Management support and early warning system for national biodiversity databases in a network of national, regional (EAAP) and international (FAO) structures

THESIS
Submitted in partial fulfilment of the requirements for the degree

DOCTOR OF PHILOSOPHY
-Ph.D.-

In the field of Animal Breeding
at the University of Veterinary Medicine Hannover

by
Zhivko Ivanov Duchev
Stara Zagora, Bulgaria

Hannover, Germany 2006
Supervisor: Dir. & Prof. Dr. E. Groeneveld

Advisory committee: Dir. & Prof. Dr. E. Groeneveld
Prof. Dr. L. Kreienbrock
Prof. Dr. O. Distl

First Evaluation: Dir. & Prof. Dr. E. Groeneveld (Institute for Animal Breeding, Federal Agricultural Research Centre, Mariensee, Germany)
Prof. Dr. L. Kreienbrock (Institute for Biometry, Epidemiology and Information Processing, University of Veterinary Medicine Hannover, Germany)
Prof. Dr. O. Distl (Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Germany)

Second Evaluation: Prof. Dr. Gustavo Gandini (Department of Veterinary Sciences and Technologies for Food Safety, Faculty of Veterinary Medicine, University of Milan, Milan, Italy)

Date of oral exam: 30.10.2006
To my parents and grandparents.
Parts of the thesis have already been published or submitted for publication:

Duchev, Z.I. and Groeneveld, E.
Synchronization of APIIS based farm animal biodiversity systems.

Duchev, Z.I. and Groeneveld, E.
Improving the monitoring of animal genetic resources on National and International level.
Archives of Animal Breeding (Accepted for publication)

Duchev, Z.I., Distl, O. and Groeneveld, E.
Early warning system for loss of diversity in European livestock breeds.
Archives of Animal Breeding (Submitted for publication)
Further aspects have been presented at national or international conferences as oral presentations or as posters:

Duchev, Z.I. and Groeneveld, E.
Data synchronization among APIIS databases.

Duchev, Z.I. and Groeneveld, E.
Synchronization of farm animal databases in APIIS.

Duchev, Z.I. and Groeneveld, E.
Management support for national and regional coordinators in FABISnet.
Proceedings of the 8th World Congress on genetics applied to livestock production, Belo Horizonte, Brazil, 33-12 (2006).

Duchev, Z.I. and Groeneveld, E.
Fast establishment of animal data collection with RapidAPIIS.
Proceedings of the 8th World Congress on genetics applied to livestock production, Belo Horizonte, Brazil, 27-05 (2006).

Duchev, Z.I. and Groeneveld, E.
Fast establishment of animal data collection with RapidAPIIS.
Contents:

CHAPTER 1 INTRODUCTION ......................................................................................................................... 9
CHAPTER 2 SYNCHRONIZATION OF APIIS BASED FARM ANIMAL BIODIVERSITY SYSTEMS 31
CHAPTER 3 EARLY WARNING SYSTEM FOR LOSS OF DIVERSITY IN EUROPEAN LIVESTOCK BREEDS ..................................................................................................................................... 54
CHAPTER 4 IMPROVING THE MONITORING OF ANIMAL GENETIC RESOURCES ON NATIONAL AND INTERNATIONAL LEVEL ......................................................................................................................... 74
CHAPTER 5 GENERAL DISCUSSION ........................................................................................................ 96
CHAPTER 6 SUMMARY ................................................................................................................................1 0 6
CHAPTER 7 ZUSAMMENFASSUNG ........................................................................................................... 110
APPENDIX
Chapter 1

Introduction
Introduction

Domestication of the animal species began more than 12000 years ago beginning with dogs. Goats and sheep followed around 10000 years ago in the Near East and Asia (Zeder and Hesse, 2000), pigs around 9000 years (Guiffra et al., 2000) and cattle around 8000 years ago. Molecular studies suggest an even earlier start of the domestication (Dobney and Larson, 2006), for instance placing that of dogs at 135 000 years ago (Vila et al., 1997). This long process resulted in around 40 domesticated species, meeting around 30% of the food requirements of the human population (FAO, 1999). Livestock is a source of high value animal protein and micro-nutrients (Delgado et al., 1999). In the developing countries domesticated animals are also source of clothing, tools, transportation and farm power. In 1992 the Convention on Biological Diversity (United Nations, 2006), identified the domestic animal diversity as a genuine and important component of the global biodiversity, placing farm animal genetic resources under the custody of national governments.

Threats for erosion of animal genetic resources

With the expansion of the human population the demand for food increased dramatically - from 1971 to 1995 the consumption of meat and milk in developing countries increased by 70 and 105 millions metric tons respectively (Delgado et al., 1999). Amongst others, this increase in animal production was made possible by rapid advances in biotechnological techniques during the last century (Cunningham, 1999). Modern reproductive technologies for semen and embryo collection and artificial insemination allow selected individuals to produce large numbers of progeny. Conservation and transportation of germplasm resulted in a rapid dissemination of exotic foreign germplasm all over the world.

As a result, a small number of highly-productive breeds dominate the production all over the world in cattle, pigs and chicken (Hall and Ruane, 1993; Simon, 1984). However, the wide spread usage of these breeds has also its negative sides, as they push aside or even replace the local less-profitable indigenous breeds (de Haan et al., 1997; Philipsson, 1992; Tisdell, 2003), thereby reducing the genetic diversity.
A number of examples can be given here. In Finland the indigenous dairy cattle breeds - Eastern, Northern and Western Finncattle were almost totally replaced by the commercially more successful Finnish Ayrshire and Finnish Holstein-Friesian (Kantanen and Vilkki, 2005, Tapio et al., 2006). The pure Sahiwal cattle population in India and Pakistan was greatly reduced by crossing with Holstein and other exotic breeds (Philipsson, 1992). The crossing was aimed to improve the milk production and to create a new national strain - the Frieswal (Mudgal and Arora, 1994).

Moreover, with the globalization of breeding programs and semen sales, only a limited number of elite sires are used for insemination in dairy production, leading to increased inbreeding with a reduction in fitness (Falconer and Mackay, 1996), more specifically a decrease in survival and reproductive performance in cows (Smith et al., 1998).

A typical example of a world-dominating commercial breed is the Holstein-Friesian (HF) which represents more than 90% of all dairy cattle in USA and more than 60% in Europe. This breed also is used in production environments which it is not suitable for, for instance in smallholder dairy farms in the Kenya highlands (Bebe et al., 2003), where smaller dairy breeds are recommended. In 2004 the USA registered Holstein population comprised more than 3.7 million cows. However, the effective population size was estimated at only 60 (Hansen, 2006). Such a reduction of effective population sizes, which are result of extensive usage of a few elite bulls, were already noticed in the middle of the 1990s (Miglior and Burnside, 1995). As one possible solution of this problem avoidance of matings yielding inbreeding coefficient greater or equal to 6.25% was proposed (Miglior, 2000). In more general terms a mating strategy that keeps the inbreeding rate in the next generation below given value, e.g. 0.5% is proposed (Finland, 2006). One negative result of the increased inbreeding is the increase in stillbirths in U.S. Holstein (Meyer et al., 2001). The decline in fertility and survival in purebred Holsteins lead some commercial producers to produce crossbred cows by using imported semen from European breeds (Normande, Montbeliarde and Swedish Red) on their HF cows (Heins et al., 2006).
A similarly low effective population size has been reported for the Jersey and Brown Swiss in the USA, with an Ne estimated in 2004 as 31 and 32, respectively (Hansen, 2006).

The reduction in number of breeds due to the replacement by a small number of international commercial breeds and their increased inbreeding is not the only threat to the AnGR diversity. The changes in some production systems are leading to usage of crossbreeding (Rege and Gibson, 2003). The roles of certain livestock change or disappear in the short to medium term (Anderson, 2003). Urbanisation leads to changes in consumer’s taste for food, resulting in a reduced demand for traditional livestock products. The erosion of rural culture is another downside from the urbanisation and intensification, which leads to loss of traditional animal breeding practices.

A sudden change in environment like droughts, famine, disease outbreaks, war and other catastrophes may also lead to breed extinction or significantly reduce the population size (Anderson, 2003). Especially endangered by wars, disasters and epidemics are breeds located in a small area. Wars and other forms of socio-political instability can lead to livestock owners moving their stock out of their usual area, thus increasing the possibility of mixing with other breeds thereby potentially losing a location-specific breed. Natural disasters as floods, famine or tsunami can reduce drastically the population size. For instance the outbreak of Foot and Mouth Disease in 2001 in UK resulted in culling around 6 million animals (Scudamore and Harris, 2002).

**Benefits of farm animals diversity**

There are several reasons to preserve the diversity of the AnGR - genetic stock protection, social role, environmental role, cultural heritage, existence value, draught power (Mendelsohn, 2003; Bradford, 1999). The currently existing breeds are part of the global genetic stock and serve as a source of variation, which can be used for genetic improvement in adaptation and productivity (Rege and Gibson, 2003; Roosen et al., 2005). Even the non-profitable breeds of today may be valuable to meet future changes in the market, consumers demand or breeding objectives (Mendelsohn,
Therefore, the long-term goal of management of AnGR should be to preserve the fully adaptive potential of the species. The decisions and actions taken today, should not lead to a limitation of the possibilities to adjust future genetic levels of production and fitness (Notter, 1999).

Another reason for preserving genetic diversity of AnGR is the impact, which the breeds may have on the landscape or environment. Animal manures are used as a nutrient source, which increases the variety and activity of the soil biota (Altieri, 1999). Another important role of the livestock is the management of grassland biodiversity by grazing, e.g. (Rook and Tallowin, 2003; Pollock et al., 2005). Grazing enhances the sward structural heterogeneity by selective defoliation, nutrition cycle, treading and propagule dispersion (Rook et al., 2004). Livestock grazing can be also used to manage the habitat of other species (Tichit et al., 2005). However, grazing may have also negative impacts on some environments, e.g. heavy grazing significantly reduced vegetative cover on steep slopes in Ethiopian highlands (Mwendera et al., 1997).

The social role of the breed is more difficult to assess. Livestock is used for payments and as a wealth indication. In some societies livestock is part of the dowry and bride wealth. Some animals are still used for slaughtering on traditional feasts and religious ceremonies (Rege and Gibson, 2003) and are part of the religious beliefs. Livestock is also important in the social relations support in the poor societies. Livestock is used as loans, gifts or even as social security (IFAD, 2004). For example, the rich men in the Boran tribe in East Africa are expected to donate livestock to the poorer tribe members, thus improving the social security of the community members (IFAD, 2004). Livestock may be important in establishing the status of the farmer, as providers of employment of the farmer and/or family members, in contributing to gender equality (Riethmuller, 2002). The Sahelian farmers in Niger use livestock as a form of precautionary savings to protect themselves against the weather risk in crop production (Fafchamps and Gavian, 1996).

Livestock breeds can also be considered a cultural good, because they were for a long time an important part of the social life in the rural societies (Gandini and Villa,
Moreover, the typical products of the local breeds have become part of the religious and civic festivals and gastronomic traditions. The cultural value of a breed can be looked from two perspectives - value as 'historical witness' and as 'custodian of local traditions' (Gandini and Villa, 2003). The value of a breed can be assessed on the basis of an agricultural system and the farming techniques linked to the breed, the contribution to landscape formation, role in gastronomy, folklore and handicrafts. An example is the Valdostana breeds group - Valdostana Castana, Valdostana Pezatta Nera and Valdostana Pezatta Rossa. These breeds are linked to the Alpine summer pasture system, their economic relevance is derived from their milk, from which the branded 'fontina' cheese is produced, while their social relevance is reflected in the local tradition of selecting 'Regina del Latte' (Milk Queen) and 'Regina delle Corna' (Queen of Horns - for Valdostana Castana only) (Gandini and Villa, 2003). Another argument for preserving a breed as a part of the human culture is its existence value. The society or some of its members may want sentimentally to preserve for the future generations breeds that have been part of the past or the present (Mendelsohn, 2003).

Estimating the breeds risk status

During the years various systems for categorisation of the endangered breeds of livestock on a national (Ruane, 2000) and international level (Bodo, 1992; Avon, 1992; Loftus and Scherf, 1993; Simon and Buchenauer, 1993; Alderson, 2003; Reist-Marti \textit{et al.}, 2003) have been developed. These systems use different parameters and thresholds, which makes their comparison difficult. The EAAP-WGAGR used as a criterion the number of breeding females and males and the trend in number of breeding males (Maijala \textit{et al.}, 1984). As endangered they considered populations of less than 1000 female animals in cattle, 500 in sheep and goats and 200 in pigs. Furthermore, a breed was considered endangered if the number of breeding females is between 1000 and 5000 in cattle, 500-1000 in sheep and goats, 200-500 in pigs and the number of breeding males is decreasing or already under 20.

The Commission of the European Communities (CEC) for endangered ruminant breeds proposed criteria based on the number of reproducing females and the trend
in number of breeding animals during the last 5 years (Avon, 1992). This system has also different thresholds for the cattle, sheep and goat breeds.

The European Union changed several times the thresholds for defining breed at risk, most recently in the EEC Regulations 1257/99 and 445/02. There, the thresholds under which a local breed is considered as being in danger of being lost to farming, are (number of breeding females): Cattle 7,500, Sheep 10,000, Goats 10,000, Equidae 5,000, Pigs 15,000. The number is calculated, for all EU Member States, for females of the same breed available for purebred reproduction, included in a register recognised by the Member State (e.g. herd book or flock book).

In 1993 Simon and Buchenauer proposed criteria based on the minimum effective population size needed to limit the accumulated inbreeding in 50 years (Simon and Buchenauer, 1993). The breeds were categorised in 5 classes 'not-endangered', 'potentially endangered', 'minimally endangered', 'endangered', 'critically endangered'. Species specific thresholds were used, based on the different generation interval in the various species. Additional factors like change in number of breeding animals, absence of herdbook, number of herds, percentage of incrossing and percentage of purebreeding were also used to downgrade the breed into a worse class. Later, the thresholds for these criteria have been increased (Simon, 1999), e.g. a breed was considered not at risk if the expected cumulated inbreeding in the next 50 years will not be greater than 10% (compared to 5% limit in the previous system). The acceptable value of incrossing was also doubled - from 10 to 20%. This resulted in a higher percentage of breeds in the 'not-endangered' class and fewer downgradings by additional factors.

The Rare Breeds Survival Trust uses criteria with three groups of requirements: genetic basis, numerical basis and current trend for identifying the rare farm animal breeds within UK. Only closed breeds or 'original populations', which should have existed for forty years and six generations. Herdbook, studbook or flockbook should be present. As a numerical basis the RBST system uses the number of registered adult females alive. Five categories are defined - 'critical', 'endangered', 'vulnerable', 'at risk' and 'traditional' and the thresholds used are species specific. The maximum number of registered females in existence, which allows listing are: cattle - 1500,
introduction

Another system, used internationally, is the one proposed by Rare Breeds International (Alderson, 2003). This system is based on the number of annual registration of female young stock (on one-year basis, or three-year rolling average), and defines four categories for horses, cattle, sheep and pigs. Additional modified factors - local adaptation, distinctiveness, global population and level of inbreeding are used to increase or decrease the conservation priority.

Recently, Reist-Marti et al. proposed criteria for assessing the extinction probabilities of a breed based on the population statistics, the environment and the present breed 'value' (Reist-Marti et al., 2003). The population factors include the total population size and its change over time, distribution of the breed and risk of indiscriminate crossing. Environment factors are represented with organisation among farmers, establishment of conservation scheme and the political situation in the country. The 'value' of the breed is estimated on the basis of its special traits and cultural importance. Additionally, the factor “reliability of data” is used to correct the results depending on data quality. The results are also rescaled to a value between 0.1 and 0.9 to rule out the possibility that the breed is considered completely safe or entirely doomed to extinction (Reist-Marti et al., 2003).

One of the major proposals for breed classification according to their DE is the one used by FAO (Loftus and Scherf, 1993). This system uses as a base factor the breeding population size. Three basic classes are defined: 'critical', 'endangered' and 'not at risk', based on thresholds for number of breeding males and females. Two additional classes: 'endangered-maintained' and 'critical-maintained' are defined for breeds from the respective basic classes, for which a conservation programme is in place, or if the population is maintained by commercial companies or research institution (Scherf, 2000). Additional factors like actual population trend and the percentage of females being bred to males of the same breed are used to upgrade or downgrade the breed. The thresholds are identical for all species, e.g. thresholds for
'critical' are 100 breeding females or 5 breeding males, or the overall population size is not greater than 120 and decreasing and the percentage of purebred females is under 80% (Scherf, 2000).

