Council on Dairy Cattle Breeding
CDCB new developments

Ezequiel L. Nicolazzi, CDCB Technical Director

CDCB

COUNCIL ON DAIRY CATTLE BREEDING

From last year…

- Review and document all aspects on nominator metrics

https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/QC_Metrics_for_Genomic_Nominators

- Review QC requirements document

GENOMIC NOMINATOR APPLICATION AND CERTIFICATION INFORMATION ¶

- CDCB Genomic Nominators - Quality Certification Guidelines (pdf)
- CDCB Genomic Nominators - Application Form (docx)
- CDCB Genomic Nominators - SOP template (docx)
- CDCB Genomic Nominators - MLA (docx)
- QC Metrics for Genomic Nominators
- Approved CDCB Genomic Nominators
- Genomic Nominator Certification Procedure

- List of all animals with "usability code = N" for nominators in affiliate query

Affiliate Specific Genotype Reports

Select desired report
Select affiliate to build
Submit

✓ Lists genotypes with fee code=N loaded in the past 6 months
Fee codes for genotypes loaded since the last invoice
Lists parentage only genotypes since the previous genomic run
Reports missing animal ID for a requester
Conflicts for genotypes loaded in the past 45 days
Check for missing pedigree of animals nominated in the past 75 days
List conflicting genotypes within animal (negative keys)
Animals with unlikely grandsire (and no other conflict) in the past 6 months
Animals with genotypes that conflict with imputed dam genotypes
Parentage verification records for genotypes loaded in the past 45 days
Nomination performance metrics

+ [NOM]_all_errors.csv

From last year...

- Publish ppts and documentation of workshop

- Improve the nomination/genotype query to list all conflicts in one page.

- Proposed to include the pedigree of the dam and in the web display that will provide the fmt 1E, progeny and calving dates of the dam.

- Include timestamp when parentage was accepted by CDCB

- Include the possibility to find negative keys in the genotype query

Redmine: 2017 CDCB Genomic Nominators Workshop

In constant development (see next section)

"Check Dam" query available

"Evaluation date" was added to genotype query. A new field called "Eval Use Date" now shows the date when the genotype became usable

Added

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

From last year...

- List of latest animals with "MGS/PGS conflicts" for nominators in affiliate query

- Provide a list of herds status by nominator (only nominator herds) each official run

- Create an online tool that creates graphs and stats for all nominators and each performance indicator

Affiliate Specific Genotype Reports

Select desired report
Select affiliate to buil
Submit

✓ Lists genotypes with fee code=N loaded in the past 6 months
Fee codes for genotypes loaded since the last invoice
Lists parentage only genotypes since the previous genomic run
Reports missing animal ID for a requester
Conflicts for genotypes loaded in the past 45 days
Check for missing pedigree of animals nominated in the past 75 days
List conflicting genotypes within animal (negative keys)
Animals with unlikely grandsire (and no other conflict) in the past 6 months
Animals with genotypes that conflict with imputed dam genotypes
Parentage verification records for genotypes loaded in the past 45 days
Nomination performance metrics

