LANDESAMT FÜR UMWELT,
LANDWIRTSCHAFT
UND GEOLOGIE

Freistaat
SACHSEN

WebLOAD - A web frontend to create a consistent dataset from multiple textfiles in animal breeding

Müller, U., Fischer, R., VC Truong, C., Groeneveld, E. and Bergfeld, U.

63rd Annual Meeting of the European Federation of Animal Sience
August 27th - 31th, 2012, Bratislava, SLOVAKIA

Objective

Parameter estimates, genetic analyses, genomic selection -

all these methods require consistent data and

logically correct pedigrees.

Introduction

- information is electronically stored data

- ascii files, excel-tables, relational databases

- independent data management systems

- correctness of merged data can not be assumed

- after linking errors became obvious

Example

animal sale from Herdbook A to Herdbook B

	Herdbook A	Herdbook B
animal	2-32-10099-6	LR-1-32-55677
sire	1-32-10389	LR-1-32-10389
dam	2-32-10099	LR-2-32-10099
tatoo	6	6
breed	1	LR
owner	1032	B1
birth	15.01.2012	15.04.2012
daily gain	500	-
no born	-	12

Example - animal number system different

	Herdbook A	**Herdbook B**
animal	**2-32-10099-6**	LR-1-32-55677
sire	1-32-10389	LR-1-32-10389
dam	2-32-10099	LR-**2-32-10099**
breed	**1**	LR -> **1**
tatoo	6	**6**
animal (2nd)	1-2-32-10099-6	1-2-32-10099-6
-	breed in animal number	recode LR → 1 dam+tatoo

Example - different code for breed and owner

	Herdbook A	**Herdbook B**
breed	1	**LR**
owner	1032	**B1**

=> consistent dataset

animal;breed,breeder,daily_gain; no_born
1-2-32-10099-6;1;B1;500;12

Requirements for a tool

- result in compatible animal number systems

- possible to harmonize code of keys (DL -> 1)

- incorrect entries and logical errors have to be detected

- produce a consistent dataset

=> WebLOAD

- written in PERL
- configuration with XML/DTD

Preparation

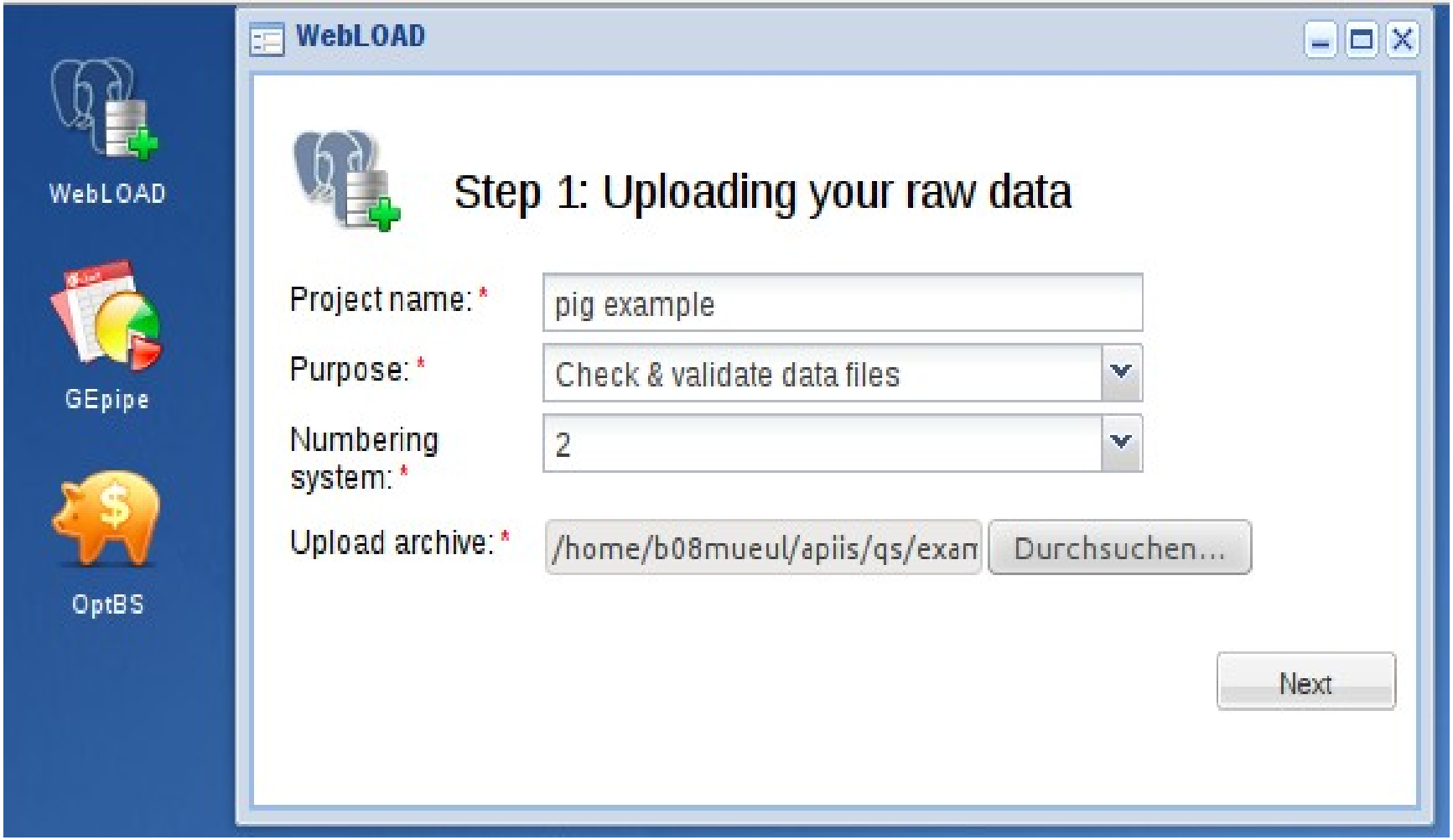
Herdbook_A.dat:

```
animal;          sire;          dam;          breed;tatoo;breeder;daily_gain
2-32-10099-6; 1-32-10389; 2-32-10099;1;      6;     1032;     500
...
```