An attempt was made by Gandini et al. (2004) towards uniform European criteria based on the data collected in the EAAP-AGDB. This proposition is based on two factors: the time to reach a critical population size and the rate of inbreeding. The time to reach critical population size is estimated on the base of the current and the projected population size and the proportional rate of population growth. Various thresholds for the population size and techniques for the estimation of the growth rate are discussed, with the main intention to introduce cohesion among the current systems for estimating DE.

**Biodiversity information systems**

One of the main preconditions for good estimation of the DE of a breed is the quality of the data used. Moreover, setting up a conservation scheme requires also knowledge of the regional situation and even the situation with the similar breeds in possibly distant regions (Bisby, 2000). These issues are addressed also in the CBD’s Clearing-House Mechanism (CHM). One of its purposes is in providing a stable data, searching and retrieval mechanism to support the obligation taken by the State Parties (Silva, 2004). The advances in information technology in the last years of the previous century changed revolutionally the way of maintaining and distribution of biodiversity information (Bisby, 2000; Soberon and Peterson, 2004). Many biodiversity information systems were developed on regional and global level for specific taxa or generally for all kind of specimens (Canhos et al., 2004; Silva, 2004), e.g. GBIF (Edwards et al., 2000; Global Biodiversity Information Facility), Species2000 (Species2000), Tree of life (Maddison and Schulz, 1996-2006), Animal Diversity Web (Parr et al., 2005), FishBase (Froese and Pauly, 2006), etc, but there is still a need for interoperable farm animal information systems on a national and supranational level.
Development of farm animal biodiversity information systems in Europe

There are several supranational biodiversity information systems specialised for domestic animals, which collects data from Europe: the Animal Genetic Data Bank of EAAP, the Nordic Gene Bank for Farm Animals (NGH) and the Domestic Animal Diversity Information System of FAO.

Following the recommendations of the FAO/UNEP Technical Consultation on Animal Genetic Resources Conservation and Management, the EAAP Animal Genetic Data Bank (EAAP-AGDB) was set-up in 1987 at the institute of Animal Breeding and Genetics in Hanover (Simon, 1990). The initial data for the EAAP-AGDB was obtained from three Europe-wide surveys issued by the EAAP-WGAGR. The first survey of AnGR was issued in 1983 and it was supposed to describe the situation in Europe in 1982, covering 5 species from 22 countries. Data for 8 more countries were added based on a literature sources. Thus, the survey contained data for 737 breeds: 181 - cattle, 77 - goat, 149 - horse, 66 pig and 264 sheep breeds (Maijala et al., 1984). The second survey took place in 1985 with response from 17 countries and covered 513 breeds: 148 - cattle, 45 - goat, 73 - horse, 64 - pig and 183 sheep breeds (Simon and Buchenauer, 1993). After this survey it was decided that the database should include all breeds or 'country populations' and not only the endangered ones. In 1988, a third survey was distributed to the countries. This survey requested more data and was developed on the basis of the previous two in cooperation with the Nordic Working Party on Animal Genetic Resources and with the Animal Production and Health Division of FAO. The replies from 12 countries covered 225 country populations, from which 68 - cattle, 9 - goat, 30 - horse, 35 - pig and 83 sheep populations (Simon and Buchenauer, 1993).

In 1984 the Nordic Gene Bank Farm Animals (Nordic Council of Ministers) was established as a body under the Nordic Council of Ministers, to coordinate the national measures in the management of AnGR in the Nordic countries.

In 1987 EAAP and FAO had agreed on a uniform questionnaire to be used in Europe and all over the world. The data collected from regions outside Europe were stored also in the EAAP-AGDB database, which by that time served as a Global Animal Genetic databank. A common working group of FAO and EAAP was created.
and on its meeting in Berlin 1991 it was decided to split the work between both parties. The EAAP-AGDB was supposed to continue collecting data from Europe, whereas the FAO will take care of collecting data from other regions (Simon and Buchenauer, 1993).

In 1991 FAO initiated a breed survey in all non-European countries. This survey covered only the livestock species - ass, buffalo, cattle, goat, horse, pig and sheep. In 1993 global surveys were initiated for domestic avian species and Camelidae (Scherf, 2000). Till 1994 all data from EAAP-AGDB was transferred into the Global DataBank. In 1996 the Global Databank and EAAP-AGDB were made accessible via Internet for browsing.

In September 1998, the second stage of DAD-IS (DAD-IS 2.0), allowing the registered users to enter and modify data via the Web-page was released (Scherf, 2000).

As a result of the separate development of FAO and EAAP, similar, but not uniform data were collected by DAD-IS and EAAP-AGDB. This situation was adverse for the informants of the both systems, which had to prepare the same data in two different formats and, more importantly, had to enter the data twice in two systems. As a natural result, there were gaps of data in either databases.

**European Farm Animal Biodiversity Information System**

With the EU funded project "European Farm Animal Biodiversity Information System" (EFABIS) a new network got developed which was to replace the old EAAP-AGDB as well as FAO’s DAD-IS 2.0 by a uniform software system (Rosati et al., 2006). The developed network is an extensive multilingual source of information about the characterisation, conservation and utilisation of AnGR in Europe.

The databases in the network collect data for breeds of domesticated animals in more than 35 mammalian and avian species. The recording is done on a breed level per country, including autochthonous breeds and country populations of international breeds. Each database is organised in five sections inherited from DAD-IS 2.0 and EAAP-AGDB: breed data, library of publications, references and Web-links, image gallery and contacts data.
The breed data section has two parts - characterisation part and population statistics part. The characterisation part contains information about the identification of the breed, origin and development, morphology and performance traits, special qualities, utilisation and conservation. The data are grouped in species specific blocks, e.g. milk performance data, horns data, eggs data, etc.

The population part contains statistics of the population size and structure for a given year. Data collected includes number of breeding males and females, number of registered and purebred females, number of herds, average herd size, trends in population size, breeding females and herd size, use of AI and natural service, presence of conservation programme and actual number of stored semen straws, embryos, oocytes and somatic cells. As the both parts in the breed data are changing with a different speed, separate update frequencies are required for each part. The population statistics part should be updated at least once during one breed generation interval.

The virtual library in each database contains around 1300 research articles, governmental and non-governmental key documents as well as links to external databases and web-sites and references to software tools covering various topics of management of genetic resources. In addition, a gallery with more than 4000 images of the various breeds, represented in the database are available.

The network operates on three levels - National, Regional and Global as shown in Figure 1.1. Within the network countries can establish their own Web-driven National farm animal biodiversity databases in one or more local languages, free of licensing costs as this was done for Poland (National Polish Biodiversity Database, 2006). The subset of data required on the upper levels in one of the official FAO languages will be automatically transferred from the National databases to the Regional one (EFABIS, 2006) via the synchronization protocol developed in EFABIS (Duchev and Groeneveld, 2006).

Several countries in Europe already have farm National biodiversity databases, e.g. Central Documentation for Animal Biological Diversity in Germany (TGRDEU, 2006), but the major part of the European countries will benefit from EFABIS for establishment of their National databases.
The software for operating the network was developed in such a general manner, that it can be used for the establishment of biodiversity networks also in other regions of the world.

Figure 1.1: *The topology and data-flow of the FABISnet*
Scope and outline of the thesis

As described in the introduction, the detailed characterisation of the AnGR, their prioritisation for conservation and their utilisation are some of the most important steps in preserving the biological diversity in domesticated animals. Furthermore, these data have to be available to a wider group of potential users - scientists, students, policy makers, Government and Non-Government organisations involved. The current work focuses on two main problems - improving the quality of data within the EFABIS network and assessing the degree of endangerment for the loss of breeds diversity. The problem of data quality and consistency is addressed within a single database [Chapter 4] and between the databases [Chapter 2]. The early-warning system for assessing DE on a national and supranational level and identifying the populations at risk is presented in Chapter 3.

In Chapter 2 the synchronization protocol for automated data exchange between the databases in EFABIS is described. This protocol solves the problem of having to enter the same data into several database, which was the situation with the previous DAD-IS and EAAP-AGDB databases and which has led to data inconsistency. The automatic data synchronization allows the National Coordinators for management of genetic resources to maintain their country data only in their National databases, ensuring that the data will be present at all other levels and kept up-to-date.

In Chapter 3 a proposition for a uniform criterion for Europe for estimating the DE of a breed is developed. The criterion is based on the trend in number of breeding females and the expected cumulated inbreeding. It takes also into account the status of global population (factor missing in the criteria applied currently in Europe) and has a novel approach to the problem of transboundary breeds and the grouping of genetically similar breeds.

Chapter 4 contains an analysis of the information requirements for data management in EFABIS and the description of the management support system developed to meet these requirements. Incompleteness of the data is still one of the biggest problems on the regional and global level. In this chapter the methodology and the tools for filling gaps in data and improving the quality of data needed for
research are described. The aim of this work is to ensure also enough data for successful application of the early-warning system described in Chapter 3.
References


Chapter 2

Synchronization of APIIS based farm animal biodiversity systems

Zhivko Duchev and Eildert Groeneveld

The extent of Zhivko Duchev’s contribution to the article is evaluated according to the following scale:

A. has contributed to collaboration (0 - 33%)
B. has contributed significantly (34 - 66%)
C. has essentially performed this study independently (67 - 100%)

1. Design of the project B
2. Analysis of the data C
3. Implementation into software module C
4. Presentation and discussion of the study in article form C
Abstract

One of the major problems in the management of farm animal and biodiversity information is the exchange of data and keeping it up-to-date, an issue that is very common with distributed information systems consisting of number of databases. This article describes the synchronization protocol developed in APIIS (adaptable platform independent information system) framework and reviews the basic considerations required when building distributed information system that has to exchange information in a network of APIIS based systems. The protocol is designed to synchronize a common part of different database structures. It is developed without any intended use of proprietary database engine and can work with a variety of RDBMS (relational database management system). The main targets of the protocol are animal biodiversity information systems without permanently connected nodes. The EFABIS (European farm animal biodiversity information system) is reviewed as an example of the implementation.

Availability: The synchronization protocol is integrated as a part of the APIIS framework, which is freely available from the authors.

Keywords: database management; data synchronization; farm animal information systems
Background:

Development of open source databases in the APIIS framework (Groeneveld, 2004) is common and the installation of identical systems becomes a financially affordable option. As outlined in elsewhere (Groeneveld, 2002), data collection in animal agriculture relies on the distributed collection of farm animal data: these may originate on many farms, on test stations and in laboratories. These different sources will have to be integrated into one central database for across herd evaluations as is typically done in Best Linear Unbiased Prediction (BLUP) genetic evaluation (Henderson, 1975) in selection programs. In principle, peripheral databases can be viewed as subsets of the central system both in terms of the database structure and also in terms of business rules. Clearly, business rules should be enforced at the initial data entry where - in case of errors - the original information is close at hand for correction. These local herd systems may be copies of the central system expanded in scope for on herd management as is done in commercial herd management packages which are available for nearly all species in animal agriculture. With this topology the task of transferring data from the periphery (e.g. farms) to the center is the last step in building a comprehensive central database.

Because all business rules have already been enforced at the periphery using a set identical to the rules at the central database, a generic data transfer can be employed which amounts to the synchronization among the central and the peripheral databases, without a need to consider business rules at this stage. The EFABIS network has a similar topology: there is a world wide central node of the biodiversity database at the United Nations FAO (Food and Agricultural Organization) in Rome (DAD-IS, 2005). On the regional level the EAAP (European Association of Animal productions) runs a database (EAAP-AGDB, 2004) with expanded information both in terms of content and structure, while countries like Poland have their own national database comprising all information from the levels above plus additional national data not to be shared with the other levels. Again: data is collected at the national and perhaps for some countries at the regional and worldwide level and will have to be propagated to all other levels. In line with the
example from the animal agriculture, we have an identical core structure of the
databases at all levels and are enforcing the same set of business rules everywhere.
Thus, after initial data entry - at whatever level - information newly added to the
database must be transferred, i.e. can be synchronized with the other databases in
the network. Development, implementation and performance of such a
synchronization procedure are described here.

**Glossary of Terms**

Data element (DE) = smallest amount of data treated as one block in the
synchronization process. DE is the list of columns from a defined subset of records in
a table;
- Node = each independent part (database) of the global information network;
- Source = any node that distributes data elements to other nodes;
- Target = set of nodes to which one source distributes a data element;
- Network manager = the management authority that will route the traffic of
  information, preventing conflicts or inconsistencies.

**Methodology:**

**Synchronization Requirements**

The synchronization requirements were derived according to the requirements
of EFABIS network.

**Requirement #1**

Each DE has a primary copy and there is only one database in the network
where it can be edited. In animal breeding information systems, data is usually
collected on different places like artificial insemination stations, farms and research
institutions. All these sources of information keep copies of data, there is someone
(human or organization) who is officially responsible for the quality of data and all
users of these data rely on its representative value. As an example the veterinary
examination of the animal can be taken and there is a paper document containing the
animal identification data, the veterinarian data and the results. Therefore, a natural
requirement is that each data element should have a primary copy at one node where this element can be changed. This is the node where the person collecting the DE always enters the data, and in all other nodes this data will be read-only. This requirement ensures also a clear responsibility for the accuracy and up-to-date status of each DE. For example, each country in EFABIS that presents its data to the European (EAAP) and global (FAO) level is responsible for the data quality and consistency.

**Requirement #2**

For each DE the "distribution target" (nodes that want to obtain this element) is defined. In general terms, the data collection process does not end in itself. Usually the collected data is intended to be used by someone and in most cases the data users are clearly defined. For example, data collected on testing stations may be sent to a research institute for calculating the breeding values and the results are returned to the farmers. Very often there is a strictly defined hierarchy in the system with one central database collecting all data as a data-warehouse. This is the situation in the EFABIS network, where each European country sends data to EAAP and EAAP distributes part of the data to FAO. Therefore, for each data element there must be a well defined target group of nodes which needs this DE. This set of nodes is actually the “distribution target”. It could be also empty if this element is only for local use and will not be propagated.

**Requirement #3**

Each DE to be included in the synchronization process has to be defined by both source and target nodes. The DE to be transferred has to be negotiated and approved by the both sides. When a reconciliation session is started it automatically synchronizes all approved DE, thus not allowing the user to refuse the changes. This principle looks very restricting, but follows from the requirement of primary copy. The idea behind this requirement is that a user who needs a certain DE is accepting by default all changes, relying on the fact that they are representative. For example, if the primary copy of DE is deleted, then this element should be deleted everywhere.
In contrast, the act of removing DE from the synchronization list has to be confirmed by both sides. Distribution sources and targets may be changed as long as this does not produce inconsistencies. This principle ensures that each node can choose the source and target nodes for a DE, unless this will disturb normal flow of data in the network. This implies that all changes in the DE path have to be coordinated by all nodes that exchange this DE.

**Requirement #4**

Each node can distribute all public data elements loaded in its database. If the primary copy node is the only source for a DE then this will produce a bottleneck in the data-flow. Therefore, each node that has received a public DE as a result of a synchronization process should be also allowed to propagate it further. This is not the case with non-public data elements and such elements can be distributed only to a subset of authorized nodes.

**Requirement #5**

Synchronization should not require human intervention. The protocol should be completely automated and be able to run on a regular basis as a scheduled task. It should not produce any inconsistencies in the target node, because such discrepancies usually require human intervention from scientific and technical persons - the former to solve the conflict, the latter to introduce the changes to the database. The process of solving conflicts is time consuming, and it requires the original data copy.

**Requirement #6**

A network regulating mechanism for the data-flow should exist. As the nodes are equal in rights and part of the requirements rely on the negotiation between two nodes, an unregulated data-flow can produce locks in the system. Therefore, if the network has no inbuilt "by design" clear data-flow, it has to be regulated by set of rules. They will prevent actions that are against the system consistency or resolve data-exchange conflicts between the nodes. The need for such rules can be seen
from the following example: Let the node A target one of its data elements to node B and node B target this DE to node C. Let also presume that by system design node B has to have always this DE. In this situation if node A wants to change the target of this DE to node C then node B will lose its source. There are two possible solutions of this conflict: Node A is not allowed to change the target, because it will produce inconsistency. It is allowed to change the target, but has to do this in cooperation with node C, which will target it to node B.

