

Population data upload in EFABIS

1. About this document

1. Introduction

The purpose of the document is to describe a protocol for mass uploading of populations statistics data in the European Farm Animal Biodiversity Information System (EFABIS)

2. What is included in the document

The document describes the data input file containing populations statistics and the steps to be taken to load these statistics in EFABIS. It also deals with the data checks and the procedures for error handling.

3. What is not included in the document

The document describes the population data input file for EFABIS, but does not provide any assistance with the creation of this file. The preparation of the input file is responsibility of the respective National Coordinator.

4. Target audience

The document is aimed at National Coordinator (NC) who will prefer to upload all the population statistics data using file and not the web interface of EFABIS. These include inter alia NCs from countries having non-EFABIS national databases for monitoring of their farm animal genetic resources.

2. Terminology

EFABIS – European regional node of the FABISnet network

NC – National Coordinator for management of animal genetic resources

3. Data format

The population data must be provided via pipe separated ('|' character is used as a field delimiter) text file. The file name can be any file name in the operating system of the user and is ignored on upload. The file content must be encoded in utf-8 character encoding. Each line in the file contains population data for a single year for one breed. There must be only one line in the file per breed per year. In case there are several lines containing population data for the same breed and year, all lines will be processed, but the lines which occurs later in the file will overwrite the changes from the previous ones. Each line must be complete, i. e. it must contain 21 positions, however some of them could be empty.

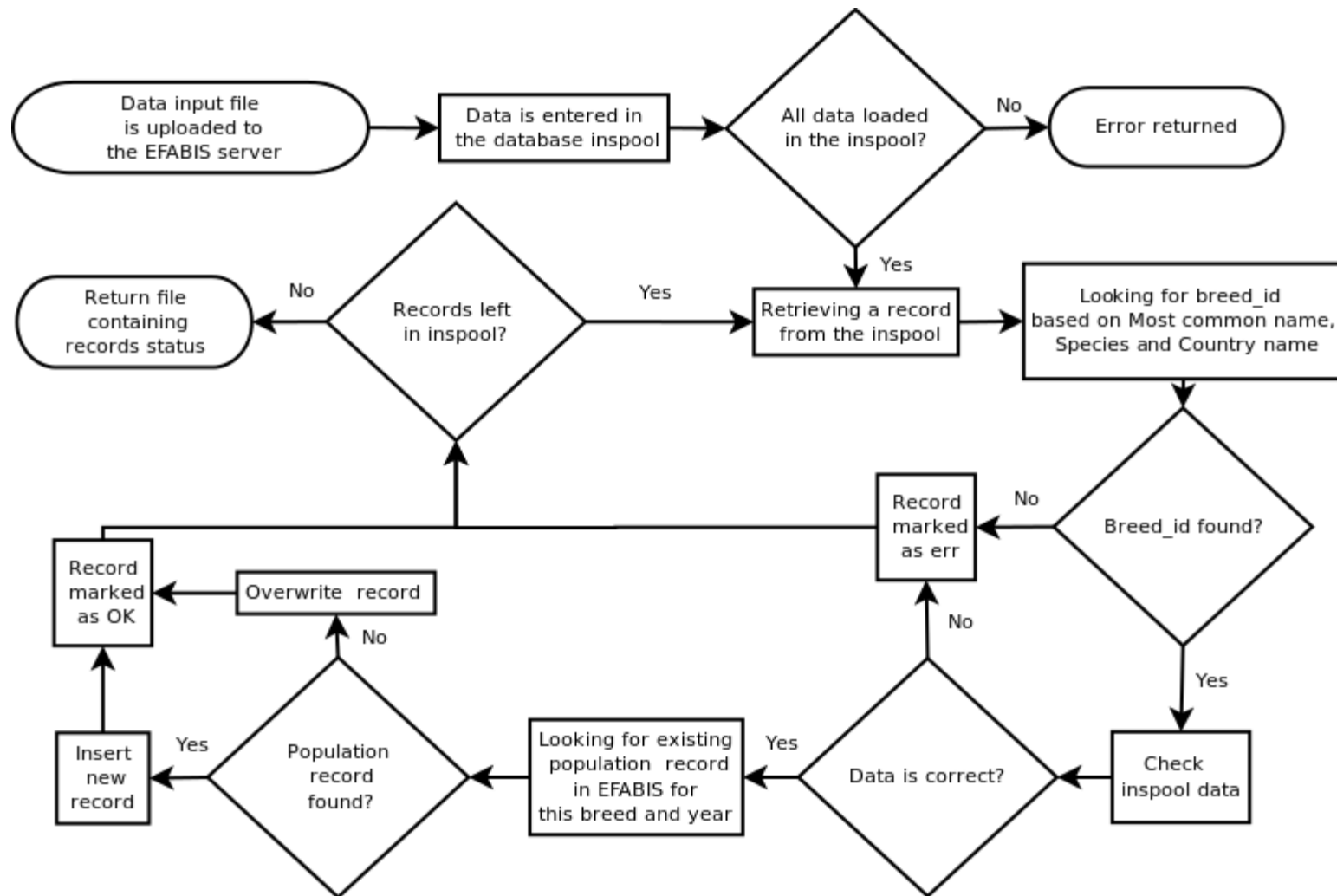
The order of the breeds and the years is irrelevant, the file is processed line by line. The file does not have a header, data starts from the first line. Each line has the following structure:

Pos	Field	Type	Description	Rules	Example
1	Most common name	String	The most common name for the breed in the country as registered in EFABIS	1. Cannot be empty	Brown Swiss
2	Species	String	The species name as registered in EFABIS	1. Cannot be empty	Cattle
3	Country name	String	The English name of the country as registered in EFABIS	1. Cannot be empty	Bulgaria
4	Year	Date	Year to which data on population size and structure refers.	1. Cannot be empty 2. A 4 digits year	2006
5	Population size (lower value)	Integer number	The total population size should include all animals in the population structure including young animals and cross-breeds. If the exact population size is known enter it both as lower and upper value. If the exact population size is not known please provide a reasonable guess. In this case provide a range and state that the total population is between the lower and the upper value of the range. For the year of extinction please enter 0 as lower range of the total population size.	1. Cannot be empty 2. Less than or equal to Population size (upper value) 3. Between 0 and 100000000 4. Greater than or equal to the sum Breeding males (no) + Breeding females (no).	5400
6	Population size (upper value)	Integer	The total population size should include all animals in the population structure including young animals and cross-breeds. If the exact population size is known enter it both as lower and upper value. If the exact population size is not known please provide a reasonable guess. In this case provide a range and state that the total population is between the lower and the upper value of the range. For the year of extinction please enter 0 as upper range of the total population size.	1. Cannot be empty 2. Greater than or equal to Population size (lower value) 3. Between 0 and 100000000	6300
7	Population figures based on	String	The source of information of population size	Allowed values: census at species level census at breed level survey at species level	census at breed level

Pos	Field	Type	Description	Rules	Example
				survey at breed level	
8	Trend	String	The trend of the total population size in the respective year	Allowed values: increasing stable decreasing unknown	stable
9	Trend females	String	The trend in the number of breeding females. This field is specific for EFABIS	Allowed values: increasing stable decreasing	decreasing
10	Trend no of breeding herds	String	The trend in the number of herds or flocks. This field is specific for EFABIS	Allowed values: increasing stable decreasing	increasing
11	Breeding males (no)	Integer number	Number of breeding males of reproductive age, not necessarily used for breeding.	1. Less than or equal to Population size (lower value) 2. Between 0 and 100000000	200
12	Breeding females (no)	Integer number	Number of breeding females in age of reproduction, not necessarily used for breeding.	1. Less than or equal to Population size (lower value) 2. Between 0 and 100000000	5000
13	Females registered in herdbooks (no)	Integer number	The number of females used for breeding and registered in herd book, studbook or similar registry.	1. Less than or equal to Population size (lower value) 2. Less than or equal to Breeding females (no) 3. Between 0 and	2434

Pos	Field	Type	Description	Rules	Example
				100000000	
14	Females bred pure (%)	Integer number	Percentage of females that are used for breeding and mated to sires of the same breed population.	1. Between 0 and 100	75
15	Herds (no)	Integer number	The number of all herds/flocks.		256
16	Herd Size (avg)	Integer number	The average number of animals per herd/flock.	1. Between 0 and 100000	85
17	AI used	String	Indicate whether Artificial Insemination (AI) is used	Allowed values: Yes No	Yes
18	Males in AI	Integer number	Enter the number of males used in artificial insemination.		45
19	Males in natural service	Integer number	Enter the number of males used in natural service.		123
20	Conservation programme	String	Is a conservation programme implemented for this breed population?	Allowed values: Yes No	Yes
21	Reliability	String	Indicate the reliability of the demographic data (population size, number of females, males, etc...)	Allowed values: very reliable reliable not reliable unknown	reliable

4. Data transfer diagram



5. Data processing notes

On the upload the file is stored in a temporary folder on the EFABIS server. All the lines from the file are loaded in a special table in the database called 'inspool'. There each line from the file is stored as one record. In case of error during the loading of a line in the inspool, the complete loading is aborted (nothing is stored in the database) and an error message is send to the web interface.

In the next step each record from the inspool is processed. If the record contains different number of columns then the defined in the protocol, a error is raised. The record is split in columns and based on the EFABIS database is searched for breed identification by Most common name, Country name and Species. If no such breed is found, a error is raised, the record is marked as erroneous and the next one is processed. No new breeds are registered by that protocol.

If there is such breed already registered in EFABIS, the data from the inspool record is checked against the business rules. In case of rule violation, the record is marked as erroneous, no updates are done in the database and the next inspool record is processed.

In the last step, the database is searched for existing population record for this breed and year. In case such record is found, it is overwritten with the new data. Only the cryo data for that year, which resides in the same record is kept intact. If there is no record for that breed and year, a new one is created in the populations_year table.

After all records from the inspool are processed, the web page returns a file containing 23 columns – the 21 original columns from the data input file and 2 more columns, the records status (OK or err), and in the last column the error message.