[NOM]_feecodes.csv in your SFTP area

Affiliate Specific Genotype Reports
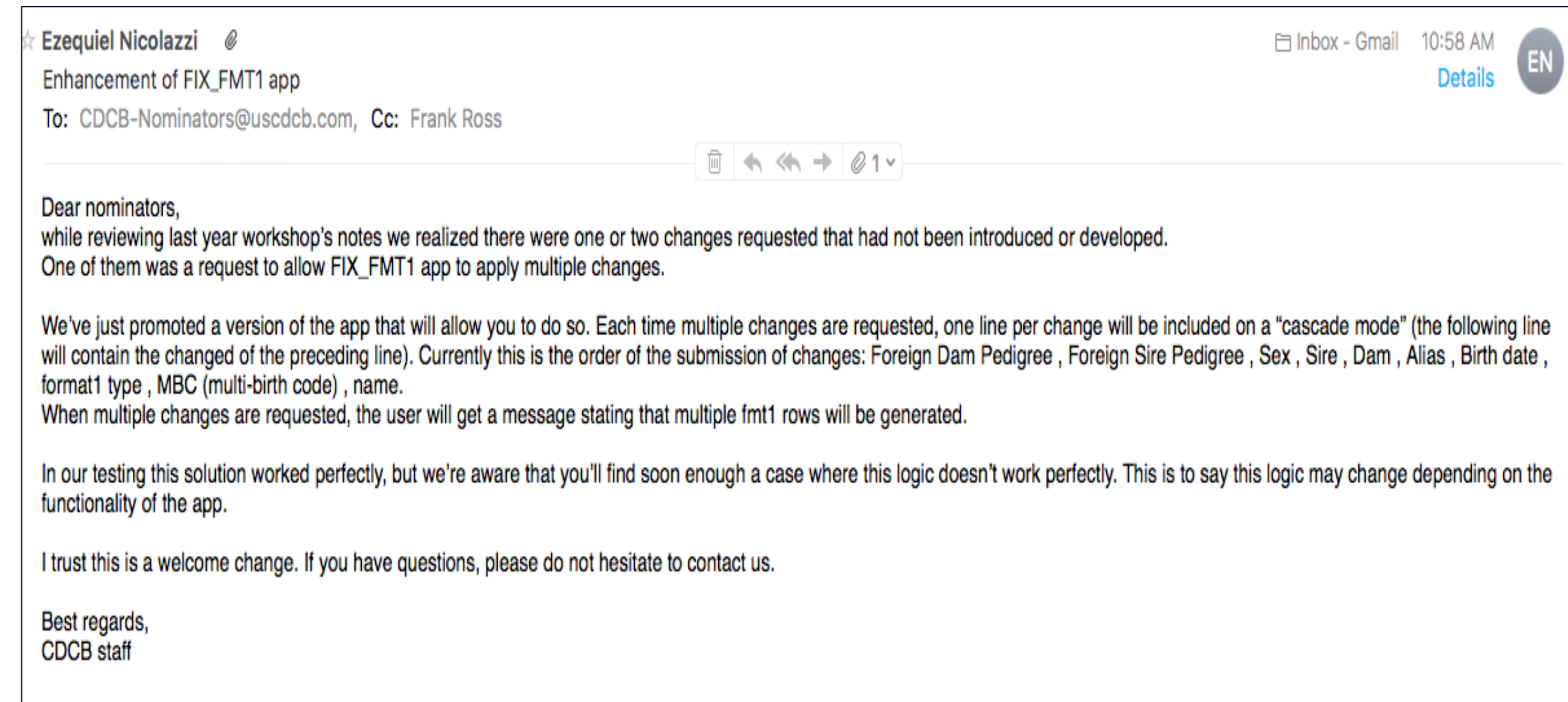
Select desired report
Select affiliate to buil
Submit

✓ Lists genotypes with fee code=N loaded in the past 6 months
Fee codes for genotypes loaded since the last invoice
Lists parentage only genotypes since the previous genomic run
Reports missing animal ID for a requester
Conflicts for genotypes loaded in the past 45 days
Check for missing pedigree of animals nominated in the past 75 days
List conflicting genotypes within animal (negative keys)
Animals with unlikely grandsire (and no other conflict) in the past 6 months
Animals with genotypes that conflict with imputed dam genotypes
Parentage verification records for genotypes loaded in the past 45 days
Nomination performance metrics

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

From last year…

- "Fix Fmt1" option in Genotype query should create 2 or more records if many changes are applied.

- Create standardized messages to be sent as monthly feedback to nominators

Ezequiel Nicolazzi
Inbox - Gmail   10:58 AM
Enhancement of FIX_FMT1 app
Details   EN
To: CDCB-Nominators@uscdcb.com, Cc: Frank Ross

Dear nominators,
while reviewing last year workshop's notes we realized there were one or two changes requested that had not been introduced or developed.
One of them was a request to allow FIX_FMT1 app to apply multiple changes.

