“SNPMace”

(International SNP Evaluations)

Next steps towards the first pilot run

Simone Savoia
simone.savoia@slu.se

Outline

Background

Interbull **SNPMace** project

- Aims
- Achievements

Towards the first **SNPMace** pilot run

- Data call
- Next steps

Background

Selection decisions are taken on the basis of gEBVs
calculated from SNP genotypes and their estimated effects.

To improve SNP effects and gEBVs accuracy,
increasing the amount of information used in the estimation process is needed.

Ideal solution:
increase reference populations' size combining across countries raw data (genotypes and phenotypes)

Restrictions and privacy regulations may limit countries in sharing such information

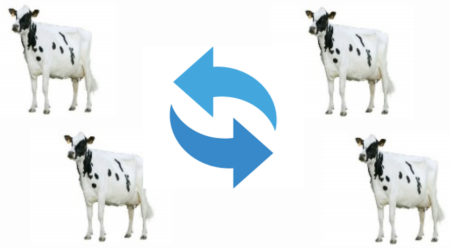
International information is used in national reference populations by using MACE EBVs (genotyped and phenotype females are not included)

Background

Combine across countries SNP effects
has been considered a valuable alternative
to make **genomic prediction more accurate**

- No exchange of raw data

- Use of females’ information from national reference populations



- “difficult to measure” traits with ad-hoc genotypes animals may benefit

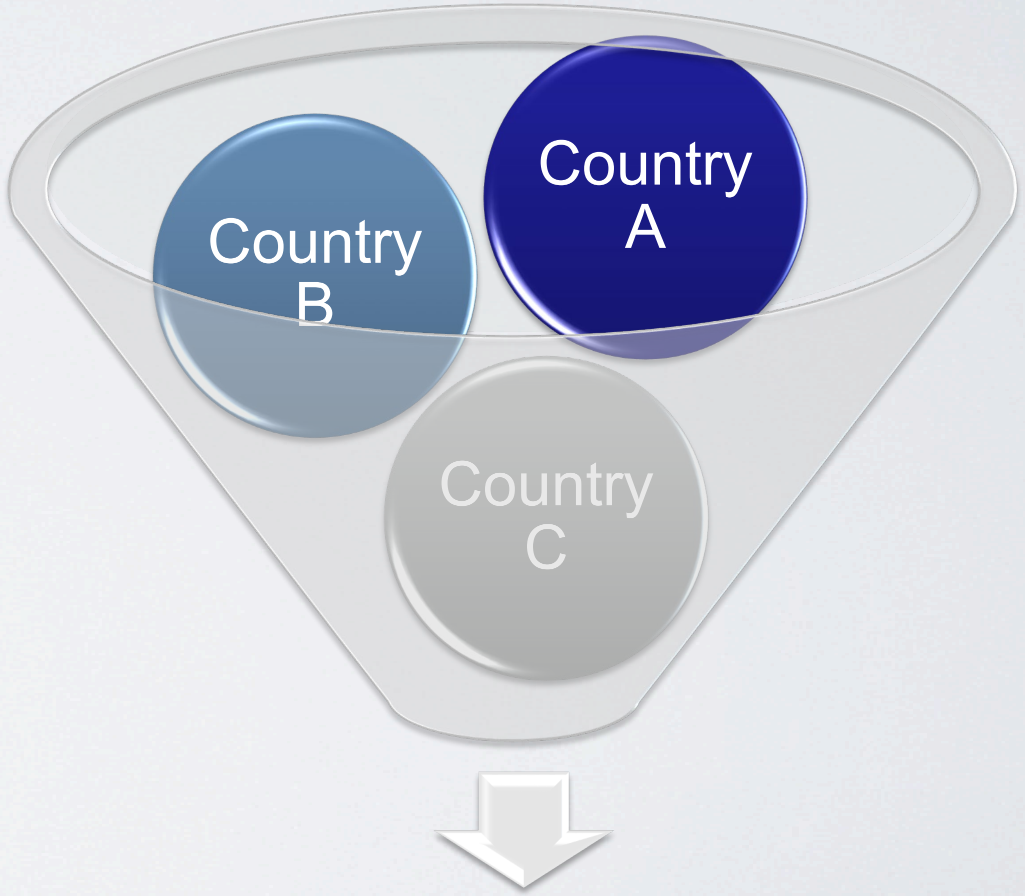
SNPMace project – aims

“International SNP Evaluations” as a new Interbull Service

“**SNPMace**” principles (Meta-analysis)