**Requirement #7**

The system is loosely coupled and not all nodes are connected all the time. Although the access to the INTERNET is getting cheaper, there are a lot of farms, even in European countries, where the only option for connection is via phone line or satellite. An example is PISSA (Pig Information System South Africa) where data are collected in the farms off-line and then sent to the center once per week via e-mail.

**Requirement #8**

The protocol should be able to synchronize data over LAN (local area network) and WAN (wide area networks) such as INTERNET. It has to ensure secure transfer of the data over the public parts of the network. The nodes of the animal information system which uses this protocol can be part of the internal network of one organization or can be connected via INTERNET. Therefore, the synchronization protocol should use network transport protocols which are applicable everywhere. And as the data exchanged can be private, the protocol has to encrypt it when transferring over a public network.

**Requirement #9**

The protocol has to be able to exchange text and binary data. The last but not least requirement is related to the type of information exchanged. We will not only synchronize data fields in the database containing quantitative values like size, milk, wool length, but also documents and multimedia data. This may look obvious, but it is important for the type and quantity of the data that will be transmitted.
Analysis of the requirements and description of the developed synchronization protocol

Analysis of the requirements

The requirement for one node where a user can change a DE puts us in situation similar to the Lazy Master Replication model from distributed systems (Gray et al., 1996). According to this model, when the user updates a DE, only the primary copy of this DE in its master node is updated. Then, in separate transactions the master node updates each replica. There is a certain time of inconsistency between the master node and the replicas. Therefore this model is called lazy or asynchronous. In our system we also use the primary copy approach, but we presume that not all nodes can connect to the master node. Therefore, the propagation of changes to the other nodes is done in a cascading manner and the nodes using the master node as a source are updated first, then their target nodes are updated and so on until all nodes are updated. Each node except the master one will be in an asynchronous state until one of its sources is updated and synchronization with that source take place. To assess the time in asynchronous state and its impact on the system functionality we have to look at the specificities of the information systems we are dealing with. In the national and supranational biodiversity systems like EFABIS a detailed breed description, morphology, performance and demographic data are collected. New data in such systems are loaded on a monthly or even on yearly basis. On the other hand, in herd management information systems, the central database receives data weekly or daily. Hence, synchronization per day [or] per week is sufficient.
The other issue to be considered here is the number of sources a node can use for a DE. Having several sources for one DE can lead to values collisions - if node C gets two different DE versions from node A and node B. Such problems are discussed elsewhere (Terry et al., 1995; Ekenstam et al., 2001). A simple restriction that solves such conflicts is the limitation of one source for each data element. Two different databases can have two different sources for the same DE, but a single database can have only one source for its DE. If the user node can establish a connection to more than one node keeping a DE then the user can choose, in accordance with the network rules, which one will be used as a source and also move from one source to another, but cannot use two sources simultaneously. The one-source requirement produces a bottleneck in the data-flow, because the possible interval to update DE is the intersection of the online time of the source and the target node. This restriction does not have big impact on the animal biodiversity databases, as already stated, the time interval between the updates is relatively long. In herd management animal information systems each DE comes to the central database usually from one source, e.g. field test data for a certain animal comes from one farm. Moreover with proper management of the network, there can be a scheduled interval of time when both nodes are online for synchronization.

Figure 2.1: Excerpt of EFABIS topology, containing National Polish database, European regional one and the world database of FAO. The dashed arrow represents the rule for one source per data element.
In each node we have information about each data element's route. This can be fully described by the expression:

\[
\text{DE} \left[ \text{Primary Copy}, \quad \text{Source}, \quad \text{Target} \right]
\]

\begin{align*}
\text{Primary Copy} &::= \text{Node Name} \\
\text{Source} &::= \text{Node Name} \mid \text{Empty} \\
\text{Target} &::= \text{List of Node Names} \mid \text{Empty}
\end{align*}

'Primary Copy' is the name of the node where this DE was initially entered. The 'Source' is the node that has supplied this element and 'Target' is the list of nodes, this element will be delivered to. These expressions are illustrated in an example from EFABIS, shown in Figure 2.1. In this example we use the DE 'breed description', which includes the general description of all Polish breeds. This DE is initially entered in the Polish database. Poland distributes it to the EAAP database and from there it will be propagated to the world database of FAO. As a result, we have the following descriptions of the DE's route in the various databases. In the Polish node (named PL), it is described as DE[PL, EAAP]. Here the 'Primary Copy' is PL because it is the first node where this DE is entered in the system. The 'Source' field is empty, since we have not received this DE as part of synchronization process. The 'Target' list consists of only one element – 'EAAP', because only the EAAP node will receive it directly from Poland. Following the same logic the description in the EAAP node is DE[PL, PL, FAO] and in the FAO's database DE[PL, EAAP, ]. The empty Target field in the last description means that FAO will not distribute this DE to other nodes.

The source and target fields of a DE’s route description in the nodes as defined above are sufficient to determine the route of DE within the whole network, resulting in a tree structure. The 'Primary Copy' node is the root while the sources and targets describe the ribs. Two checks have to be done in this structure: (1) the ribs definition is consistent, and (2) there are no cycles in the tree. The former check can be done
for each of the two nodes such that Node1 (source) and Node2 (target) exchange the element DE1:

\[(\text{Node2.DE1}[\text{Source}]=\text{Node1}) \land (\text{Node2 in Node1.DE1}[\text{Target}])\]

The main problem with both checks is that information is spread around the nodes and has to be collected in one place by the manager of the network.

There are two possible types of synchronization: (1) based on a log (journal) and (2) based on a state. The journal synchronization is based on the logging of all data modification statements that are executed on the source database and sending them to target node. It is suitable when there is a lot of data with relatively small number of changes. The log-based reconciliation is used in systems like replicated dictionary (Wuu and Bernstein, 1986), Bayou (Petersen et al., 1997), Vagabond (Nørvåg and Bratbergsengen, 1997) and StorageBox (Hupfeld, 2004). The general problem with this approach is that the node has to ensure that all targets have updated their state before removing the statement from the log file, which can lead to accumulating large amounts of unused data. This can happen for example, if a node which is in the target list does never connect to the source node. The approaches to this issue vary from discarding writes from logs in (Petersen et al., 1997) to removing long latent target nodes from the replication set (Ladin et al., 1992), but they are not suitable in our setup.

Therefore, the alternative is synchronization 'by state' - based on version vectors or time-stamping of the data elements (Parker Jr. et al., 1983; Almeida et al., 2002). The time-stamp approach requires clock synchronization as shown in (Lamport, 1978), which is practically impossible in the network of independent databases like EFABIS. Therefore, the versioning approach was chosen, where each record has an integer version attached, which is incremented on update. In the synchronization session, the source node plays the role of the server and the target node that of a client. The client sends the current version number of the DE to the
server where it is compared with the server's own version. If the server's version is new the updated DE is propagated to the client. The shortcoming of this method is that each time the versions of all DEs to be synchronized are compared, thus making the overhead proportional to the number of records. This approach is suitable for databases with relatively small number of records and this is the case with farm animal biodiversity databases collecting cumulated data on breed level. For example, the European regional database and Polish National database in EFABIS contain 21,4426 and 7,290 records respectively.

Database structure

To execute and manage synchronization between databases the following additions are made to the database structure:

Additional columns:

It is well known from the replicated databases that each record has an identifier that is unique within the information network. To ensure this independently from the RDBMS engine, which is one requirement of the APIIS design (Groeneveld, 2004), a new 'system' column for the Global Unique Identifier 'guid' has to be added to each table. Upon insertion in the primary copy node the 'guid' is automatically set from a sequence. On the other hand the synchronization process has to preserve the 'guid' in the target database.

The synchronization 'by state' requires to keep track of the changes made to records. Therefore, in each table an additional field for the record version has to be added. When initially inserted in the primary copy node all records from one DE have a version set to one and each update increments the version by one. This field should be also included in the synchronization of a DE.

By definition, each DE consists of "defined subset of records". Such classification in APIIS is done on the basis of the additional 'class' column which has been added to each table. It is up to the designer to define the classes when
designing the system. The classification of the records can be done on the basis of the location where the original data were collected. In IS collecting individual animal data, these places can be farms, breeding societies, test stations. As an example let we have a system collecting herdbook data for animals from three farms. Then each record can be classified as ‘Farm1’, ‘Farm2’ or ‘Farm3’, depending on the farm of the animal.

To have option for private data in the database, each record has a Boolean ‘synch’ field. The flag stored in this field is used to indicate if the record is targeted for synchronization. The user when entering data explicitly sets this flag. Examples of the meta-fields are shown in Figure 2.2.

<table>
<thead>
<tr>
<th>SOURCES</th>
<th>TARGETS</th>
<th>NODES</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>target</td>
<td>Roomname</td>
</tr>
<tr>
<td>idname</td>
<td>idname</td>
<td>address</td>
</tr>
<tr>
<td>class</td>
<td>class</td>
<td>last_change_dt</td>
</tr>
<tr>
<td>columnnames</td>
<td>columnnames</td>
<td>last_change_user</td>
</tr>
<tr>
<td>last_change_user</td>
<td>last_change_user</td>
<td>dirty</td>
</tr>
<tr>
<td>dirty</td>
<td>dirty</td>
<td>chk_lti</td>
</tr>
<tr>
<td>guid</td>
<td>guid</td>
<td>guid</td>
</tr>
<tr>
<td>owner</td>
<td>owner</td>
<td>owner</td>
</tr>
<tr>
<td>version</td>
<td>version</td>
<td>version</td>
</tr>
<tr>
<td>synch</td>
<td>synch</td>
<td>synch</td>
</tr>
</tbody>
</table>

**Figure 2.2:** Additional tables and meta-fields (shaded) needed for the synchronization protocol in the APIIS structure

**Additional tables:**

The management of routes requires three 'system' tables in each database. The table 'Nodes' contains the names of the nodes and their physical IP-addresses. Each node which is source or target of the current node must be registered in this table. All names must be unique within the network with each IP address linked only to one node. The other two tables – 'Sources' and 'Targets' (Figure 2.2) are used for specifying the incoming and outgoing data elements and their nodes through the ('columnnames', 'class', 'tablename', 'source'|'target') columns.
Synchronization protocol

The synchronization protocol is of client-server type. Each node, which distributes data elements, has a server daemon listening for incoming connections. Such a node will be referred in the following as ‘server’. On the other side, the node which wants to update its data from the server is the “client” and has to run the client part of the software. When a connection is initiated by the client, it starts with a handshaking to verify if the server is free for synchronization. In this case, the client reads the description of the first DE, from the server and sends this description called DED (Data Element Description) for confirmation. After successful confirmation from the server, the client reads the state of this DE (the guid and version of all records described by the DE) and sends this information to the other side. Then the server compares this information with its own state and chooses appropriate action for updating the client:

```plaintext
foreach client.record in DE1 {
    if (not exists server.record) then
        client.record.action='Delete';
}
foreach server.record in DE1 {
    if (not exists client.record) then
        client.record.action='Insert';
    else if(server.record.version>client.record.version) then
        client.record.action='Update';
}
```

The action and the data retrieved from the server (in case of insert or update) are encapsulated into a merge structure and send back to the client. There the merge structure is transformed in SQL statements in the client’s native SQL dialect and the database is updated. The functional model of the synchronization process for one DE is shown in Figure 2.3. The same steps are repeated for all other DE
expected from the server. The entire operation is treated as one transaction and changes are committed only if all DE are successfully updated. This is a weakness of the protocol, because in case of error, the synchronization has to be started from the beginning. The block diagram of the used algorithm is shown in Figure 2.4.

**Figure 2.3:** Functional model of the data synchronization process for a single data element.

**Figure 2.4:** Block diagram of the synchronization algorithm
Implementation in EFABIS

The automated synchronization protocol was implemented in EFABIS (European Farm Animal Biodiversity Information System). This system is a network of databases collecting biodiversity data from European countries and transfers it to the central European database. The European database will be used as a data source for the world biodiversity database of FAO, especially for the data that is required to build the World Watch List for endangered domestic breeds (Scherf, 2000). The data collected in EFABIS describes the farm animal breeds in terms of naming, origin and development, morphology, special qualities, performance, demographic trends and conservation programs.

In EFABIS we have clear hierarchy on three levels - National, Regional and World level. Each lower level is an expansion of the previous one in terms of content and structure. For example the National Polish Biodiversity database (NRIAP, 2006) stores data required by EAAP and additional data for species like fish and small fury animals, which are not represented at the European level. The Polish database has the structure of the European one plus extensions in terms of additional tables and fields, to handle the country specific data. The flow of data (as shown in Figure 2.5) is bi-directional. National databases have to send the data required by the European database and from there a subset will be sent to the World database. The world database will also propagate some data to the national databases, e.g. documents, images and common codes that have to be uniform within the whole network like codes for the species and sex. Such codes must be introduced only in one place, i.e. - the FAO database and accepted from the other nodes. Each node has also the option of having private codes, but these have to be used only in private data which will be not synchronized.

Each national node is named by the ISO-3166-1-alpha-2 code elements of the International Organization for Standardization (ISO, 2005), i.e. BG for Bulgaria, PL for Poland. This ensures unique name even within the whole world. Two exceptions are the names of the European database and the global database - for the former ‘EAAP’ is used, and for the latter - ‘FAO’. To ensure uniqueness of the ‘guids’ and all
other internal identifiers each database has to have a separate range for the sequence generators. Operationally, this was done by attaching to the official list of countries a predefined range for each country. This range should be set in the configuration files needed for creating a new node. All other actions required are completely automated by the software.

Figure 2.5: Topology of the EFABIS system. On the lower row the national databases are represented - BG for Bulgaria, DE - Germany, PL - Poland, VN - Viet Nam. In the middle row are the regional databases, e.g. EAAP for the Animal Genetic Data Bank of the European Association for Animal Production. On the top is the biodiversity database of FAO

Each record has to be identified as owned by a country or by one of the supranational databases. This give us the unique classification of the records and therefore the class column in EFABIS was named 'owner'. As a value the two-letter codes of the countries, along with 'EAAP' and 'FAO' are used. This marker is set automatically by the system through the country that a user is attached to. This setup allows countries that do not have their own database to load data directly in the European database. The EAAP node will be primary copy for these records, i.e. these data can be edited only there. The class column 'owner' was also used in the Access Control System of EFABIS, not only to mark the backups of DE as read-only, but also to define complete set of access rights for all users in the network.
For the transport part of the synchronization the TCP/IP protocol was used. The server daemon was set to listen on port 5433. If the node is behind firewall, it has to be set to allow incoming connections on this port. For compression and encryption of the transfer data, Mina Naguib’s Net::EasyTCP module from CPAN (2005) is used. The encryption of the data is done using the Blowfish algorithm (Schneier, 1994).

As an interesting byproduct the synchronization procedure has been used to initialize newly created national databases. By simply defining appropriate sources and targets, the national data was downloaded in one-time transfer from the global database.

**Performance**

Two nodes were established on two different machines for testing the synchronization protocol. One of the machines was with Mobile Intel(R) Pentium(R) 4, CPU 3.06GHz, 512MB RAM and the other with 4 64 bit AMD Opteron™ 850, CPU 2.4GHz, 8GB RAM. The tests were done using three types of connections: 100Mb/sec, 10Mb/sec and 128Kb/sec. In each test the time to receive merge data from the server, the time to update the database and the total synchronization time was measured. The results, rounded to the next integer are shown in Table 2.1. The value of 100,000 merge records was chosen to test the situation of initial loading of the database, while the other runs represent the amount of the updates expected in small and middle databases. The amount of memory used for the state information for 100,000 records was 5.54 MB, or 59 bytes per record. This information was read from the database in 2.35 sec and transmitted to the client in 1.128 sec using a 10 Mb/sec connection. As can be seen from the results with fast connections, the time needed to transfer data over the network is approximately the same or even less than the time needed to merge changes in the client database, i.e. the bottleneck is the communication with the database backend. Another interesting observation is the time required to receive data on 100Mb/sec and 10Mb/sec were practically the same. This means that the protocol does not use the whole bandwidth which is also confirmed by the results on 128 Kb/sec. When reducing the bandwidth 80 times (from
10 Mb/sec to 128 Kb/sec) we have only about fifteen fold increase of the receive time. A similar tendency can be observed in the total synchronization time - it increases only six time required from 100Mb/sec and 10 Mb/sec to 128 Kb/sec. This speed of the protocol is mainly due to the fact that it internally encrypts and compresses data for secure transfer over public network and usually the average record size in animal databases is relatively small - about 500 B. Therefore, the results on a 128Kb/sec connection can be considered sufficient for production systems with slow internet connections.

