SMARTER

SMAll RuminanTs breeding for Efficiency and Resilience

Research and Innovation action: H2020 – 772787
Call: H2020-SFS-2017-2
Type of action: Research and Innovation Action (RIA)
Work programme topic: SFS-15-2016-2017
Duration of the project: 01 November 2018 – 31 October 2022

Document specifying the exchange of pedigree, phenotypes and genotypes

Jean-Michel ASTRUC

IDELE

* Deliverable leader – Contact: Jean-Michel.Astruc@idele.fr

DELIVERABLE D6.1

Workpackage N°6

Due date: M36

Actual date: 15/12/2021

Dissemination level: Public

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 772787

About the SMARTER research project

SMARTER will develop and deploy innovative strategies to improve Resilience and Efficiency (R&E) related traits in sheep and goats. SMARTER will find these strategies by: i) generating and validating novel R&E related traits at a phenotypic and genetic level ii) improving and developing new genome-based solutions and tools relevant for the data structure and size of small ruminant populations, iii) establishing new breeding and selection strategies for various breeds and environments that consider R&E traits.

SMARTER with help from stakeholders chose several key R&E traits including feed efficiency, health (resistance to disease, survival) and welfare. Experimental populations will be used to identify and dissect new predictors of these R&E traits and the trade-off between animal ability to overcome external challenges. SMARTER will estimate the underlying genetic and genomic variability governing these R&E related traits. This variability will be related to performance in different environments including genotype-by-environment interactions (conventional, agro-ecological and organic systems) in commercial populations. The outcome will be accurate genomic predictions for R&E traits in different environments across different breeds and populations. SMARTER will also create a new cooperative European and international initiative that will use genomic selection across countries. This initiative will make selection for R&E traits faster and more efficient. SMARTER will also characterize the phenotype and genome of traditional and underutilized breeds. Finally, SMARTER will propose new breeding strategies that utilise R&E traits and trade-offs and balance economic, social and environmental challenges.

The overall impact of the multi-actor SMARTER project will be ready-to-use effective and efficient tools to make small ruminant production resilient through improved profitability and efficiency.

Table of contents

1 Summary

SMARTER in general and WP6 in particular aim at improving the efficacy of selection (of resilience and efficiency trait, but consequently on every trait of interest) by strengthening harmonisation and international collaboration. Across country genetic evaluation is a way to achieve this goal. This implies pooling together pedigree, phenotypes and genotypes data from different countries, different database. And this implies that if any pedigree connections do exist, we must find out them through the data. Format files for pedigree, phenotypes, genotypes, breeding values (EBV) and parameters (model used in domestic evaluations) were proposed and adopted by the partners working in WP6. This formats were used to produce the files necessary for running the international evaluation in dairy sheep, dairy goats and meat sheep. From the case-studies, feedback from the teams in charge of editing the data were collated in order to learn the lessons and to give recommendations for the future. Because of the poor connection in most of sheep and goats populations, a key point is the pedigree file and the different possible aliases of the ID to help tracking the connections. As a co-product of these formats, a proposition of codification of the breeds and the traits was proposed.

The file formats and the recommendations might be used if an international evaluation is implemented following the SMARTER project.

2 Introduction

The ambition of WP6 was basically (i) to perform the first across-country genomic evaluations in small ruminants by pooling phenotypic and genomic data and creating new shared reference populations in sheep and goats; and (ii) to create an international initiative that would facilitate, encourage and motivate the exchange of information, know-how and data (phenotype, genotype and pedigree) for international cooperation and improved breeding for resilience and efficiency in small ruminants.
To achieve this goal, WP6 undertook different actions to build the pre-requisite to enable reaching the ambition. One of this pre-requisite is the definition of common standards for exchanging data in order to implement across-country evaluations (during the project, as proof of concept, and hopefully after the project in routine evaluations).
The definition of the file format was based on the existing situation in Interbull (dairy cattle) and Interbeef (beef cattle), with an adaptation to small ruminants. We collectively defined formats for exchanging 5 kind of files:

- Pedigree (animal international ID, sire and dam international IDs, birth date, national ID, different aliases existing (genealogical ID, non-official ID). For that we first defined the structure of the international ID.
- Phenotypes (international ID, trait, value of the traits, value of the environmental effects).
- Genotypes (international ID, SNP, alleles).
- EBV (international ID, EBVs and reliabilities in the domestic evaluation).
- Parameters (for each trait: heritability and the model used in the domestic evaluation).

