



# **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 & ISO standards

- **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





# Interbull Centre Infrastructure



- 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



# Interbull Centre Infrastructure

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.



# Databases and Data at Interbull Centre

Platform	Database	Records	Breeds	First use
<b>Interbull Data Exchange Area (IDEA)</b>	Pedigree data	35 340 746	64	2009
	EBVs	40 045 829	25	2013
	Performance data	40 969 517	13	2019
	Genetic Traits	22 534	1	2019
	Genotyped records for InterGenomics	44 565	1	2014
<b>International Genotype Exchange Platform</b>	GenoEx-PSE: SNP Records for Parentage Verification	57 021	32	2019



## Interbull Centre Performance DB

- 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



# Interbull Centre Performance DB

- 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

```
<?xml version="1.0" encoding="UTF-8"?>
<interbull type="performance" subtype="interbeef" version="1.0">
  <traits>
    <trait TYPE="cae" EBREED="LIM" RCOU="CZE" ref="1" h2="9" cgn="1" twin="N" dam="Y" mpe="Y" pev="N" IDX="cae1">
      <e ETYPE="R" COL="CG" VALUETYPE="Int"/>
      <e ETYPE="F" COL="ASEXTWIN" VALUETYPE="Int"/>
      <e ETYPE="F" COL="AACA" VALUETYPE="Int"/>
      <e ETYPE="F" COL="YEAR" VALUETYPE="Int"/>
    </trait>
  </traits>
  <animals>
    <a id="LIMCZEM000000181517" twin="1" et="0">
      <trait TYPE="cae" herid="CZE008101626410" status="05" dep="1.0" IDX="cae1">
        <e>23906</e>
        <e>1</e>
        <e>3</e>
        <e>1995</e>
      </trait>
    </a>
    <a id="LIMCZEF000000187517" twin="1" et="0">
      <trait TYPE="cae" herid="CZE008101626410" status="00" dep="1.0" IDX="cae1">
        <e>23906</e>
        <e>3</e>
        <e>2</e>
        <e>1995</e>
      </trait>
    </a>
  </animals>
</interbull>
```





# Interbull Centre Performance DB

## 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



# Current Use



- **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



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



# Interbull Centre Performance DB

## 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



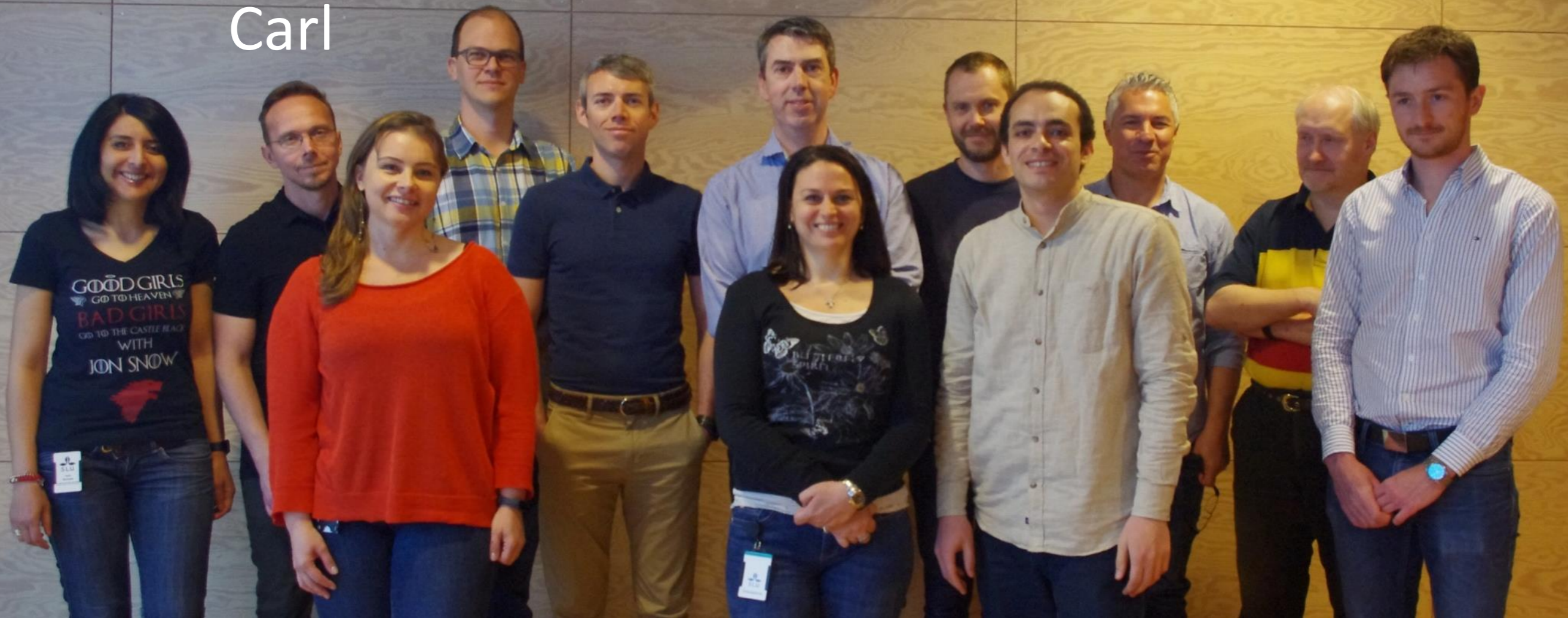
# Opportunity

- 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"



Thank you

Carl



Enabling importers and exporters to select, worldwide, the best genetics for different countries, environments or breeding goals.