We've just promoted a version of the app that will allow you to do so. Each time multiple changes are requested, one line per change will be included on a "cascade mode" (the following line will contain the changed of the preceding line). Currently this is the order of the submission of changes: Foreign Dam Pedigree , Foreign Sire Pedigree , Sex , Sire , Dam , Alias , Birth date , format1 type , MBC (multi-birth code) , name.
When multiple changes are requested, the user will get a message stating that multiple fmt1 rows will be generated.

In our testing this solution worked perfectly, but we're aware that you'll find soon enough a case where this logic doesn't work perfectly. This is to say this logic may change depending on the functionality of the app.

I trust this is a welcome change. If you have questions, please do not hesitate to contact us.

Best regards,
CDCB staff

Ready to go… program ready, waiting to finalize audits

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

From last year...

- Allow to correct breed in "Fix Fmt1" option in Genotype query

  This one will be included in a future release in development.

- Include birth code and breed source in nomination query.

  Probably added to the above

From last year…

- Include DOB limit for foreign bulls entering the official evaluations without paying the service fee (<486 days).

    - My bad (realized I left this out while reviewing last year's notes).

    - See next section for (future) good news on this

    - The *actual* rule is: Born prior to 486 days on the **1st day of the month of the release date.**

    - **Ex. Apr release was April 3rd. Foreign animals without AIS fee paid would have entered the evaluation if born before or on :**

        - **(Apr 1st – 486 days)**

        - **With a bit of help from google… Dec. 1st 2016**

Date Calculator: Add to or Subtract From a Date – Results

This calculator enables you to add or subtract to a date to calculate a past or future date and time.

Start Date

Month: Day: Year: Date:	Add/Subtract:	Years:	Months:	Weeks:	Days:
4 / 1 / 2018	(–) Subtract				486

Today

Include the time   Include only certain weekdays

☐ **Repeat**

Calculate New Date

From **Sunday, April 1, 2018**
Subtracted 486 days

Result: **Thursday, December 1, 2016**

CDCB old and new developments (some)

- Redmine fully implemented
- Report cards including metrics
- Pi=B and BBR > 90%
- Double-weekly nearby triannual releases
- MGS unlikely exclusion and new MGS strategy (George)
- Sex change and fees
- Switch to new website and query system (October 2017)
- FTP transition to SFTP

- New genotype query
- Update of PVR records to include "gets_eval" field
- Changes in evaluation formats
- (long term) New display of nominator/genotype queries
- (mid-long term) Preprocessing of files
- (short-mid term) New publication process back-end
- (short term) merging error and conflict table

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB old and new developments

- **Redmine fully implemented**

  - 1600 tickets closed in less than 1 year

  - Documentation and implementation seem to have been easily included in all nominator's pipelines

  - Still receiving support requests via email (!)

  - Questions? Requests?

- **Report cards including metrics**

  - Report cards including metrics and PASS/FAIL info have been distributed monthly since 2017.

  - Automated email request of further information for any FAILed metric will be implemented soon.

  - About that…

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

TESTING PHASE

Automated request of FAILed metrics feedback

CDCB nominator & lab portal »
Nominators_metrics_review

+ **Overview** **Activity** **Issues** **Settings**

Support #1635   ✎ Edit ⭐ Unwatch ⧉ Copy 🗑 Delete

XX nominator report card - FAILs present, please respond   « Previous | 1 of 7 | Next »

Added by CDCB automatedsystem less than a minute ago.

