Guidelines to Join Interbeef International Genetic Evaluations

Version 2.3

InterBeef

24 October 2012

Contact:

Interbull Centre

Dept. Animal Breeding and Genetics SLU
Box 7023
S-75007 Uppsala
SWEDEN

Phone : 0046-18-671994
Fax : 0046-18-672848
E-mail : interbeef@hgen.slu.se
URL   : http://www.icar.org/pages/working_groups/wg_interbeef.htm

Table of Contents

1   INTRODUCTION

Countries willing to join InterBeef must be members of Interbull and support the ICAR guidelines for data collection (**http://www.icar.org/**). This document is a description of the necessary actions and relevant files required by a new country willing to join InterBeef.

2   The Animal International Identification

2.1   Format

A unique international identification for each animal is used in the international genetic evaluation system and will be referred to as the Animal International IDentification (AIID). The AIID is constructed using the Interbull rules (referred as Interbull format in the document) and its structure is:

Breed + Country + Sex + Identification number

- The field '**Breed**' of the AIID refers to **the breed of identification of the animal in its country of first registration** (in most cases, the country of origin): 3 characters as defined in the Interbull Breed Codes, available at www.interbull.org (see Reference section).

- The field '**Country**' of the AIID refers to **the country of first registration of the animal** (in most cases, the country of origin): 3 characters or 3 digits number as defined in ISO 3166 Codes, available at http://userpage.chemie.fu-berlin.de/diverse/doc/ISO_3166.html (see Reference section).

- The field '**Sex**' of the AIID refers to **the sex of the animal (M = male ; F = female)**

- The field '**Identification number**' of the AIID refers to the identification number of the animal in the country of first registration: 12 characters right justified with left blanks filled in with zero ("0").

```
Ex: CHAFRAM006327826864 refers to
```
- A Charolais   CHA
- First registered in France   FRA
- A male   M
- With the following identification in France   006327826864

```
Ex: BAQIRL125693456875 refers to
```
- A Blonde d'Aquitaine   BAQ
- First registered in Ireland   IRL
- A female   F
- With the following identification in Ireland   125693456875

2.2   Constructing the international identification number

A country should first construct the international identifications for the animals born on its soil. The general rule for this construction would be that a given identification number is unique within a country and a breed

When an animal (live animal or issued from AI straw or embryo) is sold abroad, the importing country should avoid re-registration; in case of re-registration the original ID and the national ID should be kept.

The animal '*IE231112150014*' is, for example, a pure breed Irish limousine female named 'SHRIFFTOWN RIPPER'. Ireland has constructed its AIID as followed: *LIMIRLF231112150014*.

The animal '*FR0384106449*' is a pure breed French Charolais bull named 'VLADIMIR'. That animal is also used in Ireland under the name 'VLADIMER' and the number '0384106449' and has also been re-codified '*108376849*' and '*VLR*'. France constructs its AIID as followed: *CHAFRAM000384106449*.

In the international identification number, hyphens, slashes, commas, dots, blank spaces or any other symbols are allowed.

The animal '*UK9373 70-2/-7*', for example, is a UK Salers female animal and its international identification number would be '*SALGBRF9373 70-2/-7*'

2.3 The New Web Interface at Interbull

Interbull has developed a new database, IDEA (Interbull Data Exchange Area) that allows register users to all the pedigree handling functionalities. IDEA is available at https://idea.interbull.org/

IDEA is exclusive for database users. To become a database user you need to belong to a national evaluation centre participating to the Interbeef evaluation.
If you want to join the Interbeef evaluation and therefore have access to IDEA, your national genetic centre needs to be associated to a username and a contact email address. To get your national genetic centre listed among the ones participating to the Interbeef evaluation, you have to:

Send an email to interbeef@hgen.slu.se providing information on your national genetic centre: Full name and address of the organization and full name and email address of the designated contact person.
Clearly state the breed(s) and trait(s) you wish to participate with. If your organization provides data as a joint evaluation including other countries, you clearly have to state for which other countries and breed(s) your organization provides pedigree and performance information.

Upon reception of your request, Interbull will send you by email a link to an IDEA-test environment together with a username and password so that you will be able to get acquainted with the IDEA functionalities and the user manual which is located under "Help".

