Innovation Through Cooperative Data Governance:

The Case of the US Dairy Sector

Jared Hutchins and Brent Hueth

July 7, 2021

Introduction

Rights versus Innovation

- Data is beneficial to innovation, but firms often use data in a way the data producers do not approve.
- Farms produce vast amounts of data, but the rights of farms to their data are hazily defined.

How can we benefit from aggregating data while still protecting the rights of data producers?

A Case Study in Data Governance

Data governance: the rules and regulations surrounding the use and management of data.

Cooperative data governance: data governance which is managed and informed by the producers of the data.

The **National Cooperative Dairy Herd Improvement Program (NCDHIP)** is an example of *cooperative data governance* in agricultural data with several advantages to current approaches.

What Does the NCDHIP Do?

The system collects milk production of all dairy cows on member farms (about 40% of US dairy cows). After aggregating the data, the NCDHIP benefits dairy producers in two ways:

1. Directly: **provides benchmarking analysis** for key performance metrics.
2. Indirectly: **calculates the predicted performance of dairy sires** and releases the estimates to the public.

Dairy farmers are full or part owners of almost every part of the system.

A Tremendous Coordination Problem

To accomplish this, the system is a partnership between:

- USDA scientists
- Land grant extension
- Breed associations
- Animal genetics companies
- Dairy farmers

NCDHIP is a unique public-private partnership in agriculture which **achieves innovation without compromising the ownership rights of the farmer.**

Points of Discussion

What can NCDHIP teach us about data governance today?

1. Cooperative ownership.
   - Cooperative governance brings clarity to issues surrounding data ownership and use.
2. Coordination in data standards.
   - Leadership in data standards is key to innovation.
3. Decentralized operation.
   - Decentralized data collection is a powerful tool in the digital economy.

Outline of this Talk

1. Background
2. History of NCDHIP
3. The System Today
4. Discussion

Background

Genetic Technology Adoption

- Farmers gather information before adopting different genetic varieties.
- Historically, Land Grant Universities have conducted research on plant varieties to help inform farmers about their choices.
- No such centralized authority existed for animals, meaning farmers had to *learn from each other*.

The Problem of Animal Breeding

"Remarkably speculative and economically wasteful."
- Arend Hagedoorn (1946), referring to animal breeding.

Why is animal breeding so much more difficult than plant breeding? Animal breeding:

- has slower data collection.
- is completely decentralized!

Data Collection on New Varieties

Type	Time from breeding until first production data	
Annual Plants	2 years	1 year to produce seeds + 1 year to harvest.
Dairy Cattle	**5-8 years**	**10 month gestation** **+ 2 years to producing age** **+ average 3-5 years of production**
Beef Cattle	2-3 years	10 month gestation + 1-2 years to slaughter.
Swine	10 months	About 4 months gestation + 6 months to slaughter.
Broilers	3 months	1 month gestation + 2 months to slaughter.

Centralized vs Decentralized Breeding

- Farmers **do not need to breed** plants.
  - Plants produce nearly identical offspring, proved varieties can be replicated on the farm.
  - Breeding is **centralized** in *private firms and Land Grant Universities*.
- Farmers **must breed** animals.
  - Animals produce distinct offspring, proved varieties cannot be replicated on the farm.
  - Breeding is **decentralized** in *farmer-owned breeding associations*.

Dairy Cow Breeding Before NCDHIP

The "Purebred" Philosophy

A champion Jersey cow at the World Dairy Expo

- Based on physical appearance ("type") and whether it matches the ideal of the breed.
- Requires investment into a select few genetic lines using ancestry records ("pedigree").
- Coordinated through breed associations (Holstein, Jersey, etc.)

The Sabermetrics Critique

In baseball as much as in dairy farming, **performance matters**.

“If he’s such a good hitter, why doesn’t he hit good?”
- Brad Pitt as Billy Bean, GM of the Oakland Athletics

Statistics and Animal Breeding

The “Moneyball” Philosophy

Dr. Jay Lush
Pioneer of Quantitative Genetics

- Based on observed phenotype rather than physical traits.
- Requires repeated use of the same bulls to produce data.
- Coordinated through data collection and identification of the bull’s offspring through pedigree.

