concentrate
Feed
Milk Replacer
Half-body Weight
Conformation
FT Calf Cloud Database
Calf Feeder
Ontario Dairy Research Centre
Tie Stall
Milk
Yield
DelPro Tie Stall PC
Weekly Samples
Monthly Bulk Tank
Lactanet
Rotary Parlour
Milk
Yield
Body Weight
Conformation
Body Condition Score
DelPro & DairyComp Parlour PC
Insentec Bins
Feed Ofered
Feed Orts
Feed Intake
On-site Apollo Database
Super Data Ranger Cart
Orts Cart
Feed Ofered
Feed Orts
Dry Matter
Feed Intake
Weight
Feed Composition Analysis
Dairy Technician PC
A&L Website Database
Email
Ventilation
Temperature
Humidity
Ventilation PC
Ventilation System
Researcher's Data
GreenFeed Machine
Methane Emissions
C-LOCK Database
Cow Behaviour
Manual Notes
Animal Events
Animal Registration
Reproduction
Health
Culling
Injury
Foot Trimming

Get ready for tomorrow

1. Posit Cloud
    a. Register online if you have not done so yet
    b. If you can, get familiar with it
        i. Take a peak at tomorrow's projects on our Workspace
        ii. Don't look at the answers just yet ;)

2. Don't forget to bring your own laptop, as one will not be provided for you.