

# Design and Implementation of an Information System for National Genebanks Management

Truong Van Chi Cong, Zhivko Ducheve and Eildert Groeneveld

(Session 08 - Abstract number: 3211- Email: cong.chi@fli.bund.de)

Department of Breeding and Genetic Resources  
Institute of Farm Animal Genetics, Mariensee, FLI  
Höltzstr 10, Neustadt 31535, Germany

## Abstract

A long-term storage of genetic materials is essential in each country. In order to manage a large number of samples from different species in animal agriculture, a general data model must be considered and implemented. This paper presents an approach for the development of a sample management information system at the country level. Designed as client/server architecture, our information system is a platform independent web application based on open source software packages. As a result, the developed information system provides for information management of national genebanks. It allows users to record the origin of genetic material, keep track of samples along with the storage places, and make analyses reports. The software is freely available under the Open Source GNU public license.

## 1 Introduction

With the international acceptance of the Rio Convention [1] in 1992 conservation of biodiversity has become a major issue not only in wild life but also in animal agriculture. While conservation in wild life usually focuses on the species, the breeds are the focus of interest in animal agriculture. Already in 1979 scientists stipulated conservation of breeds of farm animals [2].

More recently a proposal has been made for a world wide emergency programme for the creation of national genebanks of endangered breeds in animal agriculture [3] based on the low cost collection and storage of somatic cells. The technique has been tested under field conditions on 6 breeds from 3 species [4]. The procedure is intended to be deployed rapidly to countries with little infrastructure and to operate identically on all mammalian species. An integral part of a physical genebank is the sufficient documentation of the stored samples, which has to be equally easy and rapid to install and operate and which only requires limited or no adaptation prior to country installation. Based on the implementation for the National Genebank in the Netherlands [5], the structure of the new system was further simplified to allow general applicability without modification.

## 2 Requirements

As pointed out above, the information system is considered to be an important component in the concept for a world wide emergency program for the setup of genebanks. Implementation of such program would entail setting up a relatively large number of national and possibly sub-national genebanks. Clearly, a standardized product, that would not require software expertize for adaptation would be of great benefit.

Often multiple locations are used in genebanks. Firstly, there are safety issues that make it advisable to store material in more than one location. Furthermore, a decentralized approach in terms of material and data collection may seem advisable. Clearly, information on all genebank material, be it centralized or not, needs to be available together.

The information available on animals represented in cryo store will be very different ranging from complete pedigrees and performance records for some populations in intensive breeding programs to close to none on animals sampled in remote low intensity environments. The system must be able to handle both cases without modification.

The basic objective is to provide a register for the material stored in the genebank. The main requirements are:

1. The identification on the vessel must be unique in the genebank and lead directly to the information about it in the database. This means, that any sample taken out of the genebank must lead solely through its vessel identification to information about this sample.
2. The information in the database about a sample in the genebank must lead directly to the actual sample in the physical store. This means, that a record about a sample in the database must through its location information lead directly to the correct vessel in the genebank store.
3. The application must have the capability to be nationalized.
4. The database should be accessible through the Internet, allowing centralized and non centralized data entry.
5. The system should be usable for all possible genetic materials (semen, oocytes, embryos, somatic cells, DNA, blood) and all possible types of used vessels (straw, tube, cane, vial, ampule) in a uniform manner.
6. For hygienic reasons, the history of all material movements in the physical store is recorded.
7. The database should store a minimum amount of information for the materials, that is expected to be needed and available for the next decade.
8. The information stored should be uniformly available across all species and populations.
9. From its content and operations the information system has to be able to be run for an extended time span as is appropriate for long term genebanks.

### 3 Implementation

We aim to use open source components and software as much as possible. APIIS, which is a framework for adaptable platform independent information systems [6], meets the above requirements. Consequently, our information system was developed within this framework. So, both the database structure and the application architecture are based on APIIS.

#### 3.1 Database Design

The data structure of the database is composed of the components of the data to be recorded which are derived from the above stipulated functionality. The reasoning for choosing them is given in [4]. It is based on the objective to have generally available information that is not specific to any species or breed and that can be the basis for later ex post data collection. On the basis of scaled colour images a number of measurements can be taken on the animal. Here, only the blocks of data are presented. These are:

- administrative data: species, breeds, codes, addresses, protocols, locations
- animal information: animal ID, sire, dam, sex, species, breed, birth date, GPS location, scaled photo, organization, and comment
- sample information: animal ID, material type, production protocol, production date, freezing date, vessel ID, number of replicates (e.g. number of straws for semen), storage location, comment

The database structure of our information system is shown in Figure 1. The administrative data and animal information are kept in the core tables of APIIS [6] (codes, unit, naming, address, animal, transfer, blobs) and one additional table (protocols). The central part of the information system, the sample management, is handled by the three tables “samples”, “vessels”, and “stocks”. A five level hierarchy (location|tank|canister|level|cell) is used to manage the storage places of samples in a genebank. This structure is covered by the stock table. To track the history of all material movements in the physical store, two extra tables (stock\_history, movements) are added.