Dairy and Information Asymmetry

Data collection can help dairy farmers *learn from their neighbors* by producing public information on bull performance.

Producing such a public good requires:

1. Measurement technology.
2. Big data.
3. **Institutions**.

The institutional aspect is a key part of the NCDHIP story.

History of NCDHIP

The Story of NCDHIP

Here we will highlight three major phases of NCDHIP:

1. Data collection
   - The founding of the Dairy Herd Improvement Associations (DHIAs) which collected data.
2. Data standards
   - The efforts of the ADSA which promoted uniform data collection standards.
3. Data scaling
   - The invention of AI and freezing technology which drastically increased model accuracy.

The Information Frictions of Dairy

Dairy suffers from two main information frictions:

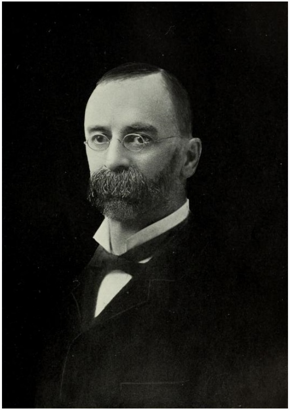
- Uncertain output quality.
  - In New England, watering down milk 25-50% was considered a “universal practice.” (Olmstead and Rhode, 2008)
- Uncertain input quality
  - Which genetics should I choose to increase profitability?

1. Data Collection

Quality measurement technology was disseminated by farmer-owned cooperatives who owned and managed on-farm data.

- 1890: The Babcock butterfat test allowed measurement of milk quality.
- 1905: The first cooperative was formed to measure cow-level milk quality.
- 1908: More cooperatives spurred by the USDA, data used in breeding research.

1890: The Babcock Test



Dr. Stephen Babcock

- In 1890, Babcock invented the first practical butterfat test for milk.
- "Made more dairymen honest than the Bible ever made." (quoted in Olmstead and Rhode (2008), pg. 344)
- Drastic change in economic incentives for dairy farm management and breeding.

1905: DHIAs and Milk Testing

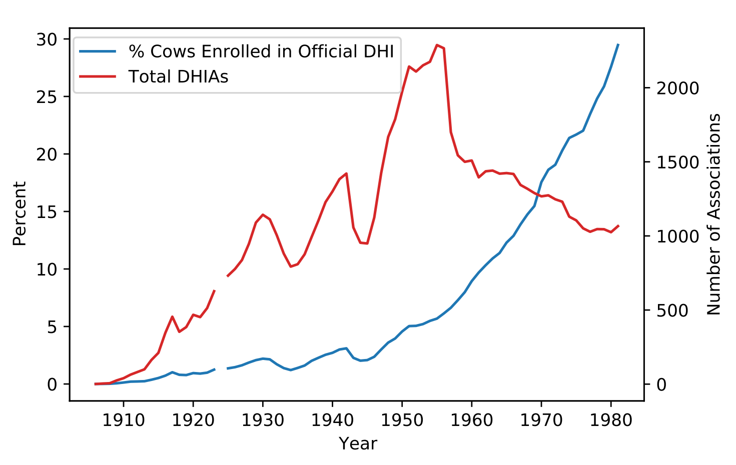
Helmer Rabild, organizer of the first U. S. dairy herd-improvement association.

- Helmer Rabild, a Danish immigrant, organized the first DHIA in Newago County Michigan in 1905.
- DHIA members collectively employed a milk tester to travel monthly to each farm to calculate butterfat yield of each individual cow.
- Originally for the purposes of farm benchmarking.