**Table 2.1:** Time to receive merge data from the server, update the database and total synchronization time

<table>
<thead>
<tr>
<th>Time to receive merge data (seconds)</th>
<th>Bandwidth / Merge records count</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Mb/sec</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>10 Mb/sec</td>
<td>1</td>
<td>2</td>
<td>13</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>128 Kb/sec</td>
<td>2</td>
<td>13</td>
<td>129</td>
<td>1279</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time to update the database (seconds)</th>
<th>Bandwidth / Merge records count</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Mb/sec</td>
<td>1</td>
<td>2</td>
<td>17</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>10 Mb/sec</td>
<td>1</td>
<td>2</td>
<td>17</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>128 Kb/sec</td>
<td>1</td>
<td>2</td>
<td>16</td>
<td>163</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total synchronization time (seconds)</th>
<th>Bandwidth / Merge records count</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
<th>100000</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Mb/sec</td>
<td>7</td>
<td>9</td>
<td>33</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>10 Mb/sec</td>
<td>8</td>
<td>10</td>
<td>36</td>
<td>270</td>
<td></td>
</tr>
<tr>
<td>128 Kb/sec</td>
<td>34</td>
<td>46</td>
<td>175</td>
<td>1451</td>
<td></td>
</tr>
</tbody>
</table>

**Conclusions**

The synchronization protocol designed for data exchange between loosely coupled nodes in farm animal information systems relies on the strict primary copy approach for each data element, thus avoiding update conflicts and the need for human intervention. The drawback of the protocol is the transfer and comparison of the DE state information on each run. Nevertheless, the protocol shows good results on medium and small databases similar to the ones used in biodiversity, national gene banks and small population management information systems. The total
synchronization time scales using a 128 Kb/sec connection, allows running the protocol as scheduled on a daily basis.
Chapter 2  Synchronization of APIIS based biodiversity systems

References


Chapter 3

Early warning system for loss of diversity in European livestock breeds

Zhivko Ducheve and Ottmar Distl and Eildert Groeneveld

Submitted for publication
The extent of Zhivko Duchev’s contribution to the article is evaluated according to the following scale:

A. has contributed to collaboration (0 - 33%)
B. has contributed significantly (34 - 66%)
C. has essentially performed this study independently (67 - 100%)

1. Design of the project B
2. Analysis of the data C
3. Implementation into software module C
4. Presentation and discussion of the study in article form C
Abstract

Towards the end of the previous century the increasing erosion of biological diversity in livestock was recognized by the animal breeders as an area which needs monitoring and conservation measures. For that purpose various national, regional and global infrastructures was put in place by governmental and non-governmental organizations like European Association for Animal Production (EAAP), Food and Agriculture Organization (FAO) of the United Nations and Rare Breeds International, etc. These organizations have developed a number of methods for estimation the risk status of the breed. However, the different systems use different parameters and thresholds and are hard to compare. Several attempts for uniform European criterion have been made, but still no consensus has been reached. In this study we continue the work of the Animal Genetic Resources group of EAAP towards the uniform criterion based on the effective population size, the expected number of breeding females and the global population. Our main contribution is the novel approach in handling the parameter global population, allowing its external parameterization. The criterion was applied to 21 pig breeds from 8 European countries and the results show clear not at risk status of the big international breeds.

Key Words: degree of endangerment, farm animals, biodiversity
Zusammenfassung

Titel der Arbeit: **Frühwarnsystem für den Verlust an genetischer Diversität bei den europäischen Haustierarten**


Schlüsselwörter: Gefährdungsstatus, Nutztiere, Biodiversität
Introduction

In the last 12000 years around 40 livestock species were domesticated resulting in a 6000 to 7000 genetically distinct breeds (Scherf, 2000). These breeds have been developed and adopted to various production environments and have their importance as a sources of food, element of the rural culture (Gandini and Villa, 2003) or a tool for ecosystems management (Gordon et al., 1990). The preservation of the farm-animal diversity is an insurance against changes in the market demands or production environment circumstances (Oldeenbroek, 1999). With the introduction of new technologies and the increasing demand for high production in the last century, a number of well-marketed breeds have spread all over the world, moving aside the indigenous ones and increasing the danger of their extinction (Hall and Ruane, 1993). As the extinction rates are bigger than the rates of creation of new breeds, a proper estimation of the breeds’ status of endangerment is required. During the years various systems for categorization of the endangered breeds of livestock on a national (Ruane, 2000) and international level (Avon, 1992; Loftus and Scherf, 1993; Simon and Buchenauer, 1993; Alderson, 2003; Reist-Marti et al, 2003) have been developed. These systems use different parameters and thresholds, which makes their comparison difficult. An attempt was made by Gandini et al. (2004) towards a uniform European criterion based on the data collected in the European Association for Animal Production (EAAP) database (EAAP-AGDB).

The aim of this study was to develop a well-defined automatically calculable criterion for classifying the livestock breeds according to degree of endangerment (DE) of their loss along the proposal presented by Gandini et al. (2004). This criterion should take into account the species specificity, state of the global population and genetic concepts. Special attention is paid to the problem of transboundary breeds. These breeds are present in more than one country and a genetic material is usually exchanged between the populations in the various countries. The currently existing criteria treat the population in each country as a separate breed, thus e.g. listing the Dexter cattle in Germany as critically endangered (Scherf, 2000). There is a big Dexter population in the United Kingdom (UK) and in the past even some of the German farmers have registered their animals in UK. We propose a method and
software tools to cope with various classifications of genetically similar breeds in estimating the degree of endangerment, which is missing in the most of the currently used criteria.

Material and Methods

The criterion is intended to be used mainly within the European farm animal biodiversity information system (FABISnet) (Rosati et al., 2006) and it is based on the data collected in this network. The criterion has four classes of endangerment - “critical”, “endangered”, “not at risk" and “unknown”, which are calculated on two levels - national and supranational (for transboundary breeds and groups of genetically similar breeds). The available data, the components of the criterion and the procedures for calculation on the various levels are described in this section.

Table 3.1: Data taken from EFABIS database in May 2006 (Verwendete Daten aus der EFABIS Datenbank im Mai 2006)

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of breeds</th>
<th>Number of population records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle</td>
<td>958</td>
<td>2319</td>
</tr>
<tr>
<td>Pig</td>
<td>509</td>
<td>1005</td>
</tr>
<tr>
<td>Sheep</td>
<td>1105</td>
<td>2437</td>
</tr>
<tr>
<td>Goat</td>
<td>314</td>
<td>737</td>
</tr>
<tr>
<td>Horse</td>
<td>750</td>
<td>1761</td>
</tr>
<tr>
<td>Ass</td>
<td>51</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>3687</td>
<td>8342</td>
</tr>
</tbody>
</table>

Data

In the databases of this network more than 5720 domesticated breeds or ‘country populations’ from 28 species in Europe are registered (data from May 2006). On regular intervals data for the population size and structure is supposed to be collected for each breed. These data include number of breeding males and females, overall population size and trend, status of registration, number of herds, average herd size, usage of artificial insemination and presence of conservation programmes. The data used in our work was taken from the European regional database in year 2006 and represent 3687 breeds from cattle, pig, sheep, goat, horse and ass
species. The number of breeds and available population data are shown in Table 3.1. In order to apply the criterion on supranational level, 8 Landrace, 3 Duroc, 4 Pietrain, 4 Large White and 2 Hampshire pig breeds were grouped on the basis of the results of the PigBioDiv program (SanCristobal et al., 2006). In this program the various populations in several European countries were genotyped for 50 microsatellite loci and genetic distances have been estimated. The breeds groups are shown in Table 3.2.

**Criterion for estimating the degree of endangerment (DE) on national level**

The criterion is based on demographic and genetic components. For each of these components a class of endangerment is estimated. The final class estimating the DE of the breed is the “worst” of the demographic and genetic classes.

**Demographic component:** Gandini et al. (2004) proposed as a first component of assessing DE of a breed the time to reach critical female population size. This time is calculated as

\[ t = \frac{\log N_x - \log N_0}{\log r} \]

where \( N_0 \) and \( N_x \) are the population size in the initial moment and the projected size, and \( r \) is the proportional growth rate. The rate of growth can be estimated as

\[ r = \sqrt[\frac{N_2}{N_1}] \]

the number of females in the population in years \( Y_1 \) and \( Y_2 \) respectively, and \( t=Y_2-Y_1 \). This proposal is well suited to assess DE of a single breed in a national context. However, its application may produce misleading results in estimating the DE in context of a group of genetically similar breeds from various countries. This will be illustrated by the following example: Let us have a transboundary breed in two countries – C₁ and C₂. In C₁ let us assume a population size of \( N_1^1 = 5000 \) and \( N_1^2 = 5000 \) in year 2000 and 2001 respectively. The population sizes in country C₂ in the same years are \( N_2^1 = 5000 \) and \( N_2^2 = 1000 \).
Table 3.2: Grouping of 21 pig populations from eight European countries. The groups are: DU- Duroc, HA- Hampshire, LR – Landrace, LW – Large White, PI – Pietrain. In the last column is the breed code from PigBiodiv Project (San Cristobal et al., 2006) (Gruppierung der 21 Schweinepopulationen aus 8 europäischen Ländern. Die Gruppen sind: DU- Duroc, HA- Hampshire, LR – Landrace, LW – Large White, PI – Pietrain. In der letzten Spalte wird der Rassencode des PigBioDiv Projekts (San Cristobal et al., 2006) angegeben)

<table>
<thead>
<tr>
<th>Most common name</th>
<th>Transboundary name</th>
<th>Country</th>
<th>Group</th>
<th>PigBioDiv code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duroc</td>
<td>Duroc</td>
<td>United Kingdom</td>
<td>DU</td>
<td>GBDU02</td>
</tr>
<tr>
<td>Duroc</td>
<td>Duroc</td>
<td>Italy</td>
<td>DU</td>
<td>ITDU01</td>
</tr>
<tr>
<td>Duroc</td>
<td>Duroc</td>
<td>Germany</td>
<td>DU</td>
<td>DEDU03</td>
</tr>
<tr>
<td>Hampshire</td>
<td>Hampshire</td>
<td>United Kingdom</td>
<td>HA</td>
<td>GBHA01</td>
</tr>
<tr>
<td>Hampshire</td>
<td>Hampshire</td>
<td>Germany</td>
<td>HA</td>
<td>DEHA02</td>
</tr>
<tr>
<td>British Landrace</td>
<td>British Landrace</td>
<td>United Kingdom</td>
<td>LR</td>
<td>GBLR10</td>
</tr>
<tr>
<td>Danish Landrace</td>
<td>Danish Landrace</td>
<td>Denmark</td>
<td>LR</td>
<td>DKLR04</td>
</tr>
<tr>
<td>German Landrace</td>
<td>German Landrace</td>
<td>Germany</td>
<td>LR</td>
<td>DELR14</td>
</tr>
<tr>
<td>DL-1970</td>
<td></td>
<td>Denmark</td>
<td>LR</td>
<td>DKLR05</td>
</tr>
<tr>
<td>Landrace</td>
<td>French Landrace</td>
<td>France</td>
<td>LR</td>
<td>FRLR01</td>
</tr>
<tr>
<td>Landrace Italiana</td>
<td>Italian Landrace</td>
<td>Italy</td>
<td>LR</td>
<td>ITLR03</td>
</tr>
<tr>
<td>Maatiaissika</td>
<td>Finnish Landrace</td>
<td>Finland</td>
<td>LR</td>
<td>FILR06</td>
</tr>
<tr>
<td>Norsk Landrace</td>
<td>Norwegian Landrace</td>
<td>Norway</td>
<td>LR</td>
<td>NOLR08</td>
</tr>
<tr>
<td>Large White</td>
<td>Large White</td>
<td>United Kingdom</td>
<td>LW</td>
<td>GBLW05</td>
</tr>
<tr>
<td>Large White</td>
<td>Large White</td>
<td>Germany</td>
<td>LW</td>
<td>DELW02</td>
</tr>
<tr>
<td>Large White</td>
<td>Large White</td>
<td>France</td>
<td>LW</td>
<td>FRLW01</td>
</tr>
<tr>
<td>Large White</td>
<td>Large White</td>
<td>Italy</td>
<td>LW</td>
<td>ITLW03</td>
</tr>
<tr>
<td>Pietrain</td>
<td>Pietrain</td>
<td>Germany</td>
<td>PI</td>
<td>DEPI03</td>
</tr>
<tr>
<td>Pietrain</td>
<td>Pietrain</td>
<td>Belgium</td>
<td>PI</td>
<td>BEPI01</td>
</tr>
<tr>
<td>Pietrain</td>
<td>Pietrain</td>
<td>United Kingdom</td>
<td>PI</td>
<td>GBP04</td>
</tr>
</tbody>
</table>

If we sum up both populations we will have population size in year 2000: \(N_1=10000\) and in 2001: \(N_2=6000\). The rate of growth of the whole population is \(r = \frac{3}{5}\) and the time to reach population size of 100 will be \(t = \left( \frac{\log 100 - \log 6000}{\log \frac{3}{5}} \right) \approx 8\) years. But if we
look at the population in country $C_1$, we will notice that it is stable ($r_1=1$) and therefore
the total population size will never go under 5000. This discrepancy shows that such
simplistic approach can be misleading in case of breeds present in more than one
country.

To keep the demographic part of the criterion simple and uniform for the cases
of autochton and transboundary breeds, we propose to calculate the population size
after two generation intervals (GI) from the last census and compare it with the critical
values. The calculation is done with the formula $N_{exp} = N_0 r^{2GI}$, where GI is the
generation interval for the species. For estimation of the growth rate we use the most
recently recorded population size (no more than one GI back from the year for which
we want to estimate DE) and the earliest population size available in the interval one
to three GI from the most recent one. Thus, we keep to some extent with the idea of
time, needed to react to the negative trend. The proposition of Gandini et al. (2004)
gives us the exact time we have for reaction and with this approach the results can
be used directly to schedule any changes in the breeding programs. In the new
criterion we presume that the minimum time for reaction is two GI and therefore
consider breeds endangered if they require immediate intervention preventing them
falling under critical value in the near future.

We have used the same thresholds as the ones used by the Food and
Agriculture Organization (FAO) of the United Nations (Scherf, 2000), with the
intention to keep uniformity with their world-wide criterion. We consider the following
classes of endangerment:

1. not at risk – $N^{2GI,f} > 1000$, where $N^{2GI,f}$ is the expected number of
breeding females after 2 generation intervals
2. endangered – $100 < N^{2GI,f} \leq 1000$
3. critical - $N^{2GI,f} \leq 100$

If there are not enough data to estimate $N^{2GI,f}$, we use the value “unknown”.

**Genetic component:** Here we use as a parameter the effective population size $N_e$, a factor which reflects the increase of inbreeding. The effective population size
assuming no selection is calculated by the Wright formula \( N_e = \frac{4N_mN_f}{N_m + N_f} \) (Falconer and Mackay, 1996), where \( N_m \) and \( N_f \) are the number of breeding males and females, respectively. As noted in (Gandini et al., 2004) the effective population size explain for mass selection, since the presence of unselected populations is not very likely in Europe. Therefore, we calculate the effective population size assuming mass selection in a simplified way as \( N_{e\text{sel}} = \frac{7}{10} N_e \) following the model described in (Santiago and Caballero, 1995). To be coherent with the EAAP criterion, we have used the same thresholds as described by Simon (1999). These thresholds are based on the maximum acceptable inbreeding \( F-50 \) after 50 years of conservation. We have transformed the five classes used by EAAP into three:

1. not at risk - \( F-50 \leq 10\% \)
2. endangered - \( 10\% < F-50 \leq 40\% \)
3. critical - \( 40\% < F-50 \)

These thresholds for \( F-50 \) are converted for \( N_e \) through the rate of inbreeding \( \Delta F \). This is done using the formula \( F_t = 1 - (1 - \Delta F)^t \) (Falconer and Mackay, 1996), where \( F_t \) is the inbreeding coefficient in generation \( t \). In this formula \( F_t \) is replaced with the thresholds for \( F-50 \) and \( t \) with the number of generations in 50 years for the respective species (Table 3.3) and the equation is solved for \( \Delta F \). The result is replaced in the formula \( N_e = \frac{1}{2\Delta F} \) (Falconer and Mackay, 1996) and the result are the thresholds for \( N_e \). The respective classes per species are shown in Table 3.3. If there are not enough data to estimate class of endangerment, we used the value “unknown”.