The national pedigree file must be first checked with the CheckPedigree.py script provided by Interbull. The checking programs are available under "Software". **The zip files produced by the Python scripts represent the only valid infiles for the database; no other files will be accepted by the system.**
When your pedigree enters the database it undergoes a series of checks called "verification process" aiming at identifying the correct authoritative organization of each and every animals listed in your pedigree. In general the system has been built in such a way that it accepts information only if these comes from the authoritative organization. Therefore, whenever you submit pedigree data, the system first checks that the combination "country-breed" in the animal ID matches with the combination your organization "owns". If so the records are considered correct and stored in the database, if this is not the case then the records will be listed in the appropriate authoritative organization account waiting for verification, i.e. waiting for the authoritative organization to provide the correct pedigree information. In case the authoritative organization is a country or organization that does not yet participate to the Interbeef evaluation, then the animals will appear as verified by Interbull until eventually that country/organization would join the service.

Example:
Ireland submits the following pedigree records to the database:

```
601 CHAIRLM458962315289 CHAIRLM789456123652 CHAIRLF369852147852 1985 OMAR
IRL
601 CHAFRAM865412398745 CHAFRAM231658479851 CHAFRAF845693274125 1988 OSCAR
IRL
```

In the first case the animal is an Irish Charolais sent by Ireland. Ireland is therefore the authoritative organization for that animal so the record gets stored in the database.

In the second case the animal appears to be a French Charolais sent by Ireland. Ireland is no longer the authoritative organization for this bull therefore the system will send the record to the French account and will wait for France to provide its correct pedigree information.

3   The required files

There are 4 groups of files to be created by a country willing to join the international beef genetic evaluation:
- The 601 file representing the national pedigree file
- The 602 file representing the national performance file (limited until now to adjusted weaning weights, adjusted carcass weight and calving ease)
- The 603 file representing the parameter file.
- The Beef Form, see Reference section.

A country willing to join InterBeef needs to include at least one breed and one trait in the process of the international genetic evaluation. The objective of the present document is to help a new member country in the construction of these files.

NB: Since the verification of the International ID is done via database, national cross-reference files are no longer needed. For more information about IID verification, please refer to the IDEA user manual.

3.1   The different InterBeef steps

The construction of these files follows several steps:
- The first step for a country that wishes to join the international evaluation system is to upload the pedigree file into the IDEA database (https://idea.interbull.org/).
- Once the pedigree file (601) is uploaded into the database, the country will have to correct its own database according to the feedback obtained from the database regarding verified pedigree information from authoritative organizations. (see corrected foreign information in the IDEA user manual)
- Once this step has been fulfilled then countries will **proceed** sending performance file (file 602) and parameter file (file 603) to the Interbull ftp server.

3.2   General principles for file preparation:

- Numeric information must be right justified.
- Characters must be left justified.
- All characters should be in upper case.

3.3   The national pedigree file (601 file)

The 601 file is a **national pedigree file** and must contain pedigree information of all animals included in the performance file in an animal-sire-dam format. Sires, dams and ancestors must also have an entry in this file as an animal; unknown ancestors should be coded 'UUUUUUUUUUUUUUUUUUU'.

To ensure sufficient pedigree information it is recommended that pedigree information of animals born within a period equivalent to a minimum of three generation intervals should be included.
The 601 file is a file issued per country or organization providing pedigree information according to the format shown in Table 1. The 601 files can contain one breed or multiple breeds.

Table 1: 601 file format (total length = 107)

Field description	Label	Format	Start	Length	Status[a]	Note	Example
Record type	RTYPE	Char	1	3	M		*601*
International ID of ANIMAL							
Breed of the animal	ABREED	Char	5	3	M	I	*LIM*
Country of birth	AIDC	Char	8	3	M	II	*FRA*
Sex	ASEX	Char	11	1	M	III	*F*
ID code of the animal	AID	Char	12	12	M	IV	*8795005065*
International ID of Animal's SIRE							
Breed of the sire	SBREED	Char	25	3	O	I	*LIM*
Country of birth	SIIDC	Char	28	3	O	II	*FRA*
Sex	SSEX	Char	31	1	O	III	*M*
ID code of the sire	SIID	Char	32	12	O	IV	*95015085*
International ID of Animal's DAM							
Breed of the dam	DBREED	Char	45	3	O	I	*LIM*
Country of birth	DIIDC	Char	48	3	O	II	*FRA*
Sex	DSEX	Char	51	1	O	III	*F*
ID code of the dam	DIID	Char	52	12	O	IV	*1111001111*
Additional information							
Date of birth of animal	BDATE	Int	65	8	O	V	*20010130*
Name of animal	NAME	Char	74	30	O		*Faust*
Country sending information	RCOU	Char	105	3	M	II	*FRA*