Status:	New	**Start date:**	05/07/2018
Priority name:	Normal (when possible)	**Due date:**	
Assignee:	Someone from your organization		

Description   💬 Quote

FAIL REPORT FOR 1805

The following FAIL metrics were found while reviewing XXX nominator report card for run:

- Total genotypes for XX 1805 : 474
- FAIL,36,Genotypes missing nomination when loaded
- FAIL,30,Genotypes not usable due to conflicts

Please note that the report card used to generate this warning is in your "out" directory of the CDCB SFTP server.
Metrics used to determine PASS/FAIL reports are available at: ⧉ https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/QC_Metrics_for_Genomic_Nominators
Also note that CDCB expects your feedback on each one of these failures within a week.
Feedback can only be accepted if a registered redmine user responds to this email (do not write below the message "Reply above this line")
Do not hesitate to contact us (open a regular Redmine ticket) if you need further explanations.

Best regards,
CDCB Staff

Subtasks   Add

Related issues   Add

Redmine ticket on sub-project "Nominators_metrics_review"

Assigned to 1 of you (no multiple assignees accepted)

Email sent to all of you (same as automated processes)

Simply reply to the email received. CDCB will close the ticket once satisfied with the response

Easier to manage and keep track of answers

Test run(s) next week (please let your colleagues know!)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB old and new developments

- **Pi=B and BBR > 90%**
  - Animals failing SNP breed test but getting BBR > 90% for the evaluation breed, will get an evaluation (one week later)
  - Confusing because these animals do not get a "***Eval Use Date"*** set. Working on a solution (actually 2, one temp, one long term)
    - Temp – Provide a "gets_eval" field in the genotype table
    - Long term – This rule will become obsolete when evaluations for crossbred animals will be available
- **Double-weekly nearby triannual releases**
  - Pending approval from CDCB Board of directors (next week)
  - Request from DSEC (NAAB). Animals becoming usable 2 weeks before the triannual release get a weekly on "old" solutions, and don't get an evaluation for other 5 weeks.
  - Other solutions were explored, but no other option seemed ideal.
  - Will distribute 2 weeklies for these animals. One the week before the triannual release, and another one on the day of the triannual release.
  - Expect differences.

CDCB old and new developments

- Unlikely MGS exclusion (George will talk about this)

- New MGS strategy (George will talk about this)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Sex change and fees

- Sex changes may determine a need to change the fee code

  - Best example is the "P" code for males not applicable to females

- An automated process was put in place to notify nominators of animals changing sex and their current fee code.

- **Procedure will change:** set to "N" the fee of any animal changing sex (email notification will change too).

- Until the change is applied, **any change you cannot do directly MUST BE submitted to CDCB through a Redmine ticket.**

- Once the procedure will be in place, you'll be able to assign any fee code desired.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB old and new developments

- **New genotype query (smoothest transition EVER)**

  - *You're welcome* ☺

- **Switch to new website and query system and IT infrastructure (October 2017)**

  - *You're very welcome* ☺☺

- **FTP transition to SFTP**

  - *You're most welcome* ☺☺☺

- **Update of PVR records to include "gets_eval" field**

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Changes in evaluation formats

- Inclusion of 7 new traits since last workshop

    - GL + 6 HTH traits

        - XML, CSV, 38 and 105 files.

- Inclusion of AH2

- Change in .zip file naming convention for some of you

    - This is linked to a full re-writing of the transfer procedure that will provide more control over the copying process (implementation expected in September 2018)

(short – mid term) Future updates on formats

- Planned complete update on CSV and official formats (38/105) process.
  - All industry will be informed and a long transition period will be accommodated.
- General (**drafted**) idea has been set up.
- STANDARDIZATION and CONSISTENCY IS KEY
- **Preview shown here for feedback, but proposal needs to be reviewed and approved by the NAAB "IP" committee:**

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Preview of GENOMIC files changes

- Same "PSV" format (pipe delimited) for all users, everywhere in CDCB system.

- Blank values for those fields currently not present in the files (e.g. "stud" in young_bulls file)

- Change in naming convention of files (next slide)

  - Still 2 separate files:

    - "**infoanim**" file: NOM_**infoanim**_YYYYMMDD.csv, NOM_**infoanim**_YYMM.csv

    - "**infoeval**" file : NOM_**infoeval**_YYYYMMDD.csv, NOM_**infoeval**_YYMM.csv

- Maintaining current naming convention of fields.