Herbook_B.dat:

```
animal;          sire;          dam;          breed;tatoo;breeder;no_born
LR-1-32-55677; LR-1-32-10389; LR-2-32-10099;LR;    6;     B1;       12
...
```

WebLOAD I


WebLOAD
GEpipe
OptBS
WebLOAD
Step 1: Uploading your raw data
Project name: *
pig example
Purpose: *
Check & validate data files
Numbering system: *
2
Upload archive: *
/home/b08mueul/apiis/qs/exan
Durchsuchen...
Next

WebLOAD II

Step 2: Defining your data files

Herdbook_A.dat

Add a function	Add a function	Add a function	Add a function	Add a function	Add a function	Add a function	Add a function
System 1 (unique	System 2 (unique	System 1	System 1	Code	Char	Code	Number
animal	animal2	sire	dam	breed	tatoo	breeder	daily_gain
2-32-10099-6	1-2-32-10099-6	1-32-10389	2-32-10099	1	6	1032	500
2-32-10099-7	1-2-32-10099-7	1-32-10389	2-32-10099	1	7	1032	520
2-32-20021-1	1-2-32-20021-1	1-32-20389	2-32-20021	2	1	20	530
2-32-20021-2	1-2-32-20021-2	1-32-20389	2-32-20021	2	2	20	530

Herdbook_B.dat

Add a function	Add a function	Add a function	Add a function	Add a function	Add a function	Add a function	Add a function
System 2 (unique	System 2 (unique	System 2	System 2	Code	Char	Code	Number
animal	animal2	sire	dam	breed	tatoo	breeder	no_born
LR-1-32-55677	1-2-32-10099-6	LR-1-32-10389	LR-2-32-10099	LR	6	B1	12
LR-2-32-52983	1-2-29-20345-1	LR-1-32-10389	LR-2-29-20345	LR	1	B1	13
LR-2-32-52984	1-2-29-20345-5	LR-1-32-10389	LR-2-29-20345	LR	5	B1	11
LR-2-32-52985	1-2-29-20345-4	LR-1-32-10389	LR-2-29-20345	LR	4	B1	13

Step 4: Checking Keys and Units

Editor of Codes

Class	File	Position	Code	Count	Target Code	Short Name	Long Name
BREEDER	Herdbook_A.dat	'6'	'1032'	2	'B1'		
BREEDER	Herdbook_A.dat	'6'	'20'	3	'B2'		
BREEDER	Herdbook_B.dat	'6'	'B1'	6	'B1'		
BREED	Herdbook_A.dat	'4'	'1'	2	'1'	'LR'	'Landrace'
BREED	Herdbook_A.dat	'4'	'2'	2	'2'	'LW'	'Large White'
BREED	Herdbook_A.dat	'4'	'3'	1	'NULL		
BREED	Herdbook_B.dat	'4'	'LR'	6	'1'	'LR'	'Landrace'

Submit

Output

- codes harmonized

- animal numbers renumbered

- text file or normalized database

=> datasets are now consistent

Usage

webservice (http://...)

+ simple usage
- reduced functionality, time of transfer

appliance (virtual machine)

+ encapsulate application
- installation of virtualbox

Conclusion

WebLOAD: web based frontend to clean up heterogeneous data from different data sources.

Possibilities to recode and check keys and animal IDs

Output: text files or normalized database.

Released under GPL

Can be used as webservice or appliance

LANDESAMT FÜR UMWELT,
LANDWIRTSCHAFT
UND GEOLOGIE

Freistaat
SACHSEN

Thank you for your attention

Project Partners

FRIEDRICH-LOEFFLER-INSTITUT
FLI
Bundesforschungsinstitut für Tiergesundheit
Federal Research Institute for Animal Health

LANDESAMT FÜR UMWELT,
LANDWIRTSCHAFT
UND GEOLOGIE

Freistaat
SACHSEN

MSZV
Mitteldeutscher Schweinezuchtverband e. V.

Project Sponsor

Bundesanstalt für
Landwirtschaft und Ernährung

“The project is supported by funds of the Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.”