These 5 files are described in section 3. The section 4 lists the actual exchange of data done within the project for the case-studies of genetic evaluation. The section 5 of the deliverables presents the feedback on the file formats and the data exchanged following their use in order to give recommendations and improve the exchanges of data in the future.

3 Format file for exchanging pedigree, phenotypes and genotypes

3.1 International ID

The international ID contains in itself the breed, the country, the sex and the ID number

- Breed on 3 letters
- Country on 3 letters
- Sex on 1 letter (“M”,”F”)
- IDNumber on 16 digits including (for European countries) the country code at the beginning (2-letters for country code: IE / FR/ IT / CA / ES / GR / HU / NO / GB /CH / UY, etc).

All ID numbers: registration numbers, right justified, leading blanks as zeros

Example: CHAFRAM000FR12345678901

In this example, the original ID is FR12345678901. This original ID on 13 digits is right justified and completed by three ‘0’ to obtain an IDNumber on 16 digits.

3.2 Pedigree file

a) Name= pedigree_*species_production_country_version_date*.txt

(ex.: pedigree_goat_dairy_FRA_v1_20200215.txt)

Species = ‘sheep’ or ‘goat’
Production = ‘meat’ or ‘dairy’
Country on 3 letters (see below)
Version = ‘v1’, ‘v2’, etc
Date=yyyymmdd

b) Flat file. With delimiter “;”.

c) Description

19 columns, some from Interbull format, some useful for SMARTER tracing back the link across country.

Columns 1, 2, 3, 4, 6, 7, 8 are from Interbull format (from Interbull format, we have removed ‘Status’ & ‘Birth date of 1st AI daughters’ which are useless for SMARTER).

Columns 5, 9 to 19 are specific SMARTER data.

N°	Data	Type & length	Comment	Example
1	International ID animal	(1)	Interbull format	Please see below
2	International ID sire	(1)	Interbull format	
3	International ID dam	(1)	Interbull format	
4	Birth date	YYYYMMDD		20120425
5	Year of birth	YYYY	Useless if birth date known	2012
6	Name of animal	Free		
7	Country sending data	(2)		FRA
8	National ID animal	Free (as it is in the country)	Official ID in the country of origin	FR45512540012
9	Animal_ID_alias1	Free	Facultative, if any	CH2507
10	Animal_ID_alias2	Free	Facultative, if any	
11	Animal_ID_alias3	Free	Facultative, if any	
12	National ID sire	Free (as it is in the country)	Official ID in the country of origin	FR45234290120
13	Sire_ID_alias1	Free	Facultative, if any	CH1247
14	Sire_ID_alias2	Free	Facultative, if any	
15	Sire_ID_alias3	Free	Facultative, if any	
16	National ID dam	Free (as it is in the country)	Official ID in the country of origin	FR45512520001
17	Dam_ID_alias1	Free	Facultative, if any	
18	Dam_ID_alias2	Free	Facultative, if any	
19	Dam_ID_alias3	Free	Facultative, if any	

(1) BreedCountrySexIDNumber = Interbull format

(2) International codification on 3 digits (CAN, FRA, IRL, ITA, ESP, GBR, NOR, GRC, HUN, CHE, URY, etc.)

IMPORTANT GUIDELINES:

- All animals are required
- All animals that are declared as sire or dam (in cols 2 and 3) must be declared as animal in col 1.
- Missing value are left as blank
- in terms of number of generations included in the pedigree file, follow the same rules as in your domestic evaluation
- If some national IDs have two sexes in practice (i.e. they are sire and dam), the suggestion is:
  - 1) If possible, avoid this situation by removing one of the 2 sexes.
  - 2) If not possible or difficult, create two animals, one with M and other with F in the international ID

3.3 Phenotypes file

a) Name= phenotype_*species_production_country_version_date*.txt

(ex.: phenotype_goat_dairy_FRA_v1_20200215.txt)

Species = ‘sheep’ or ‘goat’
Production = ‘meat’ or ‘dairy’
Country on 3 letters (see below)
Version = ‘v1’, ‘v2’, etc
Date=yyyymmdd

b) Flat file. With delimiter “;”.

c) Description

One single file is required, even if there are several phenotypes
9 columns, same as in Interbeef format + weight of performance

N°	Data	Type & length	Comment	Example
1	Trait	(1)		MYI
2	Breed of evaluation	(2)		MTR
3	Country sending data	(3)		FRA
4	International animal ID	(4)	Interbull format	
5	Herd/flock	Free (as it is in the country)	Official ID in the country of origin	FR64124012
6	Dependent variable	Free		345
7	Statistical weight of the performance	(5) 0 to 1		1
8	Number (n) of environmental effects included in the national model			4
9	Environmental effects recoded (n times)	(6) Free		4 fields (e.g. 245;14;8;67)