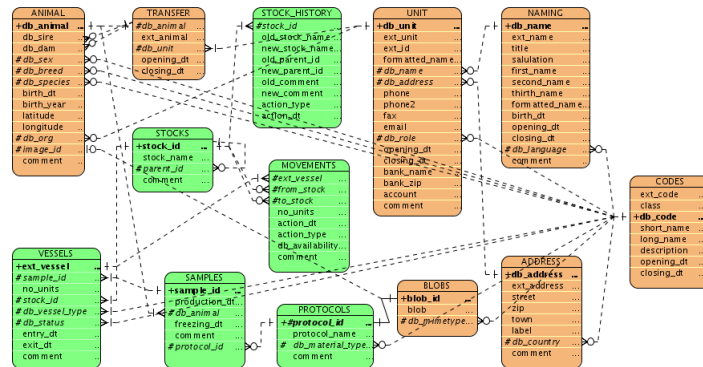


Figure 1: Database Structure for National Genebank Management

## 3.2 Application Architecture

The information system is based on the three-tier Web application architecture as shown in Figure 2. In this architecture, users (client tier) interact with the system via a web browser. In fact, the web browser is the only medium to access the database, all outputs have to be initiated through it. There is interactive output as a result of queries, as well as list outputs in pdf format. Web pages are generated in accordance with W3C in order to work with many different web browsers (e.g. Firefox, Internet Explorer, Opera, Konqueror).

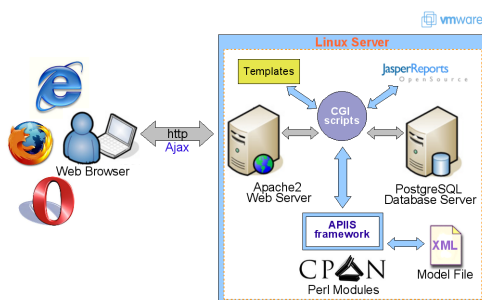


Figure 2: Web-based application architecture

The middle tier is coded in Perl and is driven by APIIS running under Apache2 web server. The application code (CGI scripts) is developed from open source packages available free from CPAN. The user interface (Figure 3) is rendered as Hypertext Markup Language (HTML) using templates. Jasper-Reports [7] which an open source reporting library is integrated into APIIS to compile and generate analysis reports automatically from given templates designed by iReport.

The backend tier is an SQL92 standard Relational Database Management System (DBMS). Currently we use PostgreSQL as a default backend. The database is built from the data model file, a XML file, created from the APIIS tool [8]. The system can be used in a multi-user environment with the user privileges being managed by the access rights management in APIIS.

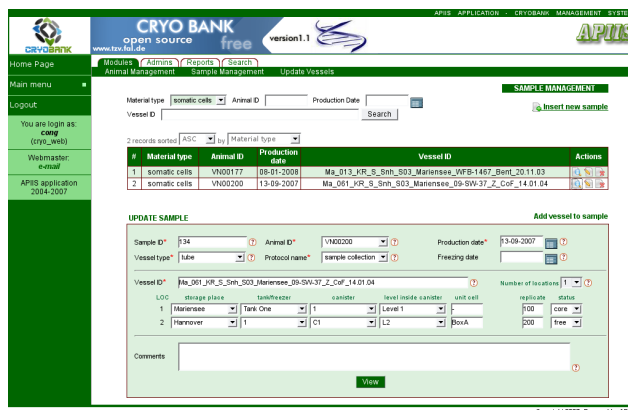


Figure 3: A web interface for sample management

## 4 Conclusions

In this paper we have presented basic requirements for national genebank information system management. Based on analyzed data model, we have successfully implemented a Web-based information system using the APIIS framework and other open source packages. It is suitable for long term management of samples in a national genebank. The deployment of our software system might be done via the virtualization technology as an appliance to reduce hardware costs and time for configuration, monitoring and maintenance. Under Open Source GNU public license, our system can be used out of the box for institutions or organizations requiring such a system.

## References

- [1] U. Nations, “United Nations Environment Programme Convention on Biological Diversity,” <http://dad.fao.org/en/refer/library/conventn/cbd.pdf>, 1992.
- [2] Genetisch-statistische Ausschuss der Deutschen Gesellschaft für Züchtungskunde, “Stellungnahme zur Bildung von Genreserven in der Tierzüchtung,” *Züchtungskunde*, vol. 51, no. 5, pp. 329–331, 1979.
- [3] E. Groeneveld, “A world wide emergency programme for the creation of national genebanks of endangered breeds in animal agriculture,” *Animal Genetic Resources Information*, vol. 36, pp. 1–6, 2005, ISSN1014-2339.
- [4] E. Groeneveld, N. H. Tinh, W. Kues, and N. T. Vien, “A protocol for the cryoconservation of breeds by low-cost emergency cell banks - a pilot study,” *Animal*, no. 2:1, pp. 1–8, 2008.
- [5] E. Groeneveld, L. Yordanova, and S. J. Hiemstra, “Organizational structure and information technological support of national gene banks,” *Livestock Production Science*, vol. 89, no. 2-3, pp. 297–304, 2004.
- [6] E. Groeneveld, “An adaptable platform independent information system in animal agriculture: Framework and generic database structure,” *Livestock Production Science*, vol. 87, pp. 1–12, 2004.
- [7] T. Danciu, L. Chirita, and G. Toffoli, *The Definitive Guide to JasperReports*. Apress L. P., 2007.
- [8] Z. Ducheve, E. Groeneveld, and R. Fischer, “Rapid implementation of integrated databases and inclusion of historic data in modern animal agriculture - an Open Source Solution,” in *Proceedings of the 6th Biennial Conference of the European Federation of IT in Agriculture and the World Congress on Computers in Agriculture, EFITA/WCCA*, C. Parker, S. Skerratt, C. Park, and J. Shields, Eds. Glasgow Caledonian University, Glasgow, Scotland, 2-5 July 2007, p. 6 pp.