- *Fixed* decimals, common to all breeds (7 will be consistently expressed as 7.00 or 7.0, if necessary)

- Merging breeds results in one file

- XML format to be gradually dismissed (starting with weeklies, then monthlies)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Words of caution

Note1: blank spaces are for demonstration, actual file will not have any (except when field has one).

Note2: These are drafted proposals intended for feedback. The IP committee needs to validate and approve.

Note3: Irrespectively of the changes decided, there will be a transition period before full implementation.

NOM_infoeval_YYMM.csv

- Evaluation file provides evaluations in "trait blocks" (following a standard format of GenPTA,GenREL,TradPTA,TradREL,DGV,GenSons).
- Adding a trait is a simple procedure for CDCB and partners (one more row per animal). Order of traits in file **is NOT** important (may change anytime).
- Traits with complete missing information will not be provided
- Example:

```
             ID|Trait|GenPTA|GenREL|TradPTA|TradREL| DGV|GenSons

HOUSA0000000001|Milk | -1000|    60|   -900|     34| 897|    830

HOUSA0000000001|STA  |  0.30|      |   0.35|       |0.31|    0.30

HOUSA0000000001|...  |   ...|   ...|    ...|    ...| ...|    ...

HOUSA0000000002|Milk | -1020|    50|   -980|     32| 900|    810

HOUSA0000000002|STA  |  0.21|      |   0.25|       |0.21|    0.20

HOUSA0000000002|...  |   ...|   ...|    ...|    ...| ...|    ...
```

NOM_infoanim_YYMM.csv (version 1)

Provides all animal information (except for evaluations) on a single file.

```
            ID|eval_breed|Sex|Sire|Dam|anim_name|naab_code|sampID|...|current|stud|herd|AH1|AH2|BH1|BH2|BHD|BHM|BHP|...
HOUSA0000000001|        BS|  M|  BB| CC|    BULL1|         |  sAA|...|      0|    |    |   |   |  0|  0|  0|  0|  0|...
```

PROS:
- All anim info in 1 line
- Same format for every breed (no changing format of any kind for different breeds)

CONS:
- Different format from infoeval
- Large number of missing values
- Other breed fields in every animal

NOM_infoanim_YYMM.csv (version 2)

Provides all animal information (except for evaluations) on a single file.

```
                ID|Inforrmation|Value
HOUSA0000000001|   eval_breed|BS
HOUSA0000000001|          Sex|M
HOUSA0000000001|         Sire|BB00000000002
HOUSA0000000001|          Dam|DD00000000003
HOUSA0000000001|    anim_name|Bullnumber1
HOUSA0000000001|    naab_code|
HOUSA0000000001|       sampID|BS0001
HOUSA0000000001|          ...|...
HOUSA0000000001|      current|0
HOUSA0000000001|         stud|
HOUSA0000000001|         herd|
HOUSA0000000001|          HH1|0
HOUSA0000000001|          HH2|1
HOUSA0000000001|          ...|...
```

PROS:
- Maintains the same "stile" and consistency of eval file across breeds
- Allows to "tailor" haplotypes
- Less missing info

CONS:
- 15 + haplo lines x animal
- Can be confusing the fields that are constant (and could be missing) and those that are not reported (but … haplo only!)

* In red: only information breed-dependant

NOM_infoanim_YYMM.csv (version 3)

Provides all animal information (except for evaluations) **in 2 files.**

```
                          ID|eval_breed|Sex|Sire|Dam|anim_name|naab_code|sampID|...|current|stud|herd
HOUSA0000000001|           BS|  M|  BB| CC|    BULL1|         |   sAA|...|      0|    |
```

```
                       ID|Inforrmation|Value
HOUSA0000000001|           HH1|0
HOUSA0000000001|           HH2|1
HOUSA0000000001|           ...|...
```

PROS:
- Clear which fields are breed-dependant
- Clear which files are affected by most common changes

CONS:
- 3 files
- 3 different formats

* In red: only information breed-dependant

(mid-long term) Pre-processing of files

- Will have a greater impact on labs than nominators

- Reprocessing genotypic files is expensive.