- exploiting summary statistics
from countries’ national genomic evaluation:
  - LD matrices
  - SNPs’ frequencies
  - SNPs’ effects
  - error variance in the population

- Considers correlations among different populations



More accurate population-specific SNP effects

SNPMace project – achievements

The **‘SNPMace’ methodology** has been developed and the underlying **model** presented to the community:

$$\begin{bmatrix} \ddots & \cdots & \cdots & \cdots & \iddots \\ \vdots & \left[Z_i' R^{-1} \mathbf{Z}_i + \mathbf{G}^{ii}\right] & \cdots & \left[\mathbf{G}^{ii^+}\right] & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \left[\mathbf{G}^{ii^+}\right] & \cdots & \left[\mathbf{Z}_{i^+}' R^{-1} \mathbf{Z}_{i^+} + \mathbf{G}^{i^+ i^+}\right] & \vdots \\ \iddots & \cdots & \cdots & \cdots & \ddots \end{bmatrix} \times \begin{bmatrix} \cdots \\ [\hat{\mathbf{g}}_i] \\ \vdots \\ [\hat{\mathbf{g}}_{i^+}] \\ \cdots \end{bmatrix} = \begin{bmatrix} \cdots \\ [Z_i' R^{-1} \mathbf{y}_i] \\ \vdots \\ \left[\mathbf{Z}_{i^+}' R^{-1} \mathbf{y}_{i^+}\right] \\ \cdots \end{bmatrix}$$

*Goddard et al. (2018)**

A. Jighly and M. Goddard (*Agriculture Victoria, Melbourne*)
developed an efficient **software (MetaGS)** to run the SNPMace model.

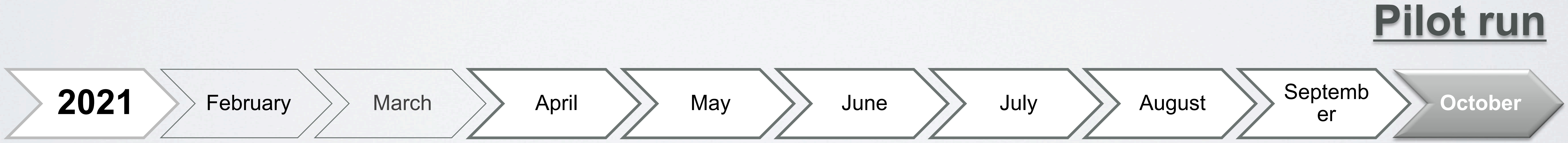
The meta-analysis generated **gEBVs** having
correlation in the range of 0.99 to 1 with those obtained by using raw information

* Additional references at the end of the presentation.

The SNPMace pilot run

A **pilot run** is considered a crucial step prior to commencing any Interbull service:

- Using data from countries interested in joining SNPMace
- Test the feasibility of the service and tune up the system, based on countries' peculiarities