[a]**Status:** M ⇔ Mandatory, O ⇔ Optional

Note

I	See Reference section for Breed codes
II	See Reference Section for Country codes – ISO 3166 Alpha-3
III	Sex : M ⇔ Male, F ⇔ Female only
IV	Identification Number in Interbull Format : Right justified, Leading blanks filled with Zero ("0")
V	YYYYMMDD format

3.4   The performance file (602)

All animals with a performance must have a line as animal in the 601 pedigree file. The performance file must be constructed in agreement with the 601 file and be consistent with the international identification of the animal.

It is expected that countries participating with data support the ICAR guidelines for beef recording.

As indicated in the ICAR recommendations, any group of animals kept for the same purpose and at the same location shall be regarded as a whole herd. For a performance record to be considered an official record, **the whole herd as defined above must be recorded**.

Each herd is identified with 15 digits identification, coded as characters:

- The first three digits correspond to the country, using the ISO 3166 Alpha-3 system, see Reference section.
- The next 12 digits correspond to the identification of the herd within the country. The original herd identification number should be used. The characters (letters or numbers) are right justified, the left blanks being filled with zeros.

All performance traits are associated with their environmental effects and complementary information (one observation line per measurement) as described in Table 2.

Table 2: 602 file format (all fields are mandatory)

Field description	Label	Format	Starting byte	Field ª	Note	Example
Record type	RTYPE	Char	1	3		*602*
Trait	FCODE	Char	5	3	I	*AWW*
Breed of evaluation	EBREED	Char	9	3	II	*LIM*
Country sending information	RCOU	Char	13	3	III	*FRA*
International ID of ANIMAL						
Breed of the animal	ABREED	Char	17	3	II	*LIM*
Country of birth	AIDC	Char	20	3	III	*FRA*
Sex	ASEX	Char	23	1	IV	*F*
ID code of the animal	AID	Char	24	12	V	*008795005065*
Twinning	TWI	Int	37	1	VI	*1*
Embryo transfer	ET	Int	39	1	VII	*0*
Herd	HERID	Char	41	15	VIII	*FRA000123456789*
Dependent variable	Y	Char	57	10	IX	*245*
Contemporary group	CG	Char	68	20	X	*333*
Number of environmental effects included in the national model	NENV	Int	89	3	XI	*4*
Environmental effect (n)ᵇ	ENV(n)	Char	93	10	XII	*2*

ª Variable total file length, depending on the number of environmental effects included in the model (Ex. ENV(1):length = 102; ENV(2):length = 113; ENV(3):length = 124; etc.)
ᵇ Repeat this field n times (n = 1,…,NENV), adding one (1) empty space between fields

Notes

I	See trait codes in Table 4 (abbreviations)
II	See Reference section for Breed codes
III	See Reference section for Country codes – ISO 3166 Alpha-3
IV	Sex : M ⇔ Male, F ⇔ Female only
V	Identification Number in Interbull Format : Alpha-numeric codes only, Right justified, Left blanks being filled with
VI	1 = single birth ; 2 = twin birth
VII	0 = no ET ; 1 = ET
VIII	Herd identification number corresponding to the herd included in the contemporary group. Format: 3 character country code + 12 digits
IX	Value observed for the trait in question. Ex. 245 Kg (no decimals)
X	Identification of the contemporary group
XI	Defines the remaining number of fields in the record. Additional fields = NENV
XII	Value of the respective environmental effect for the current record. Ex. season = 2 (out of 4 classes); birth weight = 36 (kg); sex of calf = F

3.5   Parameter file (603)

The parameter file (603) contains the name of two variables and the names of the environmental effects for each trait included in the performance file (602). This file must be provided together with the 602 file.