(1) Codification for TRAIT in dairy: MYI=milk yield, FYI=fat yield, PYI=protein yield, FCO=fat content, PCO=protein content, SCC=somatic cell count, SCS=somatic cell score, LON=longevity, CWD=chest width, FAN=feet angle, TLN=teat length, TSH=teat shape, TAN=teat angle, UFR=udder furrow, UDP=udder depth, UCL=udder cleft, UPR=udder profile, UAT=udder attachment, UFP=udder floor position, RUA=rear udder attachment, FUA=fore udder attachment, TOR=teat orientation, other…
(1) Codification for TRAIT in meat: PRW=pre-weaning weight, WEW=weaning weight, POW=post-weaning weight, GRO=growth rate between weaning and post-weaning weight, UMD=ultrasonic muscular depth, UFD=ultrasonic fat depth, LIT=litter size, A1L=age at 1st lambing, LON=ewe longevity, LMO=lamb mortality, LAM=lameness
(2) According to the traits, it will be important to precise in which way it is recorded. For example, TAN might have a note 1 for horizontal or vertical position (reversed scale across country).
(3) If crossbred, put ’XXX’
(4) International codification on 3 digits (CAN, FRA, IRL, ITA, ESP, GBR, NOR, GRC, HUN, CHE, URY, etc.)
(5) BreedCountrySexIDNumber = Interbull format
(6) Weight: weight of the performance in domestic evaluation. For example, parity 1 might be 1, parity 2 might be 0.9 or whatever.
(7) Contemporary group should come as the first effect

3.4 Genotypes file

a) Name=genotype_*species_production_country_version_date*.txt
(ex.: genotype_goat_dairy_FRA_v1_20200215.txt)

Species = ‘sheep’ or ‘goat’
Production = ‘meat’ or ‘dairy’
Country on 3 letters (see below)
Version = ‘v1’, ‘v2’, etc
Date=yyyymmdd

b) Flat file. With delimiter “;”.

c) Description

N°	Data	Type & length	Comment	Example
1	International animal ID	(1)	Interbull format	
2	SNP name	International name		OAR1_110509088.1
3	Allele A/B	Char,2		AB,--

(1) BreedCountrySexIDNumber = Interbull format

3.5 Parameter file

1. Name= parameter_*species_production_country_version_date*.txt
(ex.: parameter_goat_dairy_FRA_v1_20200215.txt)

Species = ‘sheep’ or ‘goat’
Production = ‘meat’ or ‘dairy’
Country on 3 letters (see below)
Version = ‘v1’, ‘v2’, etc
Date=yyyymmdd

2. Flat file. With delimiter “;”.

3. Description

11 columns, same as in Interbeef format

N°	Data	Type & length	Comment	Example
1	Trait	(1)		MY
2	Breed of evaluation			MTR
3	Country sending data			FRA
4	Trait heritability	Free (as it is in the country)		30

5	Min. number of obs. per CG	Free		5
6	Maternal genetic effect fitted in the model	Y/N		N
7	Maternal permanent environmental effect fitted in the model	Y/N		N
8	Permanent environmental effect fitted in the model	Y/N		Y
9	Number (n) of environmental effects included in the national model			4
10	Environmental effects recoded (n times)	(2)	Name of effect in clear	Flockxyear;mth of lambing;interval btw lambing;parity
11	Type of effects (contemporary group, fixed, covariate, random) (n times)	(2) Free (as it is in the country)	F, C, R	F;F;F;F

(1) Codification for TRAIT in dairy: MYI=milk yield, FYI=fat yield, PYI=protein yield, FCO=fat content, PCO=protein content, SCC=somatic cell count, SCS=somatic cell score, LON=longevity, CWD=chest width, FAN=feet angle, TLN=teat length, TSH=teat shape, TAN=teat angle, UFR=udder furrow, UDP=udder depth, UCL=udder cleft, UPR=udder profile, UAT=udder attachment, UFP=udder floor position, RUA=rear udder attachment, FUA=fore udder attachment, TOR=teat orientation, other …

(2) Codification for TRAIT in meat: PRW=pre-weaning weight, WEW=weaning weight, POW=post-weaning weight, GRO=growth rate between weaning and post-weaning weight, UMD=ultrasonic muscular depth, UFD=ultrasonic fat depth, LIT=litter size, A1L=age at 1st lambing, LON=ewe longevity, LMO=lamb mortality, LAM=lameness