DHIAs and the USDA

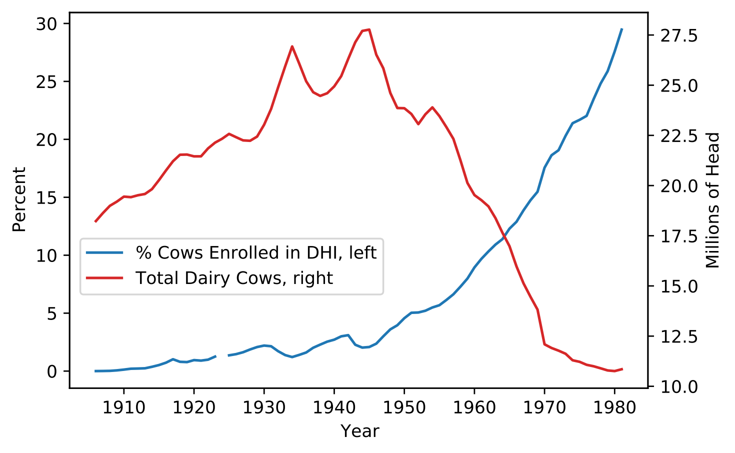
- In 1908, the USDA hires Rabild to organize DHIAs all around the country with the support of the Cooperative Extension Service.
- USDA scientists begin dairy breeding research in 1917, but soon realize the potential of DHIA data for “proving bulls.”
- By using DHIA production data and pedigree together, research begins to shift off of experiment stations on to the farm.

DHIA Growth



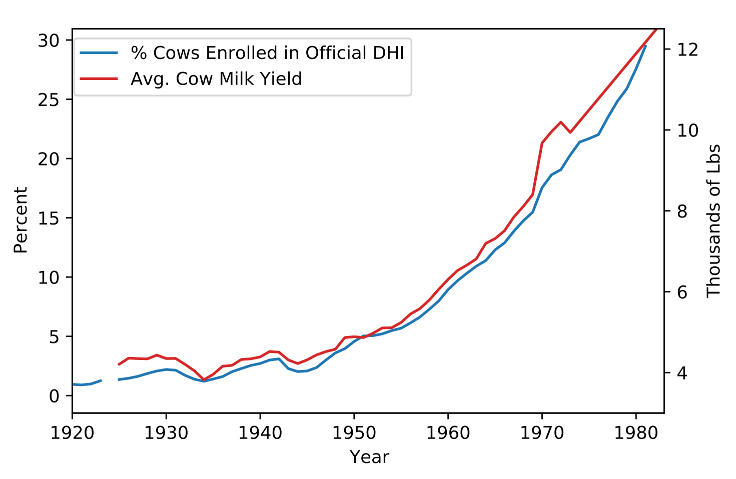
Source: Dairy Herd Improvement Letters, 1925-1980

DHIAs and Structural Change



Source: Dairy Herd Improvement Letters, 1925-1980; USDA NASS

DHIA Growth Mirrors Productivity Growth



Source: Dairy Herd Improvement Letters, 1925-1980; USDA NASS

2. Data Standards

Before bulls could be proved, number of steps had to be taken to **standardize data and data collection**:

- 1924: The American Dairy Science Association standardizes testing procedures for the DHIAs.
  - Also laid out **by-laws and governance structure** for DHIAs.
- 1936: USDA starts the National Sire Proving Program.
  - Begins researching and publishing performance metrics for comparing bulls.

DHI Letter, 1925

All of the equipment is to be owned by the cow-testing association and should be made up as follows:

1. One 60-pound milk scale.
2. Babcock tester (recommend 24-bottle tester)
3. Standard glassware as adopted by the American Dairy Science Association
   a. 30 milk-test bottles
   b. 2 pipettes
   c. 1 skim-milk test bottle
4. Acid measure or dipper
5. A standard milk-sample graduate
6. 24 sample jars (4-ounce bottles)
7. Sample dipper
8. Water bath
9. Dairy thermometer
10. Dividers
11. Computing book.
    a. Smith's computer
    b. Creller's Rechnentafeln.
12. Extra milk pail.
13. Test bottle and sample bottle brushes.
14. Lock and key.
15. A locked field box.
16. Commercial sulphuric acid 1.82 sp. gr.