Table 3: Parameter file – one record per trait-breed combination (all fields are mandatory)

Field description	Label	Format	Starting byte	Field Length[a]	Note	Example
Record type	RTYPE	Char	1	3		*603*
Trait	FCODE	Char	5	3	I	*AWW*
Breed of evaluation	EBREED	Char	9	3	II	*LIM*
Country sending information	RCOU	Char	13	3	III	*FRA*
Reference age or class	REF	Int	17	5	IV	*200*
Trait heritability	H2	Int	23	3	V	*25*
Contemporary group	CG	Char	27	20	VI	*HYS*
How CG is fitted in the model (type of effect)	CGT	Char	48	1	VII	*R*
Maternal (genetic) effect fitted in the model	DAM	Char	50	1	VIII	*N*
Maternal permanent environmental effect fitted in the model	MPE	Char	52	1	VIII	*N*
Permanent environmental effect fitted in the	PEV	Char	54	1	VIII	*Y*
Number of environmental effects included in the national model	NENV	Int	56	3	IX	*4*
Environmental effect (n)[b]	ENV(n)	Char	60	20	X	*SEAS*
How ENV(n) is fitted in the model (type of	ENVT(n)	Char	81	1	VII	*F*

[a] Variable total file length, depending on the number of environmental effects included in the model (Ex. ENV(1):length = 81; ENV(2):length = 105; ENV(3):length = 128; etc.)
[b] Repeat these fields n times (n = 1,…,NENV), adding one (1) empty space between fields

Notes

I	Use the trait codes in Table 4 (abbreviations)
II	See Reference section for Breed codes
III	See Reference section for Country codes – ISO 3166 Alpha-3
IV	Reference value (age or class) used to adjust the dependent variable. Ex. reference national weaning age = 200 days
V	Trait heritability used in the national evaluation, expressed in a scale from 1-100
VI	Description of the contemporary group used. The most common are herd-year-season (HYS) and herd-year (HY). In case a different definition is adopted, describe it briefly (20 characters).
VII	Describes if the effect should be fitted as a fixed effect, a random effect or as a covariable in the model. Use the type of effect codes in Table 4 (abbreviations)
VIII	Declare if this specific effect (DAM, MPE or PEV) is fitted into the national model (Y = yes; N = no)
IX	Defines the remaining number of fields in the record. Additional fields = 2*(NENV)
X	Describes the nth environmental effect:

- Use the environmental effect codes in Table 4 (abbreviations)
- In case the trait is not yet included on Table 4, describe it briefly (20 characters)
- In case of polynomial effects, use the effect code followed by the order. Ex. $(SEAS)^2$ should be coded SEAS2
- In case of interactions, combine the effect codes in the same field. Ex. (AAWG)*(ASEX) should be coded as

3.6 Abbreviations

Table 4: List of abbreviations used in Interbull files and documentation

Type		Abbreviation	Meaning
1	General concepts	AI	Artificial Insemination
		ET	Embryo Transfer
		Kg	Kilogram
		MGS	Maternal Grand Sire
		NS	Natural Service
		lb	Pounds
2	Genetic evaluation methods	AM	Animal Model
		BLUP	Best Linear Unbiased Prediction
		FR	Fixed Regression
		MT	Multiple Traits
		RR	Random regression
		RP	Repeatability (model)
		REML	Restricted Maximum Likelihood
		ST	Single Trait
		SM	Sire Model
3	Estimates, values	BW	Breeding Worth
		c	Environmental correlation between records within sub-classes
		EBV	Estimated Breeding Value
		ETA	Estimated Transmitting Ability
		EPD	Expected Progeny Difference
		ME	Mature Equivalent, records are adjusted to mature cow yield basis
		PD	Predicted Difference
		PTA	Predicted Transmitting Ability
		RBV	Relative Breeding Value
		REL	Reliability/Repeatability of a sire proof, given either in the scale between 0 and 1 or as percentage
		SD	Standard Deviation
		SE	Standard Error
		TMI	Total Merit Index
		VAR	Variance
4	Population parameters	rg	Genetic correlation
		rp	Phenotypic correlation
		h2	Heritability of a trait
		t	Repeatability of trait
5	Traits of interest	acw	Adjusted carcass weight
		aww	Adjusted weaning weight
		cae	Calving Ease
		stb	Sitillbirth
		bwt	Birth Weight
		pli	Productive Life
6	Type of effect	X	Covariable (expressed as a numerical variable in a continuous scale)
		F	Fixed effect (class variable)
		R	Random effect
7	Model definition	CG	Contemporary group, comparison group
		Y	Dependent variable
		HYS	Herd-Year-Season
		HY	Herd-Year
		DAM	Maternal effects
		MPE	Maternal Permanent Environmental Effect
		PEV	Permanent Environmental Effect
8	Environmental effects	AACA	Age at calving
		AAWG	Age at weighting
		AAEV	Age at evaluation
		ASEX	Sex of the animal
		CSEX	Sex of calf
		BIRW	Birth weight
		CLAS	Classifier
		HERD	Herd (renumbered)
		PARI	Parity
		SEAS	Season
		SLAH	Slaughter house
		TWIN	Twinning
		YEAR	Calendar year