- Files affecting the (re)processing of genotypes will undergo greater scrutiny before processing.

- Eg. Submissions with genotypes not nominated to be **rejected**.
  - Requires that CDCB ensures fmt1 data are processed before genotypes

(long term) New display of nominator/genotype queries (in design stage)

- Will be actually a completely new back-end AND front-end.

- Ideally, all information for the animal/sample provided in one place (no single genotype/nominator/dam queries). One-stop shop to perform all nominator queries, as much information as possible provided at once.

- Multiple animals handled at the same time.

- All functionalities will be maintained (and new added)

- Graphically appealing

- Will most likely beta version in 2019 (?)

- *Under evaluation (more for labs than nominators)*

  - *online management of processing results (nomination AND genotyping results).*

  - *Review (documentation and, when possible, standardization) of formats and data shared with nominators (and labs)*

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

(short-mid term) New publication rules back-end

- Will result in new tools front-end

- Re-writing the (internal) procedure to make decisions on animal publication

- Will simplify rules and procedures

- Will allow to develop tools for users that reflect the *actual* status of the animal at cutoff (why my animal did/did not get an evaluation?)

- Careful testing and lots of "secondary adaptation" (other programs in the procedure)

- Status – *about to start internal testing of version 1.0 (June)*

(short term) merging error and conflict table

INPUT (1) : HOUSA000014600898

Animal: HOUSA000014600898 **Sex Code:** F **Eval Breed:** HO
MGS : Unlikely ⇓ **PGS :** Likely/Unknown ⇓

Sample ID	Bar Code	Pos.	Chip Name	Use Ind	Confirmations and Errors	Parentage Only	Sire Status	Dam Status	No. Prog	Tissue Source	Lab Source	Requester	Scan Date	Eval Use Date	No. SNP read	30 SNP Genotypes	Breed SNP Conflicts
16095096	7794288049	R08C01	IIHD	N withdraw	⇑	N	4	2	0		BFGL	UMN-Yang	2012-09-05		764196	100200022205111212021020210000	20

Genotype Confirmations

confirmation type	confirmed genotype		
	Barcode	Position	Current ID

Genotype Conflicts

conflict type	conflicting genotype		
	Barcode	Position	Current ID
Dam conflicts	4454747140	I	HOUSA000013953001 [F]
Unreported progeny conflict	4028220177	H	HOUSA000015353988 [F]
Known MGS unlikely	7794288058	R06C01	HOUSA000001309465 [M]

The latest error record is 3589 hours older than the last change to this genotype.
click here to show/hide the OLD errors !

ERR_SEGMENT_RECORD proccessing date : 20150913 01:37:15		Other Genotype		
Error Code	Message	Animal ID	Barcode	Position
N4	Dam conflict (bad dam = HOUSA000013953001, current pedigree dam = HOUSA000013953001)	HOUSA000013953001	4454747140	I
O3	Discovered non-predigree parent-progeny relationship /w Bdate issues (Anim Bdate = 19911006) BDate of related Anim (HOUSA000015353988) = 19921031	HOUSA000015353988	4028220177	H
O6	MGS is unlikely (unlikely MGS = HOUSA000001309465).	HOUSA000001309465	7794288058	R06C01

Fix_FMT1: Suggestions are based on errors from the !OLD! ERROR_SEGMENT_RECORD !

Conclusions

- Last year's "wrap up" session gave us work for a whole year ☺
  - Improved the functionality and user experience
  - Still work to do.
- User interface and "back–end" expected to change completely (hopefully improving!)
  - Although uncertain of the usefulness (from recent experience with genotype query), a transition period for both systems will be in place
- Large effort in improving documentation and tool usefulness
  - Please keep sending feedback! (even if not certain)

THANK YOU FOR YOUR ATTENTION