Infrastructure developments at Interbull Centre: Opportunities for new traits

Carl Wasserman

Interbull Centre
Interbull Centre

- International Genetic and Genomic evaluations
- **Infrastructure** and **expertise** for **International Data Exchange** and **Quality Assurance**:
  - Processes, Standards
  - Protocols, CoP
- **Global Reach**
  - 35 countries
  - 5 continents
- **ISO certified**
• **Quality Assurance (Data):** All data validated by customers and Interbull Centre staff before each evaluation via a set of in-house developed programs.
  - Primary key: Interbull ID (worldwide unique identifier)

Breed country sex 12-digit number
SIM  FRA  M  000123456789

• **ISO Standard (Processes)**
• ISO 9001-2015
• High priority for Security and Privacy:
  • Firewalls, Security Enhanced Linux and Intrusion Detection Systems: integrated parts of entire infrastructure
  • Privacy is protected; only users with right permission can access data that they are supposed to
1. **Web interfaces** for users and staff to **Add, Modify** and **Display** data:
   - Interbull Data Exchange Area (“IDEA”)
   - International Genotype Exchange Platform (“GenoEx”)

2. **Database backend**: storing and processing of data in multiple - interacting – databases

3. **Calculating and Storage Cluster**:
   - Modular: Possibility to expand with increase in demand
   - Parallel, fast processing, large RAM
   - Storage cluster with multiple backup; connected with servers for persistent storage.
<table>
<thead>
<tr>
<th>Platform</th>
<th>Database</th>
<th>Records</th>
<th>Breeds</th>
<th>First use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interbull Data Exchange Area (IDEA)</strong></td>
<td>Pedigree data</td>
<td>35 340 746</td>
<td>64</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>EBVs</td>
<td>40 045 829</td>
<td>25</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Performance data</td>
<td>40 969 517</td>
<td>13</td>
<td>2019</td>
</tr>
<tr>
<td></td>
<td>Genetic Traits</td>
<td>22 534</td>
<td>1</td>
<td>2019</td>
</tr>
<tr>
<td><strong>International Genotype Exchange Platform</strong></td>
<td>Genotyped records for InterGenomics</td>
<td>44 565</td>
<td>1</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>GenoEx-PSE: SNP Records for Parentage Verification</td>
<td>57 021</td>
<td>32</td>
<td>2019</td>
</tr>
</tbody>
</table>
Flexible upload system for performance data, able to store complex and hierarchical data structures of infinite size.

Currently used for InterBeef, but possible to extend to other kinds of data.

Integrated with IDEA.

Uploads data in XML format, applies format integrity checks and other procedures to ensure data is sound.
Performance DB in combination with XML allows:

- a hierarchical and extendable format

Upload of complex structures with multiple levels of related data:
- a single data point (a birth year of an animal)
- an array of similar data points (lactation records)
- possible annotation with attributes specifying their meaning
Data checking

• Before upload:
  • Formatting (data types etc.)
  • Is structure correct

• After upload:
  • data is compared with existing data in the database (that animals exists, if they are aliases etc.)
  • verifying the client-side format checks.
  • Erroneous data is filtered out and reported back to the uploader
• Estimation of International Breeding Values for Beef bulls (Interbeef)
  • 5 Breeds: Limousine, Charolaise, Simmental, Aberdeen Angus, Hereford
  • 3 Evaluations / year (2 routine + 1 test)
  • 3 Traits: Adjusted Weaning Weight, Calving Ease and Birth Weight
## Interbeef Evaluations

<table>
<thead>
<tr>
<th></th>
<th>Jan 2017</th>
<th>Jan 2018</th>
<th>Jan 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countries</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Evaluation breeds</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Country-breed-trait</td>
<td>18</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td>combinations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animals in the pedigree</td>
<td>28 256 603</td>
<td>29 965 798</td>
<td>35 109 302</td>
</tr>
<tr>
<td>database</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submitted phenotype records</td>
<td>7 890 490</td>
<td>8 234 838</td>
<td>40 734 117</td>
</tr>
<tr>
<td>International estimated</td>
<td>61 142 008</td>
<td>63 215 818</td>
<td>257 156 657</td>
</tr>
<tr>
<td>breeding values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distributed international</td>
<td>1 344 841</td>
<td>1 407 786</td>
<td>6 805 936</td>
</tr>
<tr>
<td>estimated breeding values</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Future Developments

- System enhanced with parallel and distributed processing (Apache Spark, utilizing its Python API)
- Processing speed improvement
- Bigger data files allowed to be uploaded in a adequate time
• Phenotype exchange for specific traits where sharing of phenotypes matters more than sharing of genotypes
  • Many new traits i.e.
    • Feed Intake, Green House Gas, Meat Eating Quality
  • (Bi)lateral exchange i.e.
    • between countries
    • within specific projects
• New approaches i.e.
  • ”Beef from Dairy”
Enabling importers and exporters to select, worldwide, the best genetics for different countries, environments or breeding goals.

Thank you

Carl