**Criterion procedure**: The criterion is supposed to be calculated for each breed for a certain year. The estimated DE class should be consider reliable for no more than two generation intervals. The procedure consists of the following steps for the year \( Y \):
Chapter 3  Early warning system for loss of diversity

Table 3.3: Classes of endangerment based on the effective population size $N_e$. GI is the average generation interval for the species, NG50 is the number of generations for 50 years. (Klassen des Gefährdungsstatus auf der Basis der effektiven Populationsgröße $N_e$. GI entspricht dem durchschnittlichen Generationsintervall für die jeweilige Spezies, NG50 ist die Anzahl der Generationen innerhalb von 50 Jahren.)

<table>
<thead>
<tr>
<th>Species</th>
<th>GI</th>
<th>NG50</th>
<th>Critical(3)</th>
<th>Endangered(2)</th>
<th>Not at risk(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pig</td>
<td>1.5</td>
<td>33</td>
<td>&lt;33</td>
<td>33-156</td>
<td>&gt;156</td>
</tr>
<tr>
<td>Sheep/goat</td>
<td>2.5</td>
<td>20</td>
<td>&lt;20</td>
<td>20-94</td>
<td>&gt;94</td>
</tr>
<tr>
<td>Cattle</td>
<td>3.5</td>
<td>14</td>
<td>&lt;14</td>
<td>14-66</td>
<td>&gt;66</td>
</tr>
<tr>
<td>Horse/ass</td>
<td>4.5</td>
<td>11</td>
<td>&lt;11</td>
<td>11-51</td>
<td>&gt;51</td>
</tr>
</tbody>
</table>

Firstly, we find the maximal year $Y_2$ with available data in the interval $(Y-GI, Y]$, where GI is the species generation interval shown in Table 3.3. In the best case $Y_2=Y$, but it may happen that the data was reported one or two years earlier (we presume that data is reported at least once each GI). Based on $Y_2$ we found the minimal year $Y_1$ in the interval $[Y_2-3GI, Y_2-GI]$ with population data and estimate the growth rate as $r = e^{\frac{\log N_f^2 - \log N_f^1}{t}}$, where $t=Y_2-Y_1$ and $N_f^2$ and $N_f^1$ are the number of breeding females in years $Y_2$ and $Y_1$ respectively. The value of the expected female population size after two generation intervals $N_f^{Y_2+2GI}$ was calculated as $N_f^{Y_2+2GI} = N_f^2 e^{2GI}$. The result was compared with the thresholds 100 and 1000 as shown in the demographic part and we obtain class of endangerment $C_1$.

The next step is to calculate the effective population size in year $Y_2$ assuming mass selection: $N_e^2 = \frac{7}{10} \frac{4N_m^2 N_f^2}{N_m^2 + N_f^2}$, where $N_m^2$ and $N_f^2$ are the number of breeding males and females in year $Y_2$. This value ($N_e^2$) was compared with the thresholds for the respective breed (Table 3.3) and as a result we got class of endangerment $C_2$.

The final class of DE is the 'worst' of the classes $C_1$ and $C_2$. 

64
**Criterion for estimating the degree of endangerment (DE) on supranational level**

The criterion on supranational level has the same demographic and genetic components as the ones used on national level. However, when the degree of endangerment of a breed is estimated on supranational level the status of the global population should be taken also into account. In case of transboundary breeds, the global population is the sum of the populations in the various countries. If we have grouping of genetically similar breeds, the global population are all animals within the group. To handle this factor we have an additional parameter to the criterion - a list with grouping of the breeds. The list contains the breeds identification and the group, they belong to. With this additional parameter the following procedure is used to calculate the degree of endangerment of a breed from one group for the year $Y$.

As described before, the idea of the demographic part of the criterion is to estimate the number of breeding females in the population after two generation intervals. However, it was also demonstrated before that if we simply sum up the female populations numbers in various countries and apply to the result the national level criterion, we may obtain misleading results. Therefore, we have estimated the expected number of breeding females in each population separately (using the national criterion) and then summed up the results to obtain the expected number of females in the whole breed. Thus, for each breed $B_i$ from the group we calculated:

\[
Y_i^2 = \max\{\text{year } y \text{ with data } | \ y \in (Y - GI, Y]\}
\]

\[
Y_i^1 = \min\{\text{year } y \text{ with data } | \ y \in [Y - 3GI, Y - GI]\}
\]

\[
t_i = Y_i^2 - Y_i^1
\]

\[
r_i = e^{\frac{\log N_i^{1,i} - \log N_i^{2,i}}{t_i}}, \text{ where } N_i^{1,i} \text{ and } N_i^{2,i} \text{ are the number of breeding females in year } Y_i^1 \text{ and } Y_i^2 \text{ respectively.}
\]

\[
N_i^{Y_i^{1+2GI,i}} = N_i^{2,i} r_i^{2GI}
\]
The expected number of breeding females in the whole population after two
generation intervals was calculated as \( N_{f,exp} = \sum_i N_{f,i}^{Y_2 + 2GI} \). On the regional level we
used the same thresholds as on the national one, i.e. 100 is the upper limit for
critically endangered breed and 1000 – for endangered. In this manner we obtained
from the demographic component class of endangerement \( C_1 \).

For the genetic component we summed the number of breeding females and
males for each breed in year \( Y_2 \): \( N_{f,Y_2} = \sum_i N_{f,i}^{Y_2} \) and \( N_{m,Y_2} = \sum_i N_{m,i}^{Y_2} \). Further, we treated
\( N_{f,Y_2} \) and \( N_{m,Y_2} \) as a number of breeding animals from single breed and applied to
these numbers the demographic part of the criterion on national level. In this manner
we obtained the class of endangerment \( C_2 \). Similar to the single breed case we took
the ‘worst’ of the \( C_1 \) and \( C_2 \) and assigned it to the whole group. In such manner each
breed in the group has the same DE class as the others in the group.

One issue that needs attention is the way how breeding females and males
numbers are selected. It may happen that the numbers for the various breeds are not
taken from the same year. However all years with data should be in the interval \( (Y-
GI, Y) \). The intervals length is one GI, and we supposed that during such interval of
time the population numbers will not change too much. On the other hand, if data for
one or more breeds are missing in this time interval, this may influence the results
significantly. Therefore, we consider important as part of the results, also the number
of breeds from the group, which have been used in the calculation of DE.

**Discussion**

The criterion proposed here uses a small number of parameters (number of
breeding males and females), which have been also used in the FAO and EAAP
criteria. Many more parameters can be added, e.g. presence of organization
monitoring the breed, number and average size of herds, average age of the farmers,
distance between the animals, etc. Although they have an impact on the DE of the
breed, this impact is somehow arbitrary. The first attempt for creating a
comprehensive system for estimating breeds extinction probabilities was done by
Reist-Marti et al. (2003). This system operates on similar set of data, taken from the Domestic Animal Diversity Information System (DAD-IS) (FAO) and the Domestic Animal Genetic Resources Information System (DAGRIS). It uses 10 parameters, including from total population size and organization of the farmers to political situation in the country and sociocultural importance of the breed. Unfortunately, many of the variables used by Reist-Marti et al. (2003) require arbitrary assignment of values by a human, which makes them not suitable for automated calculation.

The estimation of the growth rate is also simplified and no regression techniques like the ones used by Bennewitz and Meuwissen (2005) are applied. This is done mainly due to data insufficiency. We also wanted to keep the calculation procedure comprehensive to a wider group of users as the main intention in using this criterion is to raise attention to the process of breed extinction.

In the demographic part of the criterion we used as a main parameter the number of breeding females. Further parameter in this section could be the number of annually registered female animals. As pointed out by Alderson (2003), this number is a more accurate indication of the viability of the population. Unfortunately, these data are not collected in the current databases in the FABISnet, and we cannot include new data in the data collection process.

Table 3.4: Assessment of endangerment of the European breeds in year 2005. (Schätzung des Gefährdungstatus der europäische Rassen- Jahre 2005.)

<table>
<thead>
<tr>
<th>Species/class</th>
<th>Unknown</th>
<th>Not at risk</th>
<th>Endangered</th>
<th>Critical</th>
<th>Total per species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle</td>
<td>639</td>
<td>90</td>
<td>46</td>
<td>27</td>
<td>802</td>
</tr>
<tr>
<td>Sheep</td>
<td>812</td>
<td>52</td>
<td>34</td>
<td>36</td>
<td>934</td>
</tr>
<tr>
<td>Goat</td>
<td>257</td>
<td>22</td>
<td>5</td>
<td>13</td>
<td>297</td>
</tr>
<tr>
<td>Pig</td>
<td>380</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>399</td>
</tr>
<tr>
<td>Horse</td>
<td>456</td>
<td>85</td>
<td>61</td>
<td>66</td>
<td>668</td>
</tr>
<tr>
<td>Ass</td>
<td>31</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>47</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2575</strong></td>
<td><strong>256</strong></td>
<td><strong>161</strong></td>
<td><strong>155</strong></td>
<td><strong>3147</strong></td>
</tr>
</tbody>
</table>

The criterion was applied on a national level to 3147 from the 3687 European breeds as the other 540 breeds were extinct. The results for year 2005 (Table 3.4) raise on the first place concern about the quality of data collected in the database.
For 82% of the breeds there are no enough data to estimate status of endangerment. The most worrying is the lack of data for pigs, where 95% of the breeds have status “unknown”. We have to remind that the search for available data is done not only in the target year, but also one GI back. In case of ass breeds, this means a serious lack of data in the time interval 2000-2005. The EAAP and FAO criteria were applied to the same time interval and the results shown in Table 3.5 confirmed that data are incomplete. The application of the criterion to other years resulted also in a relative big proportion of a breeds with unclear status. This is a strong indication for a need of a special methodology and tools to complete the data collected in the FABISnet.

Table 3.5: Comparison between the various criteria used in Europe. FAO- criterion used by FAO as defined in (Scherf, 2000); EAAP – a modified version of EAAP criterion as used in EFABIS.(Vergleich der verschiedenen Kriterien für den Gefährdungsstatus in Europa. Das FAO-Kriterium entsprach der Definition von Scherf (2000); EAAP – ein modifiziertes Kriterium, das in der EFABIS Datenbank verwendet wird.)

<table>
<thead>
<tr>
<th>Class/criteria</th>
<th>FAO</th>
<th>EAAP</th>
<th>The uniform criterion(national level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>2476</td>
<td>2537</td>
<td>2575</td>
</tr>
<tr>
<td>Not at risk</td>
<td>263</td>
<td>332</td>
<td>256</td>
</tr>
<tr>
<td>Endangered</td>
<td>230</td>
<td>278</td>
<td>161</td>
</tr>
<tr>
<td>Critical</td>
<td>178</td>
<td>0</td>
<td>155</td>
</tr>
</tbody>
</table>

When applying the criterion on supranational level it is important to notice that there is no uniform grouping of breeds. Such grouping can be done based not only on the genetic distance between breeds, but also on phenotypic traits, sociocultural and environmental role of the breed. The grouping can be done also with a special conservation interest and as stated by Ruane (1999) the genetic distance is not the only important factor. Therefore, different organizations and scientists may have different opinions on the grouping of the breeds. As an example we can use the
Pustertaler Sprinzen breed. One grouping can be based on the presence of the same breed in several countries, e.g. the populations in Germany, Austria and Italy. Another grouping can be based on the closer relation of the German Pustertaler to the Pinzgauer breed as reported by Edwards et al. (2000), or we can use the whole Pinzgauer group as defined by EAAP on their web page (EAAP-AGDB).

Table 3.6: Assessment of the DE on a supranational level for 21 pig breeds in year 2002. The classes of the demographic and genetic component are shown together with the number of breeds used to estimate the group class. Classes of endangerment: 0 - unknown; 1 - not at risk; 2 - endangered; 3 – critical. (Schätzung der Gefährdungsstatus (DE) auf einem übernationalen Level für 21 Schweinerassen in Jahre 2002. Zu den Klassen der demographischen und genetischen Komponente wird die Anzahl der Rassen angegeben, die für Einordnung der Gruppen in Klassen verwendet wurden. Klassen für den Gefährdungsstatus: 0 – unbekannt; 1 – nicht gefährdet; 2- gefährdet; 3 – kritisch gefährdet.)

<table>
<thead>
<tr>
<th>Breed name</th>
<th>Country</th>
<th>Class $N_f$</th>
<th>Class $N_e$</th>
<th>Class $N_{ef}^{group}$</th>
<th>Breeds used</th>
<th>Class $N_{e}^{group}$</th>
<th>Breeds used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hampshire</td>
<td>Germany</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1/2</td>
<td>2</td>
<td>1/2</td>
</tr>
<tr>
<td>Hampshire</td>
<td>UK</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1/2</td>
<td>2</td>
<td>1/2</td>
</tr>
<tr>
<td>Duroc</td>
<td>UK</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2/3</td>
<td>1</td>
<td>2/3</td>
</tr>
<tr>
<td>Duroc</td>
<td>Germany</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2/3</td>
<td>1</td>
<td>2/3</td>
</tr>
<tr>
<td>Duroc</td>
<td>Italy</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2/3</td>
<td>1</td>
<td>2/3</td>
</tr>
<tr>
<td>Large White</td>
<td>Germany</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0/4</td>
<td>1</td>
<td>2/4</td>
</tr>
<tr>
<td>Large White</td>
<td>Italy</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0/4</td>
<td>1</td>
<td>2/4</td>
</tr>
<tr>
<td>Large White</td>
<td>UK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0/4</td>
<td>1</td>
<td>2/4</td>
</tr>
<tr>
<td>Large White</td>
<td>France</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0/4</td>
<td>1</td>
<td>2/4</td>
</tr>
<tr>
<td>Pietrain</td>
<td>UK</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>Pietrain</td>
<td>Belgium</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>Pietrain</td>
<td>Germany</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>Piétrain</td>
<td>France</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>British Landrace</td>
<td>UK</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Dansk Landrace</td>
<td>Denmark</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Deutsche</td>
<td>Germany</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Landrassse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL-1970</td>
<td>Denmark</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Landrace</td>
<td>France</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Landrace Italiana</td>
<td>Italy</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Maatiaissika</td>
<td>Finland</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
<tr>
<td>Norsk Landrace</td>
<td>Norway</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2/8</td>
<td>1</td>
<td>4/8</td>
</tr>
</tbody>
</table>
To illustrate the application of the criterion the results for year 2002 for five pig groups are shown in Table 3.6, together with the results of the respective breeds on the national level. In the case of the Landrace group data for only two to four out of the eight breeds in the group were available, but even these data were enough to classify the breeds from the group as not endangered. Similar is the situation in Pietrain, where only the German population (6445 females and 4413 males) was beyond the thresholds for being endangered.

In the Duroc group data only from Germany and UK were available. As a single breed the German Duroc population is classified as endangered by our system and by FAO criterion. On the other hand the status of the Duroc group is clearly “not endangered”, which is the real status of this international commercial breed.

On the opposite side as data is the Hampshire group. Here only the German population has recorded data which allows estimation of the risk status. As an individual breed the German Hampshire is critically endangered and this is also the status of the group. The problem here was that the estimation of the group class \( N_r \) was based on only one out of two breeds, which makes the results unreliable.

Therefore, the reliability of the group status estimation should be always presented as a part of the results. This can be the number of breeds used in the calculation and the total number of breeds or the percentage of the used breeds. The former has some advantages, as some additional conclusions about the DE can be made from the size of the group. Moreover, in case of incomplete data we strongly suggest that the results of the particular classes are shown together with the final status. This will allow the end-users to critically evaluate the reliability of the results and thus contribute for more objective interpretation.

