Summary of Discussion Sessions from 2024 CDCB Genomic Nominators and Laboratories Workshop

Discussion/Questions 1(Onboarding session):

Presentations included in this session:
- CDCB Overview and System
- Roles and Responsibilities of Nominators and Labs
- CDCB Evaluations

Questions:
- What resources are the most helpful for nominators and labs to train their staff to interact with CDCB?
  - Reaching out to CDCB Staff, Redmine documentation, experience (developing an intuition for handling conflicts), connecting with other Labs/Nominators.
  - SOPs developed in the CDCB Certification Process are also helpful for onboarding new employees.
- For inclusion in evaluations, the deadline on the evaluation calendars is Friday, but you mentioned 6 PM EST on Sundays as being the time of extraction?
  - Fridays are the last day that CDCB staff can respond to questions/concerns prior to evaluations, which is why they are set as the calendar deadline.
  - 2024 evaluation calendar: https://uscdcb.com/genetic-evaluation-schedule/2024/

Discussion/Questions 2 (Onboarding session):
Presentations included in this session:
- CDCB Data Flow and File Exchange
- Resources: WebConnect
- Resources: Redmine
- Data Conflict Resolution

Questions:
- Is there a feature in WebConnect like the old CDCB Query that can show all of a dam's progeny with IDs, birth dates, MBCs, etc.?
  - There is not a report for this currently, as WebConnect is not 1:1 with the old Query system.

- How do you find animals that had genomic conflicts when their genotype batch is loaded?
  - After genotypes are loaded by collaborator Labs, a zip file containing a genomic errors file can be found in the "out" directory of the FTP with the following name scheme: $LAB.batchname.$NOM.zip containing file $NOM_Genomic_Errors.csv
    https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_general_files_distributed_to_nominators
  - A report on WebConnect that can be used to find genotypes that had conflicts after loading: Special Section > Reports > Report dropdown "Conflicts for Genotypes Loaded in the last 45 days"
- What is the turn-around time for requesting Redmine and WebConnect accounts?
  - Should be ~1 day, it depends on if we have adequate permission for the user to access the features of the account.
  - Documentation regarding Requests for CDCB Accounts:
    https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Request_for_CDCB_accounts
- Can you attach files to Tickets submitted to the Redmine Collaborator Portal via Email?
  - Yes, but winmail pictures are not supported.
  - Also, please know that emojis are not supported in Redmine and will cause an email to fail to submit.
- Is it necessary to block records changing birth dates by +/- 2 years?
  - We do not want conflicts to arise in other related animals, so we are cautious when approaching these changes.
  - WebConnect only supports birth date changes by 90 days at a time.
  - For animals with estimated birth dates:
    - Estimated birth dates can be changed up to 5 years at a time.
    - Cross-referencing IDs with estimated birth dates will not work, must enter the birth date as an actual YYYYMMDD.
  - If you have an animal with a birth date conflict that you need assistance correcting:
    - Open a ticket including the animal ID, the Format 1 file name, and the errors that were returned after the last submission of a Format 1.
    - CDCB can assist in complex birth date conflicts and with changing birth dates >5 years.
    - If there are any related animals in conflict with the target birth date, CDCB can let you know who that animal is and why they are preventing the update.
  - Documentation regarding birth date rules:
    https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Birth_Date_Rules