4 Filenames

All transferred files should be written in ASCII format and use character encoding: UTF-8

Specifically for the pedigree files, organizations will find access in IDEA to two Python scripts, CheckPedigree.py and CheckLinks.py. Aim of these scripts is to assure that pedigree and linking information are correct prior entering the database environment.
To run the script type at the prompt:

Python2.6 CheckPedigree.py –m {organization code[1]} –f {pedigree file name}

For more information about the CheckPedigree.py and the CheckLinks.py see Appendix I (a and b).

If the scripts detect no errors in the pedigree or link file then a zip file "**IB-ORGCODE-yyyymmddThhmmss.zip**." or "**IB202-org_code-yyyymmddThhmmss.zip"** would be created. The zip file produced by the Python scripts represent the infile for the Interbull database; no other file would be accepted.

For the time being before the database would be ready to accept performance and parameter file the nomenclature for these types of files would be:

COUNTRY.BREED.FILE.TRAIT_TYPE.DATE[2]

For example, a file from Ireland with performance for Limousine sent on 2010-01-18 would be called:
IRL.LIM.F602.AWW_20100118
A parameter file from the same country and breed would be called
IRL.LIM.F603.AWW_20100118

(1)   For the Org.Code, see Reference section.
(2)   date *yyyymmdd* format

References

ISO 3166 Country Codes: http://userpage.chemie.fu-berlin.de/diverse/doc/**ISO_3166**.html

Breed Codes for International Genetic Evaluation of dairy and beef cattle:
http://www.interbull.org/index.php?option=com_content&view=article&id=56&Itemid=77

Guideline to Form Beef and Form_Beef:
http://www-interbull.slu.se/Interbeef/General_info/framesida-general.htm

InterBeef link on Interbull website:
http://www-interbull.slu.se/Beef_Gen_Ev/framesida-beef.htm

Information about the program:

CheckPedigree.py will perform a series of checks on your pedigree data to insure that the data is all right. If no errors are detected then a zip file will be created. The zip file represents your checked pedigree file to upload to the Interbull Centre IDEA database.

The checks relate to:

- Check the international identification numbers (animal, sire and dam)
  - Correct three digit country code as in the ISO 3166 standard (no missing countries allowed)
  - Correct three digit breed code according to the Interbull breed codes
  - Correct construction of the numerical part of the ID (registration numbers, right justified, leading blanks as zeros)
  - Missing sires and dams shall be coded as UUUUUUUUUUUUUUUUUUU (i.e. with 19 U)
- Check the animal's birth date
  - Has to be reported in the format YYYYMMDD
  - If you know only the year of birth then enter it as YYYY0000
  - If you know year and month of birth then enter them as YYYYMM00
  - Missing birth dates are coded as 00000000 (or blanks or a single 0)
- Check that a male (or female) animal will eventually appear only as sire (or dam)
- Check for inconsistent duplicate records (different sire, dam or birthdate)
- Check that an animal is always younger than its parents and grandparents

Before Running the Program:

1. Install Python (version 2.5 to 2.7) if necessary
2. Create a working directory/folder
3. Download the CheckPedigree.py program from https://idea.interbull.org/software and copy it to your new directory
4. Copy your pedigree file to the working directory

Running the Program:

- Ensure there is a working network connection
- Use the command: **python CheckPedigree.py -m <ORGCODE> -f <filename>**
- Use your uppercase ORGCODE as shown on the upper right hand side of the IDEA page. Your organization code is reported within brackets beside the "**Logged in as**" information.
- The program checks its internal version with the value stored on the Interbull server. You will have to download the most recent version if there is a mismatch.