(3) Contemporary group should come as the first effect

3.6 EBV file

a) Name= EBV_*species_production_country_version_date*.txt

(ex.: EBV_goat_dairy_FRA_v1_20200215.txt)

Species = 'sheep' or 'goat'
Production = 'meat' or 'dairy'
Country on 3 letters (see below)
Version = 'v1', 'v2', etc
Date=yyyymmdd

b) Flat file. With delimiter ";".

c) Description

One single file is required, even if there are several phenotypes

10 columns

N°	Data	Type & length	Comment	Example
1	Trait	(1)		MY
2	Breed of evaluation			MTR
3	Country sending data	(2)		FRA
4	Genetic/genomic evaluation	Num,1		1=genetic 2=genomic
5	International animal ID	(3)	Interbull format	
6	Date of evaluation	YYYYMMDD		20190901
7	(G)EBV			368
8	Exact reliability	Num,2	If available	
9	Approximated reliability	Num,2	If available	52
10	Number of daughters (if males)	Num,5		590

(4) Codification for TRAIT in dairy: MYI=milk yield, FYI=fat yield, PYI=protein yield, FCO=fat content, PCO=protein content, SCC=somatic cell count, SCS=somatic cell score, LON=longevity, CWD=chest width, FAN=feet angle, TLN=teat length, TSH=teat shape, TAN=teat angle, UFR=udder furrow, UDP=udder depth, UCL=udder cleft, UPR=udder profile, UAT=udder attachment, UFP=udder floor position, RUA=rear udder attachment, FUA=fore udder attachment, TOR=teat orientation, other …

(5) Codification for TRAIT in meat: PRW=pre-weaning weight, WEW=weaning weight, POW=post-weaning weight, GRO=growth rate between weaning and post-weaning weight, UMD=ultrasonic muscular depth, UFD=ultrasonic fat depth, LIT=litter size, A1L=age at 1st lambing, LON=ewe longevity, LMO=lamb mortality, LAM=lameness

(6) International codification on 3 digits (CAN, FRA, IRL, ITA, ESP, GBR, NOR, GRC, HUN, CHE, URY, etc.)

(7) BreedCountrySexIDNumber = Interbull format

3.7 Breed code

Breed wording	Breed code
Suffolk	SUF
Blanche du Massif Central	BMC
Tsigai	TSI
Charollais	CHA
Belclare	BEL
Hungarian Merino	hMER
Vendéen	VEN
Texel	TEX
Norwegian White Sheep	NWS
Creole	CRE
Merino	MER
Corriedale	CRR
Lacaune	LAC
Manech tête rousse	MTR
Manech tête noire	MTN
Basco-Béarnaise	BAB
Corse	COR

Latxa cara rubia	LCR
Latxa cara negra	LCN
Assaf	ASS
Alpine	ALP
Saanen	SAA
Crossbreds	XXX

3.8 Trait code

Trait wording	Trait code
Milk yield	MYI
Fat yield	FYI
Protein yield	PYI
Fat content	FCO
Protein content	PCO
Somatic cell count	SCC
Somatic cell score	SCS
Longevity	LON
chest width	CWD
feet angle	FAN
teat length	TLN
teat shape	TSH
teat angle	TAN
udder furrow	UFR
udder depth	UDP
udder cleft	UCL
udder profile	UPR
udder attachment	UAT
udder floor position	UFP
rear udder attachment	RUA
fore udder attachment	FUA
teat orientation	TOR
Pre-weaning weight	PRW
Weaning weight	WEW
Post-weaning weight	POW
Growth rate between weaning and post-weaning weight	GRO
Ultrasonic muscular depth	UMD
Ultrasonic fat depth	UFD
Litter size	LIT
Age at 1st lambing	A1L
Lamb mortality	LMO
Lameness	LAM

4 Files actually exchanged

4.1 Dairy sheep

French (Manech breeds) data sent to INRAE

	Manech tête rousse	Manech tête noire
Pedigree	573,447	158,009
Phenotypes (milk yield)	543,929	146,132
Phenotypes (Somatic cells count)	236,592	33,655
Genotypes	5,567	956

Spanish (Latxa breeds) data sent to INRAE

	Latxa cara rubia	Latxa cara negra
Pedigree	153,765	68,830
Phenotypes (milk yield)	144,993	65,060
Genotypes	1,306	543