- Discussion Regarding Fee Codes:
  - WebConnect has tools that can assist in assigning fee codes:
    - Special Section > Get Fee
      - Can Query by Animal ID (17 and 18 byte) and by US Herd Code
    - Special Section > Bulk Nomination > Fee dropdown "CDCB Determined"
    - Data Exchange > Special Section >
      - Search for dropdown "Animal ID (17 byte)" and "Animal ID + Sex Code (18 byte)" > API Request dropdown "Get Herd Code and CDCB Fee for Specified Animal ID"
      - Search for dropdown "Herd ID" > API Request "Get CDCB Fee from Herdcode"
  - Remember that fee is a characteristic of the herd code, and either the animal or the animal's dam must be identified with a herd code; CDCB will apply whichever fee code is lower.
  - Fees can be updated before invoicing.
  - Before monthly billing, CDCB places a file in the FTP of nominators containing the fee codes assigned to animals set to bill that period: New/Old$NOM_Check_Fee_Code_YYMM.csv
  - Documentation regarding the fee code credits: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_Fee_Schedule_for_Genomic_Evaluation_Fees
- What is the best recommendation to get pedigree info, or notify the nominators to collect adequate pedigree info, for those animals who don't have full pedigrees at CDCB?
  - Having a deeper pedigree is beneficial for evaluation accuracy.
  - Pedigree of the dam is important, especially as it connects to genotyped maternal ancestors.
- Can you withdraw a genotype with Usability=Y through WebConnect?
  - No
  - If the animal has received an evaluation, we prefer not to remove it. But for cases where the genotype is assigned incorrectly, please submit a ticket explaining what happened and why the genotype needs to be withdrawn, and CDCB will review your request.
  - Documentation regarding Genotype Withdrawal: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/Genotype_Withdrawal
- Discussion on Crossbred cattle evaluations:
  - Explanation of how CDCB manages a crossbred evaluation can be viewed here: https://usacattlegenetics.com/crossbreed/

Discussion/Questions 3:

Presentations included in this session:
- Summary of 2023 Metrics and Reviews
- Introduction to Audit Process with Brevitas

Questions:
- Discussion regarding Brevitas's role in the CDCB Audits:
  - Brevitas will be reviewing the Annual Review and Certification Process that CDCB does for nominators and labs (includes SOPs, metrics, data proficiency testing).
  - Bevitas will not be included directly in communication with nominators or labs, CDCB will be the point of contact between these groups.
- Discussion regarding the "Changes in Pedigree" metric of the Nominator Performance Metrics:
  - Many collaborators agreed that customers are pooling oocytes and use genotyping for parent discovery.
    - Usual to exceed this metric and the response on monthly report cards is generally canned.
  - Discussed potential solutions to make this metric more useful:
    - Change the metric from "Minor" to "Info."
    - Increase the threshold of the metric from 30%.
    - Track the metric using change over time, or Critical: Exceeding previous 3 mo average >5%.
    - Identify the difference between animals with incorrect information submitted vs. animals submitted with blank parentage for parentage discovery.
    - Measure how quickly changes are being made when conflicts are discovered.
  - Collaborators do take note of the metrics – thanks for your work on these, CDCB!
- Discussion regarding XXY Conflicts:
  - Many nominators have reported an increase of animals coming back with XXY conflicts in recent months:
    - Some collaborators pull XXY samples before sending genotypes to CDCB but have also noticed an increase.
  - Animals that are retested rarely have their XXY conflicts resolved.
  - HAUSA offers karyotyping, but to start it is $300, so producers rarely take them up on this.
  - Animals with XXY conflicts do not receive evaluations until they are proven to be fertile and have been entered into a CDCB exceptions table.

     ○ Would be helpful if CDCB could help track this issue?
- ▪ Noting that results may be biased, as some collaborators pull XXY samples before sending genotypes to CDCB.
- ▪ Check pedigree similarities between animals with XXY to see if patterns exist to explain the increase in XXY cases.

Discussion/Questions 4:

Presentations included in this session:
- Genotyping by Sequencing (GBS) at CDCB

Questions:
- Discussion on the cost of implementing GBS vs SNP Arrays (Chips):
  - Currently GBS is more expensive than SNP arrays for domestic labs.
  - China has been able to make the costs comparative, but have also developed the infrastructure to do so, producing their own reagents/materials for GBS in-house.
  - Multiplexing will help spread out the costs for labs.
- What are the benefits of receiving GBS data for CDCB?
  - Research potential increases when there is a flexible number of SNPs available.
- What is the benefit of GBS for labs/nominators?
  - Can respond to discovered haplotypes more quickly, don't have to wait for a new chip to be developed.
- What are the requirements to send GBS data to CDCB?
  - We are currently still developing the validation of GBS data and recognize that GBS methods differ from lab to lab.
  - The pipeline to ingest and validate GBS technology has not been established, as we are still working to validate our first GBS dataset.
  - Future pipelines/validation processes may include sharing the VCF files.
- What is the benefit of using GBS to the end customer?
  - A comprehensive genomic solution: customers can include any gene/haplotype tests of interest when genotyping their animals.