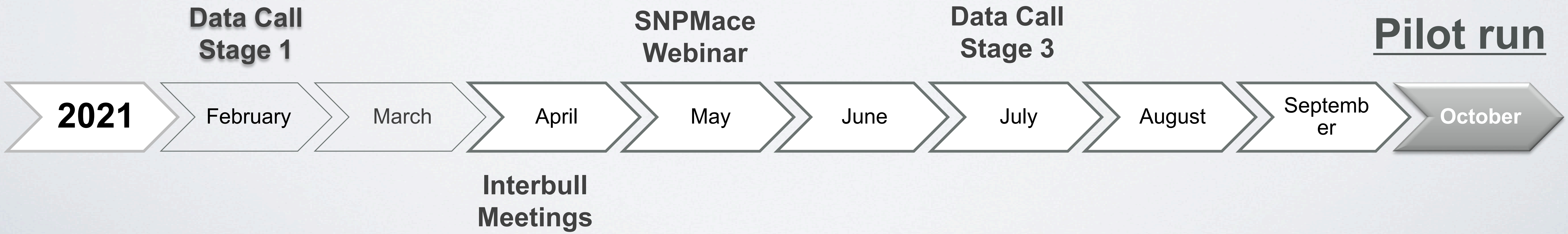
SNPMace data call

Feb 2021 – Data call: focused on **production** and **cma** traits

Mar 2021 – Stage 1: Collecting information on genomic models and set of SNPs used at national level by countries interested in joining the pilot run

May 2021 – Stage 2: Review the received information and provide recommendations regarding the data preparation - SNPMace webinar

Jul 2021 – Stage 3: Participating countries providing input data to Interbull Centre



Data Call Stage 1

SNPMace Webinar

Data Call Stage 3

Pilot run

2021

February

March

April

May

June

July

August

September

October

Interbull Meetings

Data call – Stage 1 (completed)

8 HOL Organizations

- National genomic evaluation model
- Set of SNPs used at national level
- animals’ categories in reference population and criteria for inclusion

Data Call Stage 1

SNPMace Webinar

Data Call Stage 3

Pilot run

Interbull Meetings

Country	N SNPs	Genotyped and phenotyped Females*	
		PROD	CMA
A	45,613	**250,000**	**185,000**
B	46,342	**100,000**	**100,000**
C	53,469	270,000	180,000
D	50,855	**55,000**	**23,500**
E	42,275	63,000	n.a.
F	46,216	n.a.	n.a.
G	37,996	**270,000**	n.a.
H	46,161	36,000	15,000
Total		1,044,000	503,500

* In bold females included in reference population

Data call – what’s next: stages 2 & 3

Interbull Centre is currently working on the **creation of a reference sets of SNPs** (avoiding duplicates), comparing the national sets ones:

- ✓**“Intersection” set**: including only SNPs that are present in **all** the countries’ sets
- ✓**“Union” set**: including SNPs that are present in **at least one** country set

Several different possible **SNPMace scenarios**

Data call – what’s next: stages 2 & 3

Stage 3 provides countries to submit of the following **data**:

I. **National SNP** effect estimates gi (for each of the investigated trait);

II. **$Zi\ 'Ri-1Zi$ matrix** for a measure of prediction error (co)variances of the SNP effect estimates (using an *ad-hoc* version of MetaGS);

III. **Marker allele frequencies** of the reference SNP allele;

IV. **Variance** of the direct genomic values.

The data received from countries participating in the pilot run will not be shared by Interbull Centre among the participants.

Possible SNPMace scenarios

Basic – use of “*intersection*” set of SNPs

- Countries’ Z’Z and Z’y reduced to the list of “*intersection*” set of SNPs
- Predictions provided by Interbull Centre rescaled to national sets