Source: Cow Testing Association Letter No. 1

The First Proven Sire List, 1937

MISC. PUBLICATION 277, U. S. DEPT. OF AGRICULTURE **33**

HOLSTEIN SIRES

Name	Animals	Records averaged	Milk	Test	Fat
	Number	*Number*	*Pounds*	*Percent*	*Pounds*
AAGGIE CREAMELLE PRINCE 307021	6 daughters	6	9,987	3.3	327
Born, 3-12-20; proved, 9-30-36; ——; Pa.	6 dams	6	8,897	3.3	289
Sire, Delaware Tidy Abbekerk Creamelle 239111.	Difference	(5-4-5)	+1,090	.0	+38
Dam, Roxy Elmwood Aaggie 3d 232408.					
AAGGIE INKA MAY 499959	10 daughters	15	10,551	3.3	349
Born, 10-29-25; proved, 9-25-36; dead; Minn.	10 dams	37	13,583	3.2	435
Sire, Sir Inka May 422078.	Difference	(1-9-2)	−3,032	+.1	−86
Dam, Walkeracres Colantha Bess Aaggie 969044.					
AAGGIE PONTIAC KORNDYKE HARTOG 433562	7 daughters	10	11,960	3.6	436
Born, 2-7-24; proved, 11-16-36; alive; Va.	7 dams	19	12,983	3.5	453
Sire, King Hartog Aaggie Korndyke 350686.	Difference	(2-4-4)	−1,023	+.1	−17
Dam, K P B A McKinley Queen 252246.					

Source: List of Sires Proved in Dairy Herd Improvement Associations (1937), USDA Misc Pub 277

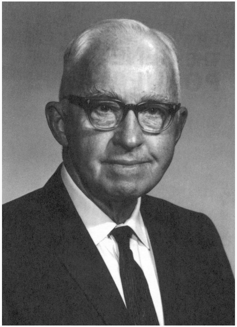
USDA and Scientific Collaboration

Founder of population genetics

Dr. Sewall Wright
USDA ARS/U of Chicago

Studied animal trait inheritance at the USDA.

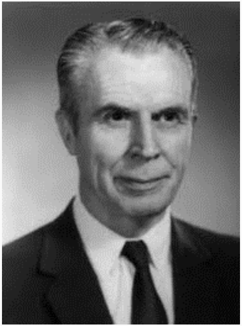
Father of modern animal breeding.



Dr. Jay Lush
Iowa State University

Used of quantitative data for to predict traits.

Pioneer of mixed model equations.



Dr. Charles R. Henderson
Cornell University

Developed a mixed model to predict traits.

3. Data Scaling

Artificial insemination (AI) and freezing drastically **increased the number of data points for each bull.**

- 1933: Commercialization of AI.
  - Before AI, a bull could produce **56 daughters in a lifetime.**
  - After AI, it could produce almost **5,000 daughters a year**.
- 1946: AI cooperatives form the National Association of Artificial Breeders (NAAB).
- 1953: Freezing semen commercially viable.

Adoption of AI in the US

Number of Dairy Cows Bred with AI

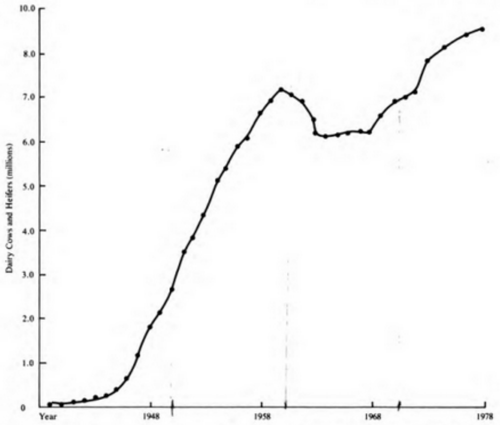
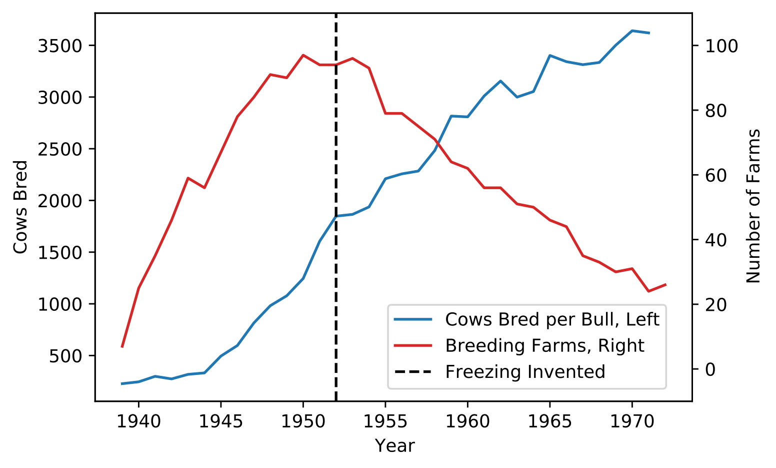