4.2 Meat sheep

French data sent to TEAGASC

	Charollais	Vendéens
Pedigree	697,528	917,486
Phenotypes (growth)	128,922	8,576
Phenotypes (litter size)	152,230	160,441
Phenotypes (post weaning weight)	128,922	8,576
Phenotypes (weaning weight)	207,555	195,152
Genotypes	-	-

Irish data sent to TEAGASC

	Belclare	Charollais	Suffolk	Texel	Vendéens
Pedigree	41,501	120,751	85,231	186,659	20,120
Phenotypes (40 day weight)	23,183	25,477	17,427	26,201	4,193
Phenotypes (wean weight)	17,707	27,313	29,806	39,679	5,256
Phenotypes (scan weight)	8,660	22,269	15,977	33,775	5,452
Phenotypes (US muscle depth)	3,167	18,834	25,617	40,881	6,002
Phenotypes (US fat depth)	3,167	18,739	25,402	40,703	5,972
Phenotypes (ewe mature weight)	2,525	3,814	2,849	7,232	686
Phenotypes (single lambing ease)	3,618	10,956	14,049	25,313	2,045
Phenotypes (multiple lambing ease)	25,334	41,472	42,693	66,111	9,748
Phenotypes (birth weight)	27,727	49,065	53,039	84,253	10,658

Phenotypes (lamb survival)	28,952	52,428	56,742	91,424	11,793
Phenotypes (number of lambs born)	16,823	47,697	28,871	80,973	8,190
Genotypes	-	-	-	-	-

British data

	Charollais	Suffolk	Texel
Pedigree			
Phenotypes (early life weight)	16,452	12,302	21,480
Phenotypes (scan weight)	7,778	7,736	13,219
Phenotypes (muscle depth)	6,971	7,519	12,619
Phenotypes (fat depth)	6,504	7,383	12,527
Genotypes	-	-	-

4.3 Dairy goats

French data sent to INRAE

	Alpine	Saanen
Pedigree	1,086,563	876,178
Phenotypes (milk yield)	2,590,558	1,956,235
Phenotypes (udder traits)	326,424	203,667
Genotypes	2,968	2,009

Italian data sent to INRAE

	Alpine	Saanen
Pedigree	68,359	82,479
Phenotypes (milk yield)	170,963	189,500
Phenotypes (udder traits)	22,756	26,549
Genotypes	1,061	338

Canadian data sent to INRAE

	Alpine	Saanen
Pedigree	7,938	4,207
Phenotypes (milk yield)	7,414	3,947
Phenotypes (udder traits)	5,235	2,443
Genotypes	793	903

Swiss data sent to INRAE

	Alpine	Saanen
Pedigree	25,379	25,748
Phenotypes (milk yield)	66,795	64,389
Phenotypes (udder traits)	-	-
Genotypes	1,280	503

5 Feedback from the exchanges

5.1 Dairy sheep

A document describing all the issues regarding the exchanged files found during the process of data edition and while running the across-country genetic evaluation in dairy sheep was produced. Below are the most important points.

1. Some genotyped animals were not in the pedigree.

In the case of the Latxa breeds, roughly 25 % of the genotyped animals were not in the pedigree. This is due to the fact that the pedigree extraction is based on the animals that have phenotypic information. Consequently, old animals and young animals with no information are not included.

For the case of the Manech breeds, roughly 12 % of the genotyped animals were not in the pedigree. These were all animals with the year of birth in 2020. The pedigree was built from data up to end of 2019 (hence without the year 2020), while the genotype file was built from the only version available on line, which was the more recent (including the year 2020).

In order to avoid this situation, it is important to take care that all genotyped animals are known in the pedigree file (as it should be for phenotyped animals).

2. IDs with spaces in the pedigree file of LCN.

There were IDs with spaces in between in LCN in the pedigree file.

Example: 'LCNESPF00ES1500 001745'

The spaces were removed (or filled?) to comply with the international format adopted by SMARTER ("BreedContrySexIDNumber").

3. Animals shared between Manech and Latxa have different IDs in the French and Spanish files.

Some animals in common (e.g. sires used both in Latxa and Manech) are identified differently in the French and Spanish pedigree. There is a table of equivalences provided by IDELE/Confelac/Neiker to deal with this.

4. Error in the second column of the phenotype file of Latxa (the one identifying the breed)

Column number 2 in the file of phenotypes of LCN should be "LCN". It was corrected.