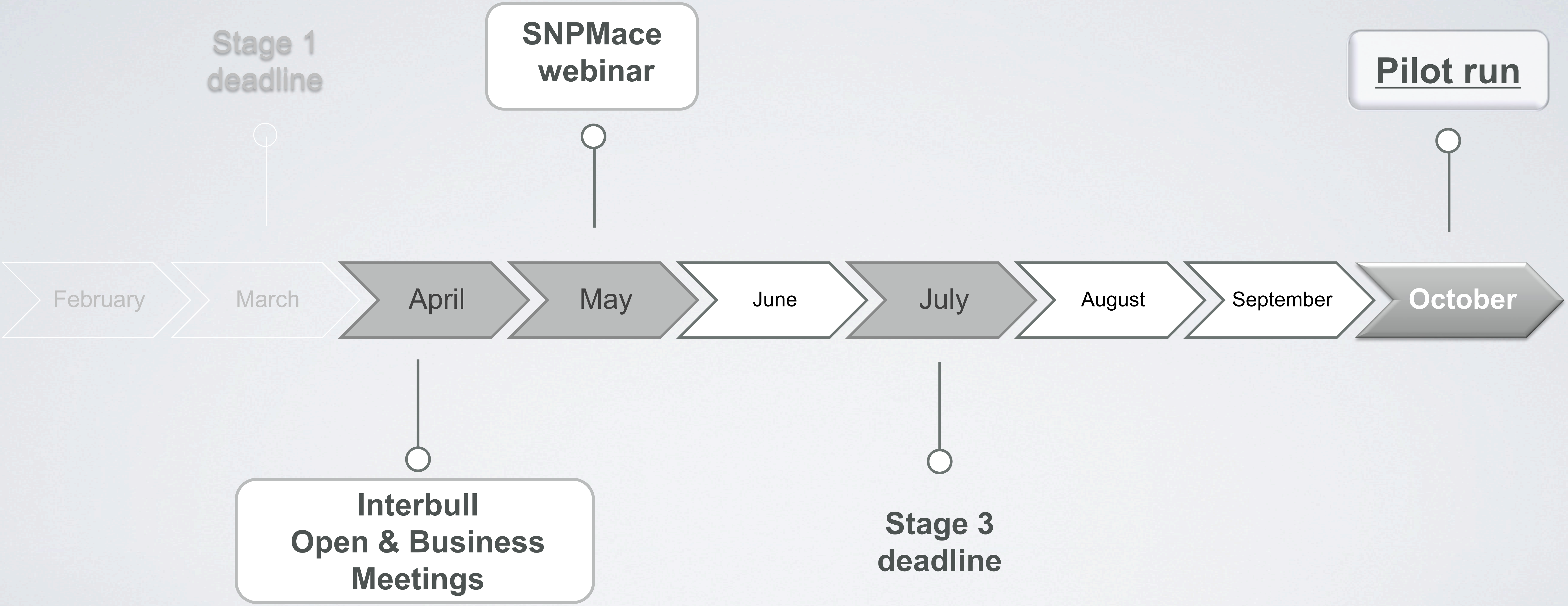
Improved – use of “union” set of SNPs and “T” matrix

- The genotypes of “n” bulls provided by each country used to design the SNPMace reference T matrix and for imputation of SNPs missing from the “union” set.
- These genotypes will not be shared with other Service Users.

Optimal - use of “union” set of SNPs only (investigation ongoing)

- Approximation of the Z’Z full matrix merging the Z’Z provided by each country, without using any additional genotypes.

SNPMace Time Line 2021



Stage 1 deadline

SNPMace webinar

Pilot run

February

March

April

May

June

July

August

September

October

Interbull Open & Business Meetings

Stage 3 deadline

SNPMace, expected benefits

September

October

November

December

2022
…

Expected benefits from SNPMace:

- Improves the accuracy of SNP effect estimates compared to within-country SNP solutions;
- No need of sharing/exchanging/pooling genotypes on individual bulls or cow;
- Use at International level of females’ information from national reference populations.

Please, feel free to contact Interbull@slu.se if you are interested in the SNPMace project.

References

- SNPMace – A meta-analysis to estimate SNP effects by combining results from multiple countries. Goddard, Jighly, Benhajali, Jorjani and Liu, 2018. INTERBULL BULLETIN NO. 54, Dubrovnik. https://journal.interbull.org/index.php/ib/article/view/1464
- SNPMace – A meta-analysis to estimate SNP effects across countries. Jighly, Benhajali, Liu and Goddard, 2019. INTERBULL BULLETIN NO. 55, Cincinnati. https://journal.interbull.org/index.php/ib/article/view/1493