Figure 3.1. Dairy cows and heifers bred artificially to dairy bulls, 1938–1978. Source: *Dairy Herd Improvement Letter* ARS (1939–1979), USDA; and NAAB reports (1947–1979).

Structural Changes in Genetics Markets



Source: Dairy Herd Improvement Letter 48:4, 1972 USDA

The System Today

The Members of NCDHIP

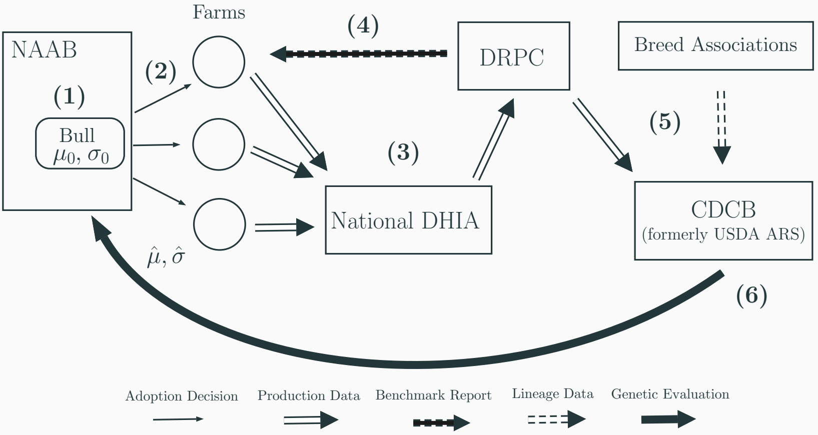
Primary members:

- **National DHIA**: represents dairy farmers.
- **National Association of Animal Breeders (NAAB)**: represents genetics companies.
- **Breed associations**: provide pedigrees.
- **Council on Dairy Cattle Breeding (CDCB)**: cooperative which produces bull evaluations. Members include all of the above.

Secondary members:

- Dairy Records Processing Centers (DRPCs): process and collect data.
- USDA: maintains a cooperative agreement with the CDCB and assists evaluations.

The Roles of the Members



Adapted from Wiggans (1991), pg. 3,856

Governance of NCDHIP

Farmer ownership in the NCDHIP:

Owned Directly	Owned Indirectly	Not Owned
National DHIA	NAAB (mostly)	USDA
DRPCs	CDCB	ADSA
Breed Associations		

At nearly every stage of data governance, the owners of the data have input and control.