Discussion/Questions 5:

Presentations included in this session:
- Latest updates to CDCB evaluations
- Research updates
- Integrated Communication Project/Follow up from 2023 Workshop
- CDCB Brand Mark Guidelines
- Lactanet Exchange Changes

Questions:
- What genetic conditions are being exchanged internationally for Brown Swiss?
  - Syndrome of arachnomelia and arthrogryposis (SAA), Spinal dysmyelination (SDM), Spinal muscular atrophy (SMA), Weaver Syndrome.
  - https://uscdcb.com/haplotypes/
- Will the updates to the clone workflow extend to identical twins?
  - Yes, as identical twins are considered genetic clones at CDCB.
- What haplotypes does CDCB offer gene tests for?
  - CDCB does not offer gene tests.
  - Labs offer a variation of gene tests, it depends on the lab.
  - Reporting differs for the results that are shared with customers. It may or may not be indicated if the results of genetic conditions are from haplotype calling vs. gene testing.
  - The question was posed: are customers being educated on the difference between a gene test and a haplotype?
- What is the difference between an Unknown Parent Group (UPG) and Constructed IDs?
  - Constructed IDs are used to connect an animal via pedigree to maternal-male ancestors within the database.
  - UPG is a concept in genomic modeling, used to represent the parental contribution to the evaluation for animals with unknown parents.
  - Parent groups are created to calculate the average PTA of parents based on parental birth year.
  - CDCB had previously kept domestic and foreign UPGs separate, but now they are combined in processing as of the April 2024 triannual run.
- Is there a long-term goal to collect data from herds that don't test and will never test?
  - We are working to enhance our incentives program to create additional value for producers that contribute data to the National Cooperator Database. This includes both existing collaborators and those that are not yet sending data. Producers need to see a clear benefit from participating and it is our job to provide that.

- o We need to target herd demographics that are not well proportionately represented in the CDCB database (e.g. Western large size herds, robotic).
  - o The strategy is still in the works and stakeholders will be part of the conversation as it continues to move forward.
- The incentives mentioned, will they have to be managed by nominators as the CDCB fees are?
  - o We are currently in the concept phase at this point – a task force is being assembled by the board of directors.
  - o The form of the incentive has not been determined, but it is possible that it may be managed by nominators.
  - o Data may not be sourced from herds that participate in genomics, so other collaborators may need to be involved.
- What happens when recipient codes among multiple nominators for the same animal do not match (e.g. some are L, some are not)?
  - o The table stores animal and requester, so we will only share based on requester – so when a requester uses 'L', then all genotypes for that animal with that requester will be shared.
  - o A 'D' code should only be sent to remove an existing 'L' code.
  - o Documentation of valid Recipient codes: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/REFERENCES#Ref163
- If we send a 'D', but then the producer changes their mind, what is the transfer schedule after the change in code?
  - o Sharing with Lactanet is done weekly (new genotypes) and monthly (all genotypes not previously shared in a monthly).
  - o In the case that a weekly release has already occurred after the genotype was loaded, the 'L' will not generate a genotype transfer until the next monthly release.
  - o Should remove the recipient code "L" using a "D" before extraction deadlines on Sundays at 6PM EST.
- Is there an agreement with Lactanet on their side that if the code changes from L to D that they will delete the data on their end?
  - o We do not ask them to delete the data on their side, but we won't continue sharing data with them once a D is provided.
- What about sharing pedigree info with Lactanet?
  - o Pedigree data and genomic data are coupled together, so whatever Recipient Code is applied (L or D or blank) for the genotype is applied to the sharing of pedigree.
- Will you consider the same idea of a recipient code for sharing with other organizations?
  - o Currently, CDCB is not considering it for other organizations.
  - o Lactanet is a special situation.