**Conclusion**
The proposed criterion is based on a minimum number of parameters that are already in the data collection scheme of FABISnet. However, the data incompleteness in the databases of this network, will prevent the regular usage of this criterion at the present moment, but it is a step towards common criterion in Europe. The major effort at this point of time should be targeted for completing the data. This will allow more complex techniques to be applied in the future for the estimation of the degree of endangerment of breed extinction.

The criterion has a novel approach to the status of endangerment of the breeds on a supranational level. It allows various grouping of populations of transboundary and international breeds and genetically similar breeds. The results in a single and group context, presented together with their reliability, provide more objective information about the degree of endangerment.

**Availability**

The criterion was implemented as OpenSource software module and is available from the authors.
References


Chapter 4

Improving the monitoring of animal genetic resources on National and International level

Zhivko Duchev and Eildert Groeneveld

Archives of Animal Breeding (Accepted)
The extent of Zhivko Duchev’s contribution to the article is evaluated according to the following scale:

A. has contributed to collaboration (0 - 33%)
B. has contributed significantly (34 - 66%)
C. has essentially performed this study independently (67 - 100%)

1. Design of the project C
2. Analysis of the data C
3. Implementation into software module C
4. Presentation and discussion of the study in article form C
Abstract

The Farm Animal Biodiversity Network (FABISnet) is a new biodiversity network for collecting domestic animal breeds data from the European countries. Data are collected on National, Regional and Global level and can be automatically transferred between the levels. As a successor of the Animal Genetic Data Bank of the European Association for Animal Production (EAAP-AGDB) and the Domestic Animal Diversity Information System (DAD-IS) of the Food and Agricultural Organization (FAO), it has taken their historic data and integrated them in a network of databases. However, the data are incomplete. For example the general description of 57% of the registered European breeds are very sparse and for more than 3900 breeds the population size and structure statistics are outdated. A set of 13 management support reports and a methodology for their application has been developed. The reports present summarized information about the degree of completeness of the breeds descriptions by country, missing or incomplete population records, reminders for updating data and the status of data translation. Various monitoring/reporting tasks of the National and Regional Coordinators for management of animal genetic resources can be organized in a systematic manner using these reports. Such an organizational scheme can reduce the time spent in completing data and improve the content of each database in the network.

Key Words: Data management, Farm animal biodiversity, Information systems
Zusammenfassung

Titel der Arbeit: Strategie zur Verbesserung des Monitoring Nationalen und Internationalen Tiergenetische Ressourcen


Schlüsselwörter: Datenmanagement, Biodiversität landwirtschaftlicher Nutztiere, Informationssysteme
Chapter 4  Improving the monitoring of animal genetic resources

Introduction

In the past efficient use of modern techniques of genetic evaluation and selection along with achievements in biotechnology have lead to remarkable improvements in performance of animal production. As a consequence some populations and breeds fell behind as they got under the pressure from their improved competitors. This issue was already addressed in 1979 (Züchtungskunde, 1979), when the creation of a genebank was suggested. To monitor the status of farm animal genetic resources (AnGR) the database of the European Association for Animal Production (EAAP) was set up in Hanover (EAAP-AGDB, 2004). When FAO developed its world wide Domestic Animal Diversity Information System (DAD-IS), the EAAP-AGDB served as a starting point for development. With the EU funded project "European Farm Animal Biodiversity Information System" (EFABIS) a new network got developed which will replace the old EAAP-AGDB as well as FAO's DAD-IS 2 (FAO, 2005) by a uniform software system (Rosati et al, 2006). The developed network is an exhaustive multilingual source of information about the characterization, conservation and utilization of AnGR in Europe. The software for operating the network was developed in such a general manner, that it can be used for the establishment of biodiversity networks in other regions of the world.

The network operates on three levels - National, Regional and Global. Within the network countries can establish their own Web-driven National farm animal biodiversity databases in one or more local languages, free of licensing costs as this was done for Poland (NRIAP, 2006). The subset of data required on the upper levels in one of the official FAO languages will be automatically transferred from the National databases to the Regional one (EAAP, 2006) via the synchronization protocol developed in EFABIS (Duchev and Groeneveld, 2006). After the implementation of the new network the focus is now on the improvement of its data quality.

The databases in the network collect data for breeds of domesticated animals in more than 35 mammalian and avian species. Each database is organized in five sections: breed data, library of publications, references and Web-links, image gallery and contacts data. The breed data section contains information about the origin and
development of the breed, morphology and performance traits, utilization and conservation. An important part of the breed section is the population size and structure, which should be entered on a regular basis. This information is used for example in monitoring the population dynamics and trends, as well as in estimating the degree of endangerment of the breed according to the FAO (Scherf, 2000) and the EAAP criteria (Simon and Buchenauer, 1993).

In this study the information requirements for successful management of domestic animal genetic resources data in the FABISnet are analysed. The use of the set of reports developed in the EFABIS is presented and illustrated with examples. Several propositions are made on how to use this management support system worldwide in the FAO structure of National and Regional Coordinators to improve the quality of Animal Genetic Resources (AnGR) data.

Material and Methods

In this section management problems in terms of data completeness for the areas of breed and population data on one hand and translation/synchronization on the other will be described. This is followed by a proposal for a solution.

The Problem

The current content of the European database is based on the merger of the data from EAAP-AGDB with the National data stored in DAD-IS. In May 2006 the European data represented 5720 breeds from 28 species in 47 countries. Within each country, breed data are collected, summarized and entered by possibly several persons. This is especially the case in countries with large number of breeds, e.g. France - 493, Germany - 459, United Kingdom - 578. The information required (more than 100 fields) is collected from various sources - breeding societies, farmers, scientific institutions and is not always available at once. Therefore, entering data in the database is probably done in an incremental mode leading to gaps in the data. Moreover, the new network of multilanguage databases presents added challenges for administration and monitoring of data translation and exchange.

General Breeds Data: The general breeds database block contains the description
of the live and extinct breeds. The degree of completeness (DC) of the European breeds description is shown in Table 4.1. The DC was calculated as the ratio of the number between filled and all possible data fields per data block. Additionally, the DC was calculated across all blocks. More than half of the breeds - 57% have less than 10% DC which means that almost no data are recorded for these breeds. The description of another 2063 breeds has DC less than 30%. Only one breed has a degree of fill more than the 60% in the description fields. Although 100% DC is impossible, such gaps indicate the need for a systematic management approach to improve data completeness.

Table 4.1: Degree of completeness of the European breeds description. (Grad der Vollständigkeit der Europäischen Rassenbeschreibungen)

<table>
<thead>
<tr>
<th></th>
<th>&lt;10</th>
<th>10-29</th>
<th>30-59</th>
<th>&gt;59</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>642</td>
<td>2877</td>
<td>1740</td>
<td>461</td>
</tr>
<tr>
<td>Origin and development</td>
<td>85</td>
<td>3056</td>
<td>1719</td>
<td>860</td>
</tr>
<tr>
<td>Morphology</td>
<td>1959</td>
<td>2046</td>
<td>689</td>
<td>1026</td>
</tr>
<tr>
<td>Performance</td>
<td>4970</td>
<td>656</td>
<td>78</td>
<td>16</td>
</tr>
<tr>
<td>Additional information</td>
<td>5025</td>
<td>556</td>
<td>138</td>
<td>1</td>
</tr>
<tr>
<td>All blocks</td>
<td>3277</td>
<td>2063</td>
<td>379</td>
<td>1</td>
</tr>
</tbody>
</table>

Population Counts Data: In addition to the detailed breed description we also need information about the population size and structure over the years. Such time series are required to investigate patterns in population dynamics and extinction probabilities as it is done by Bennewitz and Meuwissen (2005). A "long" time series of population data is needed also for development of a uniform criterion for assessing the degree of endangerment of breed, as done by Gandini et al. (2004). In this regard parameters like effective population size and rate of inbreeding, estimated on the number of breeding males and females, can be compared with the results of the population analysis.(e.g. Biedermann et al., 2004 for the Vorderwald cattle).

Here data are seriously lacking: as shown in Table 4.2, from 5720 European breeds registered in the database 1736 do not have any population records and therefore their status of endangerment is unknown. From the rest, 2208 breeds have population data only for a single year, 1731 - between two and nine years and only
45 breeds have population data for ten or more years. Thus, only a minor part of the database can be used for studying population dynamics.

Table 4.2: Distribution of the population records of the European breeds. (Verteilung der Populationsdaten Europäischer Rassen)

<table>
<thead>
<tr>
<th>Breeds/Populations</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>&gt;9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheep</td>
<td>262</td>
<td>355</td>
<td>151</td>
<td>82</td>
<td>37</td>
<td>68</td>
<td>80</td>
<td>34</td>
<td>12</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Pig</td>
<td>112</td>
<td>211</td>
<td>41</td>
<td>43</td>
<td>28</td>
<td>27</td>
<td>19</td>
<td>12</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Goat</td>
<td>80</td>
<td>81</td>
<td>40</td>
<td>28</td>
<td>18</td>
<td>27</td>
<td>14</td>
<td>17</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Horse</td>
<td>189</td>
<td>240</td>
<td>92</td>
<td>31</td>
<td>28</td>
<td>48</td>
<td>27</td>
<td>67</td>
<td>7</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Cattle</td>
<td>207</td>
<td>317</td>
<td>100</td>
<td>71</td>
<td>60</td>
<td>52</td>
<td>65</td>
<td>36</td>
<td>26</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Ass</td>
<td>23</td>
<td>15</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>All species</td>
<td>1772</td>
<td>2118</td>
<td>540</td>
<td>310</td>
<td>179</td>
<td>223</td>
<td>206</td>
<td>166</td>
<td>58</td>
<td>40</td>
<td>45</td>
</tr>
</tbody>
</table>

Translations and Data Synchronization: National databases may contain textual descriptions in one or more of the National languages. However, data to be send to the Global database must be in one of the official FAO languages - English, French, Spanish, Arabic or Chinese. Therefore, we need a mechanism to monitor and manage the translation of data and ensure that all new or updated data are translated.

Another area, which should be monitored, is the data exchanged among databases via synchronization. An example is the European database, which contains the English version of the World data. The synchronization between the levels is automatic, but the there will be a certain time of inconsistency between EAAP database and the National databases on one hand and EAAP database and FAO database on the other. Therefore we need to check the current content of each database for consistency.

The Solution

The problems described above can be solved by presenting the database managers with the appropriate tools for monitoring the database content and a schedule for the systematic organization of this monitoring.

Monitoring: As already stated above, the general breed description contains information about the breeds characterization, conservation and utilization. As most
of the European breeds are already registered in the EAAP and FAO databases the
remaining task is the completion of this data block. To be able to fulfil this task, the
persons responsible for data entry have to be informed about the degree of data
completeness. They should be able to quickly find which items of the description are
missing, so that they can make a targeted effort to fill the gaps.
The other important part of breed data contains the population counts. There are
three tasks in this regard, which require management support:

- Completing data for already reported years
- Reporting population data for antecedent years
- Regular data reports for the current year

In the population data area it is important to identify years for which demographic
statistics are missing or incomplete and to have an overview of all population data
entered for a certain year. Such information about gaps should be available by breed
and also by year. This will allow two different approaches to the task - completing
data breed by breed or year by year for all breeds.

Another area, which requires monitoring, is the data translation. In the National
databases data to be translated includes breeds description, publications and
reference information, image captions and contacts data. The situation in DAD-IS is
similar, data entered in one language will be translated in the other official ones. This
will probably be done by an external translator on a contract basis. In both cases the
responsible persons have to monitor the translation process and keep track what has
been translated up to now and what is left for translation.

If the country has established a National Farm Animal Biodiversity Database it may
include public data from other databases via synchronization. For example, the
publications and links to useful biodiversity Web pages can be automatically loaded
from FAO database into the National one. In such case the person in charge has to
be informed about the amount of ‘foreign’ data, currently stored in the National
database. The situation at the Regional and Global level is similar as these
supranational databases contain subsets of synchronized data from the National
ones. Therefore, the managers of the Regional and Global database should have
summary of data loaded via synchronization. Moreover, they should keep track of the countries, which have not synchronized their data for a long time.

**Organizational Structure:** Most European countries have officially nominated National Coordinator for Management of Animal Genetic Resources (NC). The NC is responsible for reporting their National genetic resources to the international level and therefore she/he must monitor all national data entered in the database.

In multilingual National databases, the translation process should always be monitored by the National Coordinator. This is an obligation of the NC, because she/he is responsible for the data presented to Regional and Global level in an official language. The situation on the Global level is similar: the person monitoring translations should receive information on the current status on a regular basis. Thus she/he will have an idea what National data are available to the world audience.

FAO has created the structure of a Regional Coordinator (RC) for the support and communication of NCs in a region. This structure seems to be ideally suited to also improve AnGR data from a region by having RC take a moderating and encouraging role. In such a network the RC can organize the dissemination of feedback to countries about the amount and completeness of the country data. The RC can also help countries in solving data improvement problems, sharing the experience from other countries in the region. Specifically, it is suggested that the European Regional Focal Point (ERFP) takes over this task for Europe using the EAAP database. It is suggested that the ERFP (as a model for the other RC in the world) invites the countries to update their data and points the attention to the gaps in the reporting of the population statistics on regular intervals. To be efficient in this task the RC should have an overview of the status of the breed data loaded from each country and should also monitor the presence of population data. If breed data are translated in one or more languages on Regional level, we propose that the Regional Coordinator also monitor this process.

Results

In this section the 13 reports developed in EFABIS for monitoring database content along with work schedules for the routine tasks of the National and Regional
Coordinators are presented. The data used in the examples are taken from the European Regional database in April 2006 and are intended for illustration purposes only.

**Monitoring**

The reports for the management support in the FABIS network can be divided in two groups according to their type. The first group contains Web reports, which are accessible only to the National or Regional Coordinators after login in the Web page of the respective database. These reports are generated on the fly and represent the actual status of the data in the database. For National coordinators the reports are generated on the basis of their National data, whereas the Regional coordinators see cumulated statistics over all countries from their region. The reports are created in tabular format and can be copied from the screen into a spreadsheet software, e.g. OpenOffice or MS Office for presentation. This has been done also in some of the examples in this section.

The second group contains reports in Portable Document Format (PDF), which are automatically generated and disseminated via e-mail to the NCs, RC and other participants in the network. These reports are scheduled for automatic execution on a regular basis. For example National Coordinators may receive an e-mail containing the report for the status of data translation once at the end of each year.

**General Breeds Data:** On a country level, the Web reports summarize in one place the DC of the breed data, allowing the NCs to find rapidly incompleteness or gaps. These reports are:

1. Degree of completeness of breed description (Web)
2. Presence of images per breed (Web)
3. Status of reporting by country (Web)

The most important report for completing the general breed description is the "Degree of completeness of breed description by screen". An example for this report is shown in Figure 4.1. In this report the DC of the data loaded for each breed from one country is calculated. To explain the meaning of this report data entry process in a FABIS database needs to be understood. The "Entry data" section consists of five
Chapter 4 Improving the monitoring of animal genetic resources

description blocks - "General", "Origin and Development", "Morphology", "Performance", "Additional information". Each of these blocks represents a group of forms, e.g. the "General" block consists of four forms - "Names", "Other names", "Uses" and "Images". In the report "Degree of completeness of breed description by screen" the degree of fill of data is calculated for each separate form as the ratio between filled and all possible fields on this form. If a form is not applicable to a certain species, e.g. there is no "Eggs" form for cattle, then it is marked with "_" in the report.