After Running the Program:

If no errors are detected, the pedigree file will be written into a zip file called **IB-ORGCODE-yyyymmddThhmmss.zip**. Upload the zip file to Interbull's data exchange site: https://idea.interbull.org/.

In case of errors, no zip file will be created. Please correct your data and re-run the program until the data successfully pass all required checks.

Specific information about your pedigree data, descriptive statistics and a summary of errors are written to the file **CheckPedigreeLog.txt**.

All errors are listed in detail in the file called **CheckPedigreeErrors.txt**. The following table describes the brief error messages more fully:

Error message	Description
Inconsistent duplicates	An animal appears twice with different sire, dam or birth date
Warning duplicates	An animal appears twice but with same sire, dam and birth date
Illegal character errors	The numerical part of the international ID is not valid
Breed-country error	The breed-country combination is not recognized - see file CheckPedigreeAuth.txt (created by the program)
Sex coding error	The sex code is neither M nor F
Parent sex error	A male animal (or a female) appears in the dam (or sire) column
Birth date errors	Malformed entry for birth date
Ancestor check	Animal appears older than its parents or grandparents - if a parent's birth date is unknown, grandparents are checked

Note

Please do not modify the program to circumvent any checks. Doing so would be pointless because the same checking routine is used again inside IDEA to double-check the pedigree file uploaded in the zip file.

If you need assistance, please do not hesitate to contact us at interbull@slu.se .

5   APPENDIX I b: CheckLinks.py

Information about the program:

CheckLinks.py will perform a series of checks in your format 202 link file to ensure that the data is all right. If no errors are detected then a zip file will be created. The zip file represents your checked links file ready to upload to the Interbull Centre IDEA database.

The checks relate to:

- Record type: must be 202
- Record length: must be 44
- International identification numbers (first ID, second ID)
  - Correct three digit country code as in the ISO 3166 standard
  - Correct three digit breed code according to the Interbull breed codes
  - Allowed breed-country combination
    - see file CheckLinksAuth.txt (created by the program)
  - Correct construction of the numerical part of the ID (registration numbers, right justified, leading blanks as zeros)
  - UUUUUUUUUUUUUUUUUUU (19 U's) is only allowed for the second ID
- Sex code of the two IDs (they must be the same and M or F)

Before Running the Program:

1. Install Python (version 2.5 to 2.7) if necessary
2. Create a working directory/folder
3. Download the CheckLinks.py program from https://idea.interbull.org/software and copy it to your new directory
4. Copy your pedigree link file to the working directory

Running the Program:

- Ensure there is a working network connection
- Use the command: **python CheckLinks.py -m <ORGCODE> -f <filename>**
- Use your uppercase ORGCODE as shown on the upper right hand side of the IDEA page.Your organization code is reported within brackets beside the "**Logged in as**" information.
- The program checks its internal version with the value stored on the Interbull server. You will have to download the most recent version if there is a mismatch.

After Running the Program:

If no errors are detected, the pedigree file will be written into a zip file called **IB202-org_code-yyyymmddThhmmss.zip**. Upload the zip file to Interbull's data exchange site https://idea.interbull.org/

In case of errors, no zip file will be created. Please correct your data and re-run the program until the data successfully pass all required checks.

A summary of errors is written to the file **CheckLinksLog.txt**.

All errors are listed in detail in the file called **CheckLinksErrors.txt**.

Note:

Please do not modify the program to circumvent any checks. Doing so would be pointless because the same checking routine is used again at the Interbull Centre to double-check the link file uploaded in the zip file.

If you need assistance, please do not hesitate to contact us at interbull@slu.se .