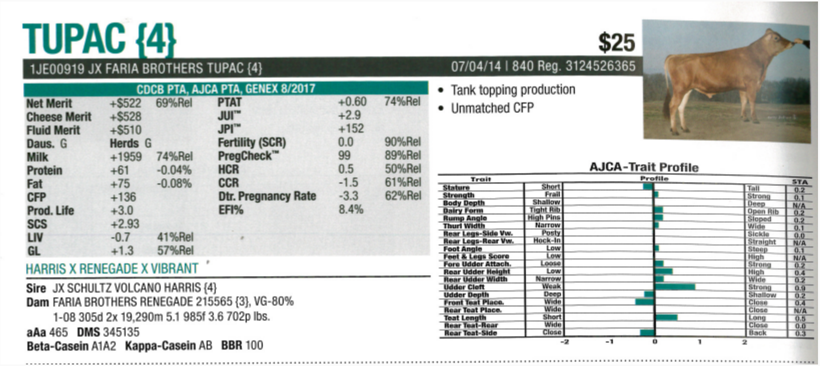
Bull Proving

What is produced by the CDCB...

```
================
PUBRUN=2008
      Bull                 Sire                 Dam          Birth    Ped_Comp% Itb_ID
JE840 3124526365       JE 118157731      JE840 3011206183  2014/07/04     91        1

                                                            Exp_Fut
Bull Name                         Reg St           Inbrd     Inbrd   Dau_Inbrd  Recessive Codes
JX FARIA BROTHERS TUPAC {4}        04    Pedigree   8.2       9.4       7.3
                                         Genomic   12.7       7.9

Current   Cntrl   Sampling   Orig     Entered AI                             Primary
Status    Stud     Status    Stud       Yr Mo    Short Name                 Stud Code
   I        1                  0        15/08    TUPAC {4}                  001JE00919
Eval Breed JE

      PTA   Rel   Daus  Herds Src Mean  DauDev      PA   RelPA
Mlk   1507  .99   1966    61    23660   1515      1048   .46   Yield Rel      0.99
Fat     45  .99   1966    61     1114     46        39   .46   Fat PTA%      -0.12
Pro     42  .99   1966    61      843     43        38   .46   Prot PTA%     -0.06
PL     0.9  .95   1332    38     29.6    1.1      -0.4   .44   Age wt         1.01
SCS   2.99  .99   1966    61     2.84   0.36      3.02   .44
DPR   -4.4  .95   1702    59     38.2             -3.9   .43
HCR    0.9  .92   2005    54                      -0.1   .40
```

Bull Proving

…then is used by genetics companies.



TUPAC {4}

$25

1JE00919 JX FARIA BROTHERS TUPAC {4} — 07/04/14 I 840 Reg. 3124526365

CDCB PTA, AJCA PTA, GENEX 8/2017					
Net Merit	+$522	69%Rel	PTAT	+0.60	74%Rel
Cheese Merit	+$528		JUI™	+2.9	
Fluid Merit	+$510		JPI™	+152	
Daus. G	Herds G		Fertility (SCR)	0.0	90%Rel
Milk	+1959	74%Rel	PregCheck™	99	89%Rel
Protein	+61	-0.04%	HCR	0.5	50%Rel
Fat	+75	-0.08%	CCR	-1.5	61%Rel
CFP	+136		Dtr. Pregnancy Rate	-3.3	62%Rel
Prod. Life	+3.0		EFI%	8.4%	
SCS	+2.93				
LIV	-0.7	41%Rel			
GL	+1.3	57%Rel			

- Tank topping production
- Unmatched CFP

HARRIS X RENEGADE X VIBRANT

Sire JX SCHULTZ VOLCANO HARRIS {4}
Dam FARIA BROTHERS RENEGADE 215565 {3}, VG-80%
1-08 305d 2x 19,290m 5.1 985f 3.6 702p lbs.
aAa 465 **DMS** 345135
Beta-Casein A1A2 **Kappa-Casein** AB **BBR** 100

AJCA-Trait Profile

Trait			STA
Stature	Short	Tall	0.2
Strength	Frail	Strong	0.1
Body Depth	Shallow	Deep	N/A
Dairy Form	Tight Rib	Open Rib	0.2
Rump Angle	High Pins	Sloped	0.2
Thurl Width	Narrow	Wide	0.1
Rear Legs-Side Vw.	Posty	Sickle	0.0
Rear Legs-Rear Vw.	Hock-In	Straight	N/A
Foot Angle	Low	Steep	0.1
Feet & Legs Score	Low	High	N/A
Fore Udder Attach.	Loose	Strong	0.2
Rear Udder Height	Low	High	0.4
Rear Udder Width	Narrow	Wide	0.2
Udder Cleft	Weak	Strong	0.9
Udder Depth	Deep	Shallow	0.2
Front Teat Place.	Wide	Close	0.4
Rear Teat Place.	Wide	Close	N/A
Teat Length	Short	Long	0.5
Rear Teat-Rear	Wide	Close	0.0
Rear Teat-Side	Close	Back	0.3

Animal Genetics Improvement Laboratory (AGIL)/CDCB

"The Beatles"