<table>
<thead>
<tr>
<th>Breed name</th>
<th>Species</th>
<th>General</th>
<th>Origin and development</th>
<th>Additional information</th>
<th>Morphology</th>
<th>Performance</th>
<th>Milk</th>
<th>Prolificacy</th>
<th>Fries</th>
<th>Eggs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Braunschweig</td>
<td>Cattle</td>
<td>75</td>
<td>75</td>
<td>65</td>
<td>0</td>
<td>85</td>
<td>58</td>
<td>16</td>
<td>71</td>
<td>33</td>
</tr>
<tr>
<td>Entfernter</td>
<td>Cattle</td>
<td>75</td>
<td>85</td>
<td>80</td>
<td>0</td>
<td>85</td>
<td>65</td>
<td>33</td>
<td>66</td>
<td>100</td>
</tr>
<tr>
<td>Mittelbäder</td>
<td>Cattle</td>
<td>75</td>
<td>75</td>
<td>80</td>
<td>0</td>
<td>85</td>
<td>16</td>
<td>50</td>
<td>66</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 4.1: Degree of completeness of the breed description - Switzerland (excerpt) (Grad der Vollständigkeit der Rassenbeschreibung für die Schweiz (Auszug))

The values in the various groups are marked in three colours, depending on the class of completeness of data: "almost empty" if the DC is under 30%; "sparse" between 30 and 60% and "well filled" if the DC is greater than or equal to 60%. Using the red, orange and black colour schema helps the NC to spot quickly the incomplete blocks of data. However, this DC information is only useful together with knowledge of the data fields in the various blocks and should not be interpreted on its own. For example the "Name" form consists of four data elements - "most common name of the breed", "language of the most common name", "description of the most common name" and "transboundary or brand name". It is obvious that 100% degree of completeness is not possible, e.g. the indigenous breeds do not have a transboundary name. In any case, describing the most common name and its language, results in 75% DC for this sub-block.

In our example, which is an excerpt from the report for Switzerland, the "General" and "Origin and development" blocks are generally "well filled". One exception is the
"Images" sub-block for cattle breeds, where no data are present at all. There may be two reasons: there are no images in the database or there are images, but the image descriptions are missing. The "Presence of images per breed" report (Figure 4.2) can be used to check if there are any images loaded for these breeds in the database.

**Figure 4.2: Presence of image data from Switzerland (excerpt) (Bilddaten aus der Schweiz (Auszug))**

On a global level the degree of completeness of data per region can be obtained from the "Status of reporting per country" report. The degree of completeness of data is calculated as the ratio between the filled and all possible data fields in all records entered from a single country. The chart produced on the basis of this report is shown in Figure 4.3. In all regions the DC of data is on average 32%, ranging from 28% in Latin America and the Caribbean region to 35% in Near East and in Europe.

**Figure 4.3: Degree of countries data completeness by region (Grad der Vollständigkeit von Länderdaten nach Regionen)**

From the same report the Regional Coordinator can obtain the DC of data per country in the respective region. An example for region Europe is shown in Figure

---

**Table 4.1:** Count of images per breed for country Switzerland

<table>
<thead>
<tr>
<th>Breed name</th>
<th>Species</th>
<th>Image count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Braunvieh</td>
<td>Cattle</td>
<td>0</td>
</tr>
<tr>
<td>Eringer Rind</td>
<td>Cattle</td>
<td>0</td>
</tr>
<tr>
<td>Evoléner</td>
<td>Cattle</td>
<td>0</td>
</tr>
<tr>
<td>Fibesurgis</td>
<td>Cattle</td>
<td>0</td>
</tr>
</tbody>
</table>
4.4. The average DC of data in Europe is 35% - with minimum of 27% in Bulgaria and in Republic of Moldova up to 45% in Switzerland.

*Figure 4.4: Degree of data completeness in region Europe (Grad der Vollständigkeit in Europa)*

**Population Counts Data:** Six reports can be used in monitoring population data. These are:

4. Breeds with missing population records in the last 10 years (Web)
5. Degree of completeness of the populations records (Web)
6. Countries with missing population records for a year (Web)
7. Population data entered for one year (Web)
8. Annual NC report (pdf)
9. Reminder for entering population data (pdf)

From the report "Breeds with missing population records in the last 10 years" (Figure 4.5) the NC can produce a list of years with no annual statistics. This list can be used to collect the missing data from the breeding organizations or farmers.
Figure 4.5: Missing population data for the German breeds in the last 10 years (excerpt) (Fehlende Populationsdaten Deutscher Rassen der letzten 10 Jahre (Auszug))

<table>
<thead>
<tr>
<th>Breed Name</th>
<th>Species</th>
<th>Year</th>
<th>Total</th>
</tr>
</thead>
</table>

Figure 4.6: Degree of completeness of the population records for four Switzerland breeds (Grad der Vollständigkeit von vier Schweizer Rassen)

The other issue to be addressed is the improvement of the completeness of already existing records in the database. For this purpose the "Degree of completeness of the population records" report can be used. The example shown in Figure 4.6 is created with data from this report. There, several typical situations in reporting of yearly data from the database are presented. The Räthisches Grauvieh and Braunvieh cattle breeds are two examples where data have been reported on a regular basis. In the case of the Braunvieh the degree of completeness in the last five years is over 60% - in the "well filled" zone. This is not the case for the Räthisches Grauvieh, where the DC is under 50%, reaching its minimum in year 2003 with 30% DC. This is also true for the CH-Warmblut horse in 1983, which is an example for gaps in reporting: it has only two years with population data - 1983 and 1999. Such a gap of fifteen years
between the two records and seven years from the last record to our days is too long, even taking into account the long generation interval in horses. A similar situation can be found in the Skudde sheep breed, where only data in years 1998, 1999 and 2001 are recorded in the database. Here the regular reporting is even more important, because this breed is under the endangered-maintained category according to FAOs criteria, with only 259 breeding females in 2001.

In the next report "Population data entered for one year" all population data reported from a country for a chosen year are presented together. The NC can use this report to complete the population statistics in a certain year. Another usage of this report is monitoring the process of loading data for the current year. As the report represents the actual status of the data, it can be printed or stored in spreadsheet at regular intervals. Comparing the difference between two consecutive printouts will give us the changes and data updates made in the mean time.

The report "Reminder for entering population data" (Table 4.3) follows our proposition that Regional coordinator should invite NCs to update their data at regular intervals. The report contains a list of all breeds from a single country, which have outdated population data. Each year the NCs will receive by e-mail the list of national breeds with last population data older than the 2-3 years and invitation from the RC to update the database. A copy of the report can be also send to the RC.

**Table 4.3: Invitation to Bulgaria for updating the population data (excerpt).**

*Erinnerung an Bulgarien zum Aktualisieren von Populationsdaten (Auszug)*

Population data reminder for year 2006

<table>
<thead>
<tr>
<th>Breed name</th>
<th>Species</th>
<th>Most recent year with population data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayrshire</td>
<td>Cattle</td>
<td>2002</td>
</tr>
<tr>
<td>Carakachanski Kon</td>
<td>Horse</td>
<td>1994</td>
</tr>
<tr>
<td>Istochnobalkanska Svinia</td>
<td>Pig</td>
<td>1994</td>
</tr>
<tr>
<td>Askaniiskaya</td>
<td>Sheep</td>
<td>1980</td>
</tr>
</tbody>
</table>
The reporting of population data for the current year can be monitored by the NC also via the "Annual NC report" (Table 4.4). This report is supposed to be automatically sent each year to the NCs who enter data in one of the network databases. The report contains information about the number of breeds per species, the degree of completeness of the populations records also calculated per species and the trend in degree of completeness in comparison to previous year. All population data entered during the year is also included in this report.

**Table 4.4: Annual report for population data entered by Germany for year 1994 (excerpt). (Jahresbericht der Deutscher Populationsdaten für das Jahr 1994 (Auszug))**

<table>
<thead>
<tr>
<th>Species</th>
<th>Breeds count</th>
<th>Degree of fill</th>
<th>Trend from the previous year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle</td>
<td>59</td>
<td>55%</td>
<td>2%</td>
</tr>
<tr>
<td>Chicken</td>
<td>97</td>
<td>27%</td>
<td>27%</td>
</tr>
<tr>
<td>Deer</td>
<td>4</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Duck (domestic)</td>
<td>3</td>
<td>37%</td>
<td>37%</td>
</tr>
<tr>
<td>Goat</td>
<td>16</td>
<td>47%</td>
<td>-3%</td>
</tr>
<tr>
<td>Goose (domestic)</td>
<td>6</td>
<td>31%</td>
<td>31%</td>
</tr>
<tr>
<td>Horse</td>
<td>118</td>
<td>51%</td>
<td>11%</td>
</tr>
<tr>
<td>Pig</td>
<td>17</td>
<td>53%</td>
<td>-5%</td>
</tr>
<tr>
<td>Rabbit</td>
<td>75</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Sheep</td>
<td>62</td>
<td>46%</td>
<td>8%</td>
</tr>
<tr>
<td>Turkey</td>
<td>2</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Population statistics**

<table>
<thead>
<tr>
<th>Breed name</th>
<th>Species</th>
<th>Min</th>
<th>Max</th>
<th>Breeding females</th>
<th>Breeding males</th>
<th>Trend in population size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angler</td>
<td>Cattle</td>
<td>30153</td>
<td>30000</td>
<td>153</td>
<td>Stable</td>
<td></td>
</tr>
<tr>
<td>Angus</td>
<td>Cattle</td>
<td>13985</td>
<td>13500</td>
<td>485</td>
<td>Increasing</td>
<td></td>
</tr>
<tr>
<td>Ansbach-Triesdorfer</td>
<td>Cattle</td>
<td>30</td>
<td>27</td>
<td>3</td>
<td>Decreasing</td>
<td></td>
</tr>
</tbody>
</table>

90
The population size and structure statistics can be monitored from the Regional Coordinator via the "Countries with missing population records for a year" report. An excerpt of this report for Europe for year 1986 is shown in Figure 4.7.

<table>
<thead>
<tr>
<th>Country name</th>
<th>Breed name</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>Comune</td>
<td>Ass</td>
</tr>
<tr>
<td></td>
<td>Buffaloes</td>
<td>Buffalo</td>
</tr>
<tr>
<td></td>
<td>Black and White</td>
<td>Cattle</td>
</tr>
<tr>
<td></td>
<td>Blanc Bleu Beige</td>
<td>Cattle</td>
</tr>
</tbody>
</table>

**Figure 4.7: Breeds with missing population data for year 1997 from Europe (excerpt) (Rassen mit fehlenden Populationsdaten für Europa im Jahr 1997 (Auszug))**

**Translations and Data Synchronization Area:** The reports in this group are intended to support the administrators of the National and Regional database in organizing the translations of data and synchronization with other databases. The group includes four reports, which are listed below.

10. Data received via synchronization (pdf)
11. Status of data translation (pdf)
12. Number of breeds records updated in a year (Web)
13. Last update date per country (Web)

The "Data received via synchronization" report contains information about the number of breeds, images and publications received from other databases. Other reports like "Status of translation" inform the coordinators for the status of data, which are not directly entered by them, e.g. translations of breed description or publications in an official language. The last two Web reports - "Number of breeds records updated in a year" and "Last update date per country" are intended to be used by the RC to monitor the activity of the countries in updating their data.
Work Schedule

Readily available infrastructure is only useful for data improvement when embedded in a well-defined work schedule. Here we propose a workflow for the NC and RC in their management tasks.

**National Coordinator:** The first task of the National Coordinator is registering all National breeds in the database and completing their general description. The NC should print the "Degree of completeness of breed description" report to identify which breeds are already registered in the database and what is their description DC. Then she/he may send the breed description forms (available from the Web page) to the breeding association, the institutions or the farmers monitoring these breeds for completion.

The other important task is completing the population data and keeping it up-to-date. The population data in Europe should be entered on a yearly or Bi-yearly basis. Such interval is reasonable, taking into account the possible disease outbreaks and their consequences on the population size.

There are two approaches to this task - completing data per breed or completing data per year. In the "per year" approach the NC chooses the year for which you want to complete population data. Then she/he should print the "Population data entered for one year" report for this year and send the printed forms to the breeding association, the institution or farmers for completion. Data from the returned filled forms can be entered directly in the database without need to reformatting.

With the "per breed" approach the NC first chooses a breed to work with. Then she/he should identify gaps in reporting data for this breed from "Countries with missing population records for a year" report and "Reminder for entering population data" report. From the report "Degree of completeness of the populations records" the NC will identify incompleteness in existing population records. As a result the NC will have a list of years, for which she/he should enter data and also a list of years where data is incomplete.

**Regional coordinator:** Here, several tasks should be organized by the Regional Coordinator on a yearly basis. These tasks will require a one or two weeks working time per year. In the beginning of each year the RC should send invitation e-mails to
the countries to update their data. Each of these mails should contain the "Reminder for entering population data" report for the respective country.

At the end of the year the RC should send annual feedback emails to the countries for their improvement in data completeness. For this purpose data from the "Status of reporting by country" report should be stored each year and differences in data completeness between two consecutive years calculated. The proposed working schedule for the Regional coordinator is shown in Table 4.5.

**Table 4.5: Regional coordinator schedule proposition. (Vorschlag zum Arbeitsablauf für den Regionalen Koordinator)**

<table>
<thead>
<tr>
<th>Report</th>
<th>Timing</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report 9</td>
<td>January each year</td>
<td>Disseminate to all countries</td>
</tr>
<tr>
<td>Report 6</td>
<td>January each third year</td>
<td>Invite the countries to fill gaps in population data</td>
</tr>
<tr>
<td>Report 3</td>
<td>December each year</td>
<td>Summarize and send feedback to the countries</td>
</tr>
<tr>
<td>Report 13</td>
<td>December each year</td>
<td>Invite the countries to update their data</td>
</tr>
</tbody>
</table>

**Conclusions**

The management of the farm animal biodiversity data and its regular update can be very time-consuming task depending on the organization and number of breeds to be monitored. Translation of the data in several languages and distribution of the data across the network also adds to the complexity of the system and increases the informational requirements of the data providers and database managers - the National, Regional and Global Coordinators. The system of tools developed in the EFABIS project for meeting these requirements can be used from the National Coordinators all over the world to reduce the time to discover gaps in data and filling them. The tools also help in preventing new gaps during the regular data loading and give the National and Regional coordinators clear overview of the content, data updates and translation and synchronization activities in the network.
Involving RC in the monitoring and updating process of countries is in line with FAO’s intended function for RFP and should help improve data in DAD-IS.

Acknowledgements
This work is part of the EU funded project EFABIS, contract number QLRT-2002-00026
References


95
Chapter 5

General Discussion
Discussion

The increasing erosion of biological diversity was recognised by the animal breeders already in the previous century (Simon and Schulte-Coerne, 1979) and propositions for monitoring and conservation measures were made (Genetisch-statistische Ausschuss der Deutschen Gesellschaft für Züchtungskunde, 1979). In 1992 the importance of domestic animal diversity was also recognised with the Convention on Biological Diversity (United Nations, 2006). Since then, one of the main efforts on national and supranational level has been the creation of complete inventories of farm animal genetic resources, containing detail breeds characteristics and population statistics. Such information is needed for the estimation of the breed risk status of extinction and for prioritization of breed conservation. As it may not be economic possible to preserve all breeds, an accurate estimation of the degree of endangerment is of great importance.

The aim of the current work was to bring about a cohesion of criteria used in Europe [Chapter 3] on the basis of the data collected in the new biodiversity information network developed in the EFABIS project. The results from the application of these criteria on national and regional level can be used as an early warning system for loss of biodiversity. Furthermore, the information requirements for data management in the national and regional databases in the network were analysed [Chapter 4] and, because of data insufficiencies, appropriate tools for monitoring database content were developed to improve the data quality. The protocol for data exchange between databases developed in EFABIS [Chapter 2] ensures that the national data is propagated at regular intervals to the regional and international databases.

The synchronization protocol was developed in a general manner to allow data synchronization in all kind of farm animal information systems based on the APIIS (Groeneveld, 2004) platform. However, if the business rules in the various databases differ it is better to export data from the source database and load in the target database via the standard data stream channels (Groeneveld, 2004). A physical data transfer approach was used, because data in FABISnet are needed on international
level for evaluation, e.g. for creating the World Watch List of endangered breeds by FAO (Loftus and Scherf, 1993; FAO, 1995; Scherf, 2000). Moreover, the presence of data in several databases increases its availability and reduces the load on a single database. These arguments, together with the lack of national databases in many countries had spoken against the approach taken by GBIF for federating data from various collections and sources (Edwards et al., 2000). As the national databases are autonomous nodes in the network, which may not be connected all the time, the synchronization of data is based on the current version of the data elements in both nodes (Parker Jr et al., 1983; Almeida et al., 2002). The use of synchronization based on log (or journal) like in StorageBox (Hupfeld, 2004) was not suitable in EFABIS, because it may lead to accumulation of huge log files due to long latency of the nodes. In such case in the replicated systems the latent nodes are removed from the replication set (Ladin et al., 1992), but in EFABIS this is not acceptable. Synchronization by state has also its drawbacks, e.g. the transfer and comparison of the DE state information on each run.