Discussion/Questions 6:

Presentations included in this session:
- WebConnect and API status update
- Data Ingestion System

Questions:
- What is the limit for the number of queries via API?
  - Limits are specific to the query.
    - There is a limitation of 50 animals per query in the "Queries" browser section, and 2,000 total animals per day.
    - For the API, there is a limit of 2,000 animals per request and 15,000 total queries a day.
  - The limits are in place to prevent bottlenecks because of large files and single user requests affecting WebConnect's performance.
- Are the performance enhancements being implemented for WebConnect also going to improve performance of the API?
  - Yes, all hardware performance will be improved
- You said the API is not a "real API"; can you explain?
  - A true API interacts with a database directly (so one system is interacting with another system).
  - WebConnect's API is a data request system, in which users are only receiving full pre-formatted information (no subsets allowed) and are not (yet) allowed to submit data.
- Will there be a "real" API at some point?
  - Yes and no – there will be more control than the API has now.
- Is there a timeframe set for the real API?
  - Not really set yet, but we expect that for EDITS system (submission of pedigree and nomination records, for example), APIs will be available early/mid next year.
- Can multi-day locks be minimized?
  - Most of the activities that freeze the database have not changed much over time. The big difference is that we now inform collaborators via WebConnect.
  - Monthly evaluations take three to four days, and triannuals take a little longer. These cannot be minimized unless there are additional performance enhancements in the evaluations that shorten calculation times (they used to take 5 days not so long ago).

- o There are daily freezes for about an hour to do daily backups/exports. There are special projects that require temporary freezes, which are typically communicated in advance. Projects like 305-AA require a complete freeze because all records in the database are being modified and new data cannot be processed until that is completed. Other events may affect database accessibility but have not changed over time.
- o We will work on more specific messages to inform collaborators about what process is freezing the database (daily maintenance / monthly freeze) and the anticipated timeline of the freeze.
- Discussion on having notice or a monthly schedule of expected outages:
  - o There are three types of locks: planned, special projects, and failure or error outage.
  - o There is currently no schedule available to the public for expected (planned or special project) freezes. However, planned freezes for evaluations are known based on the evaluation schedule.
  - o In some cases, it is difficult to predict the amount of time an outage may occur – especially for failure/error outages where we are actively trying to correct a problem – we can speculate, but if we find errors the timeline will extend to ensure we are outputting reliable information.
- Can planned outages be on a day of the month not corresponding to releases?
  - o CDCB does its best to plan around this for all planned events.
- Can a code/message be added to the API to indicate the database is locked?
  - o You can still submit requests regardless of whether the database is locked, but we will work on a way to provide more informative messages.
- Can we submit an API request when we are blacklisted?
  - o No, you will need to communicate to CDCB why you went over the limit, and we will correct/remove you from the blacklist.
- Will 1E codes be retained in the new ingestion system? When will they be phased out?
  - o They will be retained initially; It would be months/years before we go to a universal format fully (or to only using a universal format).
- Will there be improved conflict correction file processing, especially around birthdates?
  - o Yes, likely not in the very first release, but this is certainly on the list with high priority for future developments.
- Is a full test environment being planned for collaborators?
  - o Once the EDITS cloud system is in production, we believe replicating the functionalities for testing purposes upon request will / should be possible (we will need to explore this once we're closer to deployment).

- Is there an update on increasing the frequency of evaluations?
  - CDCB is looking into this – a task force is discussing this topic specifically.
  - To increase the frequency, we also need to set updated publication and distribution policies.
  - We are looking to potentially increase the updating of phenotypic data to 6X a year, as this will benefit both young bulls and females.