Dr. Paul VanRaden

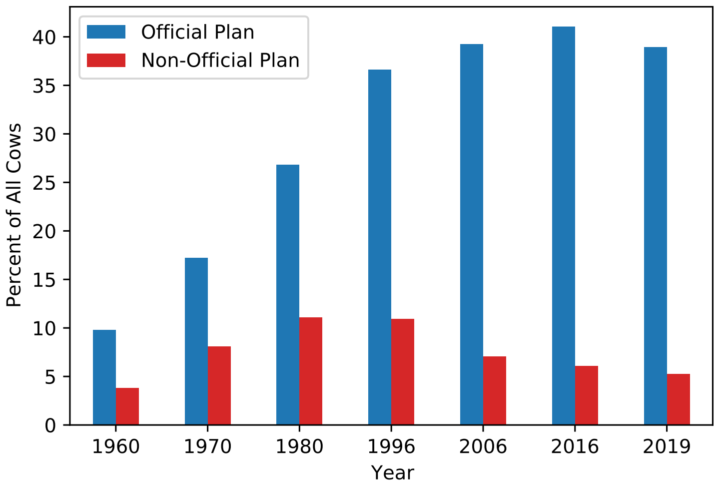
Dr. John Cole

Dr. George Wiggans

Dr. H. Duane Norman

Major figures of the current dairy sire evaluation program.

Program Enrollment



Discussion

Relevance to Agricultural Data Governance Today

Some similarities:

- The invention of a measurement technology.
- Large data volume.
- Data collection is decentralized.

What's different:

- Uncertain regulations concerning privacy and ownership.
- No uniform data standards or "inter-operability."
- Data is siloed in private firms.

Lessons for Agricultural Data Today

Three important aspects of NCDHIP that apply to our current paradigm:

1. Cooperative ownership.
2. Coordination and standards.
3. Decentralized data collection.

1. Cooperatives and Data Ownership

Cooperative institutions can better address the use versus privacy trade-off.

- In NCDHIP farmers own data but license to researchers at LGUs (e.g. yours truly) and the USDA through National DHIA and the DRPCs.
- NCDHIP has engendered 100 years of trust with producers, and participation in the "official" DHI program remains strong.

Farmers owning the means of data collection is a straightforward way to align incentives with use of data.

1. Example Institution

Grower's Information Services Coop

- Aggregates and analyzes agricultural data from member growers.
- Suggests that this kind of cooperative model may work in crops.

2. Coordination and Standards

- Agricultural data is collected through several sources: satellites, sensors, surveys, etc.
- Inter-operability is key to translating data into research and innovation.
- Currently no uniform standards concerning these data.

Leadership by the USDA or industry groups can help solve this coordination problem.

2. Example Initiatives

AG DATA TRANSPARENT

- Voluntary agreement to specify transparency for privacy and use of data.
- May be difficult to get full compliance from across the industry.

AgStack
A Linux Foundation Project

- The open-source approach to standardizing data handling and processing.
- In its infancy, unclear how it will evolve.

The industry still lacks effective coordination.

3. Decentralized Data Collection

Decentralized arrangements like NCDHIP are already catching on in other sectors.

- New technologies exist (e.g. blockchain) that can protect privacy and still coordinate transactions without the need for a central intermediary.
- Many new innovations in data collection are decentralized and effectively "crowd sourced" (OpenStreetMap, Wikipedia, etc.).

Data collection in agriculture is already decentralized and data governance should be designed with this in mind.

3. Example Initiatives

MIDATA

SALUS COOP

COOPERATIVA CIUDADANA DE DATOS
PARA LA INVESTIGACIÓN EN SALUD

- Health data cooperatives have users submit data which is then made available to researchers.
- Protects privacy and ownership while realizing benefits from scaled data.

This model recognizes the key partnership between data producers and researchers.

Concluding Thoughts

- Measurement technology increases data; institutions determine who benefits!
- The USDA has historically played a large role in setting up these kinds of institutions (and can again).
- Privacy and innovation need not be trade-offs if data producers have control over their data (Jones and Tonetti 2020).