5. Animals' identifications in the genotype files (MTR and MTN).

The initial version of the genotype files of Manech was built with the French ID instead if the international ID. The appropriate international identification of the animals must be respected.

6. Different subsets of markers kept for Latxa and for Manech.

The genotypic information received, both from Latxa and Manech, has already been through quality control checks and edition. Consequently, different subsets of markers were kept in each case: 35,579 for Latxa and 38,523 for Manech. We worked with the markers they had in common (27,063).

7. Problems with some SNPs from Latxa.

Latxa animals were genotyped using 2 different platforms (Affymetrix and Illumina). Approximately half of the animals were genotyped using the Illumina platform and the other half using Affymetrix, but for the same loci. When running a PCA there were differences between Affymetrix and Illumina platforms. There was a problem in the data edition (not any problem regarding the platform). SNP outliers had to be removed to solve the problem.

8. Problems with some genotyped Latxa animals.

Initially we received files with genotypic information for Latxa including animals with poor genotyped (low Call rate, etc). These animals should be deleted prior to exchange information. We requested the list of the animals with problems and eliminated them previous to the analyses.

9. Unknown Parent Groups (UPGs) in Latxa were made every 3 years and without differentiating the origin of the animal (whether if they were Manech or Latxa).

In the “official” Latxa evaluation unknown ancestors of Manech rams are assigned different UPGs to consider different selection progress. In the shared evaluation, a Manech individual was traced back to its Manech population. This creates a difference (that we consider to be small) between the Official Latxa evaluation and the Smarter evaluation where only Latxa records are included.

5.2 Dairy goats

In dairy goats, the main difficulty was to find the connections across countries.

The main difficulties are shown below through 5 examples.

Example 1

ALPCHEF0000FR0603705306 (Swiss pedigree)
ALPITAF0000FR0603705306 (Italian pedigree)
The country is expected be the country of origin. And thus to be identical in both cases.

Example 2

SAACHEF0000000FR4371IMF (Swiss pedigree)
SAAITAF000CHIMF00004371 (Italian pedigree)
The way the original ID was built has changed (4371 and IMF not at the same place; CH vs FR).

Example 3

ALPITAF00000ITBZ960557 (Italian pedigree)
ALPITAF000000ITBZ960557 (Italian pedigree)
Not the same length in Italian file (length 22 instead of 23)

Example 4

SAACHEM0000000CH1361212 (Swiss pedigree)
SAAITAM000000CH01361212 (Italian pedigree)
Country of origin different + '0' not at the same place.

Example 5:

SAACHEF0000000CH1198592 (Swiss pedigree)
SAAITAF0000CHGG01198592 (Italian pedigree)
Country of origin different + part of the ID number diffferent.

6 Conclusion

The first and main conclusion is that the format file that we defined globally work, were respected and allowed to run analyses with data pulled together. These formats can be used for further joint (across countries) analyses and as well in the perspectives of a routine international evaluation.

Nevertheless, the process of edition of the data allowed to underline some issues: The main one is the way the international number is built. For example, we must communicate on the fact that the country is the country of origin (thus should not change). In the future, in any exchange, it is of high importance, either not to change the ID number (for example within EU countries), or to keep the original ID in the database besides the new ID, so that it is possible to provide aliases ad thus better match same animals from different pedigree files.

We have been aware in the dairy sheep case of the importance in the edition of the genotypes, especially when different genotyping platform (Illumina vs Affymetrix), but also in the way of removing SNP because of quality control. Some recommendations (not normative but informative) should be given.

We did not discuss (because it was not really the topic in this deliverable) the phenotypes. Phenotypes, even with same "name", such as milk yield, appear to be differently computed from one country to another. This is probably one of the reason for the difficulties to estimate genetic correlation. A work for harmonisation is expected. This will be done in efficiency and resilience trait, which are more often novel traits. We should tend toward harmonisation in classical and existing traits as well.

Finally, we have proposed a codification for the breeds involved and the traits studied. This codification might be maintained and updated within ICAR and especially within the reference centre that SMARTER would like to define for the future.

7 Deviations or delays

The deliverable should initially be submitted on 31 October 2021. As the case-studies on across country evaluation were undertaken mostly in 2021, the feedback on the file format were not ready for the deadline. Therefore, we asked for and obtained a two-months delay.

8 Acknowledgements

All the people working in WP6 have participated in this collaborative work. However, we may express a special thanks to those who produced the data according to the file formats, and most of all to Carolina Garcia-Baccino, Marc Teissier and Aine O’Brien who handled the files for pooling data and who really got their hands dirty.