The synchronization protocol was not used only in regular data updates. As an interesting byproduct, the synchronization procedure is also used to initialize newly created national databases. By simply defining appropriate sources and targets, the national data were downloaded in a one-time transfer from the global database.

As described in the introduction, several systems for estimation of the degree of endangerment were developed during the years, which are hard to compare due to their specific parameters and thresholds. The criteria used in this study [Chapter 3] for the estimation of DE have to be considered on two levels - national and regional each consisting of two components - demographic and genetic.

The expected female population size after two generation intervals is used as a demographic component on the national level. For uniformity, the thresholds are the same as in FAO criteria (Scherf, 2000): 100 females for the critical and 1000 for the endangered status. Analysis of data from Europe showed that the number of breeding females are present in almost 80% of the population records. This suggests that such data are available in Europe and using only the number of breeding females in the criteria will stimulate the countries to enter the precise numbers. The
use of expected number of breeding females has also other advantages over the
criteria currently used by FAO (Scherf, 2000), which are based on current number of
breeding females, number of breeding males or the total population size. One
advantage is that the responsible parties have a ‘reaction time’ of two generation
intervals, during which they can analyse the underlying factors and change the
negative trend. Another parameter, which can be used also as rapid indicator of
negative changes, is the number of young females registered annually (Alderson,
2003). This parameter is used in the criteria of the Rare Breed International and is
more accurate indicator of the viability of the population, but such data are not
present in EFABIS and there was no possibility to include it in the data collection
process. Gandini et al. (2004) proposed as a demographic indicator the time to reach
critical female population size. Such a proposition has the benefit of giving the exact
time left for reaction, but on the other hand it can be used only on a national level for
estimating the status of a single breed. As demonstrated in [Chapter 3] this indicator
is hard to apply on a regional level, when estimating the status of a ‘country
population’ of transboundary breed, or status of a breed as part of a group of
genetically similar breeds.

The assessment of the expected number of females is done by estimating the
recent rate of growth in female population size. This is done in simplified manner as
in (Gandini et al., 2004). Estimating the growth rate by regression techniques may be
more accurate than the procedure used by Benenwitz et al. (2005), however, data
insufficiencies prevented its use.

Another major difference between the criteria described in [Chapter 3] and the
FAO and EAAP criteria is that the former two use the last available data, whereas in
this study the requirements for data are more strict. Moreover, the results produced
by the new system are considered reliable only for two generation intervals, a
constraint not explicitly included in the other systems.

For keeping conformity with the current EAAP criteria, the expected cumulative
inbreeding in 50 years was used in the genetic component of the criteria [Chapter 3].
The thresholds are the same described in (Simon, 1999), and the methodology is
similar, but no additional factors for correction of the results were used. Another
difference from the criteria described by Simon is that the effective population size is estimated from the number of breeding males and females and then corrected assuming mass selection, as unselected populations are very rare in Europe.

The application of the criteria on the national level resulted in a large number of breeds with the status ‘unknown’, e.g. 82% in year 2005. This was a strong indication for a need of a special methodology and tools to assist the data managers in completing the data [Chapter 4].

To illustrate the application of the criterion on supranational level a grouping of 21 pig populations from the Hampshire, Duroc, Large White, Pietrain and Landrace breeds was done, based on the genetic distances calculated by SanCristobal et al. (2006). The results show that Duroc, Pietrain and Landrace breeds are not endangered. Data for some of the populations in these groups were missing, but these gaps had no impact on the results. This is an indication that the criterion can produce results on supranational level even in the case of incomplete data that none of the other procedures can.

The criterion proposed here uses a small number of parameters (number of breeding males and females), which have been also used in the FAO and EAAP criteria. Many more parameters can be added, e.g. presence of organization monitoring the breed, number and average size of herds, average age of the farmers, distance between the animals, etc. Although they have an impact on the DE of the breed, this impact is somehow arbitrary. The system used by Reist-Marti et al. (2003) for estimating breeds extinction probabilities is based on 10 parameters, including everything from total population size and organization of the farmers to political situation in the country and sociocultural importance of the breed. Unfortunately, many of the variables used by Reist-Marti et al. require arbitrary assignment of values by a human, which makes them not suitable for automated calculation.

When developing the criterion for transboundary breeds, an assumption was made that the various populations in the different countries can and probably do exchange genetic material. While this is very much the case in EU, it is not always true when other regions of the world are taken into account. Moreover, the question
of objective thresholds under which the breed is not sustainable in a transboundary context is still open and requires further research.

With the existence of the synchronization protocol, the gaps of data as outlined in [Chapter 3] have to be filled, where the data are initially entered - in the National databases. The data managers of these databases -i.e. the National Coordinators for management of genetic resources need for filling the gaps a quick overview of the actual status of their data. As the breeds characterization and population data are updated on different time basis, different approaches are required [Chapter 4]. In the case of the breed descriptions only the degree of completeness of data is relevant, whereas for the population statistics the regular time-series are also important.

To summarize these data a simplistic approach was chosen by calculating the degree of fill of the various blocks as the ratio of number of complete to the number of all possible fields. The results, however, should be treated with caution and knowledge of the breed and the content of the respective block as direct comparison of the numbers may be misleading. This was one of the main reasons, to make such information available only to the National Coordinators and not to the general public.

Such infrastructure is only useful for data improvement when embedded in a well-defined work schedule, e.g. for completing the population data two approaches: 'by breed' and 'by year' were proposed. This should cover the standard situation of data collection in the countries, or serve as guidelines in countries, where this process is not yet systematized. Such a schedule is also important for controlling the tasks like data translation or identifying the data received from synchronization with other databases, which are done by other people.

Overall, this study tried to support the parties involved in the management of animal genetic resources in taking decisions about the conservation and utilisation of these resources and in improving the data needed for such decisions. However, an effective result can be achieved through the cooperation of the governments and non-government organizations, breed societies, farmers and the general public.
References


Chapter 6

Summary
Summary

Zhivko Duchev

Management support and early warning system for national biodiversity databases in a network of national, regional (EAAP) and international (FAO) structures

Animal diversity has evolved over millions of years through the process of natural selection. In the last centuries the process of selection became more human driven, thus leading to acceleration of diversity development. The increasing effectiveness of the breeding techniques led to the use of restricted number of highly productive breeds, which raised concern about the loss of genetic diversity. Towards the end of the previous century the increasing erosion of biological diversity in livestock was recognized by the animal breeders as an area which needs monitoring and conservation measures.

The Farm Animal Biodiversity Network (FABISnet) is a new biodiversity network for collecting domestic animal breeds data from the European countries. Data are collected on National, Regional and Global level and can be automatically transferred between the levels. As a successor of the Animal Genetic Data Bank of the European Association for Animal Production (EAAP-AGDB) and the Domestic Animal Diversity Information System (DAD-IS) of the Food and Agricultural Organization (FAO), it has taken their historic data and integrated them in a network of databases.

With the new network the concept of data entry and management has changed, relying exclusively on the managers on national level to supply the upper levels with quality and complete data.

The aim of this work was to analyse the information requirements of the data managers and develop tools to support them in their work. The other part of the study was intended to bring a cohesion and uniformity in the criteria for estimating the degree of endangerment of breed heading for extinction. Such criteria are suggested to be used in Europe as an early warning system for loss of biodiversity in animal genetic resources.
The first part of this work was dedicated to the development of data synchronization protocol, which automatically keeps defined parts of two biodiversity databases in the network in synchronized state. An efficient synchronization procedure is the prerequisite to establishing National biodiversity websites while being integrated in the Global FABISnet. The protocol was already used to synchronize the content of the Polish National Biodiversity database and the regional European database. As an interesting byproduct the synchronization procedure has been also used to initialize newly created national databases.

In the second part of this study a proposition for a uniform criterion for Europe for estimating the DE of a breed is described. The criterion is based on the trend in number of breeding females and the expected cumulative inbreeding. It takes also into account the status of transboundary populations (factor missing in the criteria applied currently in Europe). In this study a novel approach to the problem of transboundary and international breeds was used, allowing the user to define herself groups of genetically similar breeds. The criteria were applied to the European data from regional database and to 5 groups of pig breeds. The results show that the incompleteness of the data is still one of the biggest problems on the regional and global level.

Therefore, the last part of the study was to develop a methodology and tools for filling gaps in data and improving the quality of data needed for research and monitoring. A set of 13 management support reports and a methodology for their application has been developed. The reports present summarized information about the degree of completeness of the breeds’ descriptions by country, missing or incomplete population records, reminders for updating data and the status of data translation. Various monitoring/reporting tasks of the National and Regional Coordinators for management of animal genetic resources can be organized in a systematic manner using these reports. Such an organizational scheme can reduce the time spent in completing data and improve the content of each database in the network.
Chapter 7

Zusammenfassung
Zusammenfassung

Zhivko Duchev

Entwicklung von Managementhilfen sowie eines Frühwarnsystems für nationale Biodiversitätsdatenbanken in einem Netzwerk auf nationaler, regionaler (EAAP) und internationaler (FAO) Ebene


Das Netzwerk zur Biodiversität bei Nutztiern (Farm Animal Biodiversity Network -- FABISnet) ist ein neuerartiger Ansatz zur Sammlung von Informationen über Nutztiersorten in europäischen Ländern. Die Daten werden auf nationaler, regionaler und globaler Ebene gesammelt und ausgetauscht.


Appendix
List of figures

Figure 1.1: The topology and data-flow of the FABISnet ........................................... 22
Figure 2.1: Excerpt of EFABIS topology, containing National Polish database,

European regional one and the world database of FAO. The dashed arrow

represents the rule for one source per data element ........................................ 40
Figure 2.2: Additional tables and meta-fields (shaded) needed for the synchronization

protocol in the APIIS structure ........................................................................ 44
Figure 2.3: Functional model of the data synchronization process for a single data

element ............................................................................................................... 46
Figure 2.4: Block diagram of the synchronization algorithm ................................... 46
Figure 2.5: Topology of the EFABIS system. On the lower row the national databases

are represented - BG for Bulgaria, DE - Germany, PL - Poland, VN - Viet

Nam. In the middle row are the regional databases, e.g. EAAP for the

Animal Genetic Data Bank of the European Association for Animal

Production. On the top is the biodiversity database of FAO ............................ 48
Figure 4.1: Degree of completeness of the breed description - Switzerland (excerpt)

(Grad der Vollständigkeit der Rassenbeschreibung für die Schweiz

(Auszug)) ....................................................................................................... 85
Figure 4.2: Presence of image data from Switzerland (excerpt) (Bilddaten aus der

Schweiz (Auszug)) ...................................................................................... 86
Figure 4.3: Degree of countries data completeness by region (Grad der

Vollständigkeit von Länderdaten nach Regionen) ........................................... 86
Figure 4.4: Degree of data completeness in region Europe (Grad der Vollständigkeit

in Europa) ....................................................................................................... 87
Figure 4.5: Missing population data for the German breeds in the last 10 years

(excerpt) (Fehlende Populationsdaten Deutscher Rassen der letzten 10

Jahre (Auszug)) ........................................................................................... 88
Figure 4.6: Degree of completeness of the population records for four Switzerland

breeds (Grad der Vollständigkeit von vier Schweizer Rassen) ...................... 88
Figure 4.7: Breeds with missing population data for year 1997 from Europe (excerpt)
(Rassen mit fehlenden Populationsdaten für Europa im Jahr 1997
(Auszug)).................................................................................................. 91
List of tables

Table 2.1: Time to receive merge data from the server, update the database and total synchronization time ................................................................. 50

Table 3.1: Data taken from EFABIS database in May 2006 (Verwendete Daten aus der EFABIS Datenbank im Mai 2006) .......................................................... 59

Table 3.2: Grouping of 21 pig populations from eight European countries. The groups are: DU- Duroc, HA- Hampshire, LR – Landrace, LW – Large White, PI – Pietrain. In the last column is the breed code from PigBiodiv Project (SANCRISTOBAL et al., 2006) (Gruppierung der 21 Schweinepopulationen aus 8 europäischen Ländern. Die Gruppen sind: DU- Duroc, HA- Hampshire, LR – Landrace, LW – Large White, PI – Pietrain. In der letzten Spalte wird der Rassencode des PigBioDiv Projekts (SANCRISTOBAL et al., 2006) angegeben) ........................................ 61

Table 3.3: Classes of endangerment based on the effective population size $N_e$. GI is the average generation interval for the species, NG50 is the number of generations for 50 years. (Klassen des Gefährdungsstatus auf der Basis der effektiven Populationsgröße $N_e$. GI entspricht dem durchschnittlichen Generationsintervall für die jeweilige Spezies, NG50 ist die Anzahl der Generationen innerhalb von 50 Jahren.) .................................................. 64

Table 3.4: Assessment of endangerment of the European breeds in year 2005. (Schätzung des Gefährdungstatus der europäische Rassen- Jahre 2005.) ................................................................................................................. 67

Table 3.5: Comparison between the various criteria used in Europe. FAO- criterion used by FAO as defined in (SCHERF, 2000); EAAP – a modified version of EAAP criterion as used in EFABIS. (Vergleich der verschiedenen Kriterien für den Gefährdungsstatus in Europa. Das FAO-Kriterium entsprach der Definition von Scherf (2000); EAAP – ein modifiziertes Kriterium, das in der EFABIS Datenbank verwendet wird.) ......................................................... 68

Table 3.6: Assessment of the DE on a supranational level for 21 pig breeds in year 2002. The classes of the demographic and genetic component are shown together with the number of breeds used to estimate the group class.
Appendix


Table 4.1: Degree of completeness of the European breeds description. (Grad der Vollständigkeit der Europäischen Rassenbeschreibungen) .................. 80

Table 4.2: Distribution of the population records of the European breeds. (Verteilung der Populationsdaten Europäischer Rassen) ........................................ 81

Table 4.3: Invitation to Bulgaria for updating the population data (excerpt). (Erinnerung an Bulgarien zum Aktualisieren von Populationsdaten (Auszug)) .................................................................................................................. 89

Table 4.4: Annual report for population data entered by Germany for year 1994 (excerpt). (Jahresbericht der Deutscher Populationsdaten für das Jahr 1994 (Auszug)) ............................................................. 90

Table 4.5: Regional coordinator schedule proposition. (Vorschlag zum Arbeitsablauf für den Regionalen Koordinator) ................................................................. 93
**Abbreviations**

AnGR  Animal Genetic Resources  
APIIS  Adaptable Platform Independent Information System  
BLUP  Best Linear Unbiased Prediction  
CBD  Convention on Biological Diversity  
CPU  Central Processing Unit  
DAD-IS  Domestic Animal Diversity Information System  
DE  Degree of Endangerment  
DED  Data Element Description  
DC  Degree of Completeness  
EAAP  European Association for Animal Production  
EAAP-AGDB  European Association for Animal Production – Animal Genetic Data Bank  
EFABIS  European Farm Animal Biodiversity Information System  
EU  European Union  
FAO  Food and Agriculture Organization of the United Nations  
Kb  Kilobit  
LAN  Local Area Network  
Mb  Megabit  
MB  Megabyte  
NC  Country Official National Co-ordinator  
Ne  Effective Population Size  
PISSA  Pig Information System South Africa  
RAM  Random Access Memory  
RBST  Rare Breeds Survival Trust  
RDBMS  Relational Database Management System  
RC  Regional Co-ordinator  
sec  second  
SQL  Structured Query Language  
UN  United Nations  
WAN  Wide Area Network
Acknowledgements

I would like to thank my supervisor Dr. Eildert Groeneveld for the initiation of this study, for his continuous guidance, confidence and patience. I am especially grateful for the fact, that he always finds time to share ideas or give an advice.

I would like to thank Prof. Dr. Ottmar Distl for giving me the opportunity to be part of the EFABIS project and for supporting me in my study.

Prof. Dr. Lothar Kreienbrock I would like to thank for his interest in my work and for the valuable suggestions.

Furthermore, I thank the EAAP Working Group on Animal Genetic Resources and especially its Chairman for the interesting discussions and ideas.

Special thanks to all people from the Department of Genetic Resources and Breeding, Institute for Animal Breeding in Mariensee, who maintained a nice working atmosphere and helped me with all kind of problems.

Last but not least, special thanks to my relatives and friends for their support during the years.