



XII CONGRESSO INTERNACIONAL DO LEITE

XII Workshop de Políticas Públicas
XIII Simpósio de Sustentabilidade da Atividade Leiteira

Sistema de armazenamento para a organização e o controle do banco de DNA de bovinos de leite

Katia Cristina Lage dos Santos¹, Wagner Antonio Arbex², Marcos Vinícius Gualberto B. da Silva³,
Marta Fonseca Martins³, Isabela Fonseca⁴, Daniel Jordan de Abreu dos Santos⁵

¹ Analista, Embrapa Gado de Leite, Juiz de Fora/MG, katia.santos@embrapa.br

² Analista, Embrapa Gado de Leite, Juiz de Fora/MG

³ Pesquisador, Embrapa Gado de Leite, Juiz de Fora/MG

⁴ Doutoranda, Universidade Federal de Viçosa, Juiz de Fora/MG

⁵ Doutorando, Faculdade de Ciências Agrárias e Veterinária/UNESP, Jaboticabal/SP

Resumo: A crescente necessidade de alta disponibilidade e acesso *online* de alto volume de dados nos diversos campos do agronegócio tem demandado uma organização eficiente desses repositórios digitais. Seguindo essa tendência, o presente trabalho objetivou apresentar um sistema de banco de dados relacional confiável e robusto para controlar a tramitação de amostras genotipadas que compõem o banco de DNA de bovinos de leite organizado pela Embrapa Gado de Leite. A representação computacional pode ainda ser estendida a outras espécies, fazendo a alteração de informações específicas da natureza em estudo.

Palavras-chave: confiabilidade, disponibilidade, modelo de dados relacional

Storage system to organize and control the dairy cattle DNA datasets

Abstract: The growing need for high availability and online access to a high volume of data in many areas of agribusiness has demanded an efficient organization of these digital repositories. Following this trend, this paper intended to present a reliable and robust relational database system to control the processing of genotyped samples that make up the DNA database dairy cattle organized by Embrapa Dairy Cattle. The computational representation can also be extended to other species, making specific changes related to the nature of the study.

Keywords: reliability, availability, relational data model

Introdução

O crescimento exponencial das informações eletrônicas relacionadas ao agronegócio desperta a necessidade de se desenvolver sistemas computacionais que gerenciam o acervo de dados de maneira confiável e com alta disponibilidade. Informação eletrônica é todo dado em formato digital sistematizado que pode ser reconhecido e entendido por um usuário, por meio de um programa computacional (SOUZA et al, 2000). Além do armazenamento em formato digital, deve-se disponibilizar o acesso direto e rápido à informação para sua recuperação, visando gerar, formar e absorver novos conhecimentos.

Esse cenário pode ser identificado no controle e tramitação de dados de material para genotipagem de DNA presentes nas amostras de tecido de bovinos de leite preparadas pela Embrapa Gado de Leite. A genotipagem é a identificação de um conjunto de marcadores moleculares do indivíduo, observados a partir de seu DNA. A tramitação inicia-se com o registro e o armazenamento das amostras de material genético, que são tecidos (sangue, sêmen, etc.) enviadas pelas associações de criadores de diversas raças ou por outra entidade que auxilie na coleta das mesmas. Em seguida, são realizados os procedimentos de extração do DNA das amostras e acondicionamento do material genético. Por fim, é realizada a



XII CONGRESSO INTERNACIONAL DO LEITE

XII Workshop de Políticas Públicas
XIII Simpósio de Sustentabilidade da Atividade Leiteira

genotipagem de parte do DNA extraído. Este serviço é realizado pela Embrapa Gado de Leite e por empresas terceirizadas. Cada uma dessas etapas gera uma quantidade bem expressiva de dados que são continuamente consultados e utilizados nas ações de pesquisas e como métricas de produtividade.

Visando automatizar o processo de gerenciamento do mecanismo acima descrito, este trabalho descreve uma estrutura computacional confiável, rápida e com alta disponibilidade para armazenamento e recuperação dos dados das amostras genotipadas. Para isso, foi definido um modelo de banco de dados extensível às várias raças de bovinos, permitindo assim o registro e controle de toda a tramitação dos procedimentos para extração do DNA e genotipagem das referidas amostras.

Material e Métodos

Uma propriedade fundamental de sistemas de bancos de dados é que permitem a abstração dos dados, ocultando detalhes do armazenamento de dados, que são desnecessários para a maioria dos usuários desse tipo de aplicação. O mecanismo que é utilizado para descrever a estrutura e as operações em um banco de dados é denominado modelo de dados (ELMASRI et al., 2005). Ele busca sistematizar o entendimento dos objetos e eventos do mundo real que serão representados em um sistema informatizado.

Com base na natureza das entidades usadas para descrever a estrutura do banco de dados de DNA, foi escolhido o modelo de dados denominado relacional. Segundo (ELMASRI et al., 2005), nesse modelo, a principal construção para representação dos dados é a relação, ou seja, uma tabela com linhas não ordenadas e colunas. Uma relação consiste de um esquema e de uma instância. O esquema especifica o nome da relação e o nome e o domínio de cada coluna, também denominada atributo ou campo da relação. O domínio do atributo é referenciado no esquema por seu nome e serve para restringir os valores que este atributo pode assumir (valores numéricos, data, tempo etc). O esquema de uma relação é invariável ao longo do tempo, comportando-se assim como uma forma.

A Figura 1 é uma representação gráfica da relação Usuários. Nessa relação, há quatro atributos: identificador, CPF, nome e função. O domínio dos atributos são representados pelas cores amarelo e cor de rosa. Os atributos em amarelo são valores numéricos, já os em cor de rosa são sequências de caracteres. O atributo “identificador” é a chave primária da tabela, valor esse que é utilizado para representar um usuário de forma única. A chave primária é utilizada para correlacionar as relações existentes no modelo de de dados, não apresentando necessariamente uma representação no mundo real.

USUÁRIOS
identificador
CPF
nome
funcao

Figura 1: Representação gráfica da relação Usuário

Um sistema de banco de dados relacional é um conjunto de uma ou mais relações. O esquema do banco de dados relacional é a coleção dos esquemas de cada relação que compõe o banco de dados (SAYAO, 2001).

Resultados e Discussão

O registro e controle das informações de extração do DNA e genotipagem das amostras de bovinos de leite envolvem conceitos ou entidades variados. O primeiro passo para se definir um modelo de dados relacional consiste no levantamento das entidades a serem manipuladas. A partir de então, para cada entidade deve-se relacionar as suas características, bem como o tipo de dado que a representa (um valor numérico inteiro ou decimal, uma data ou um conjunto de caracteres). Para finalizar, é preciso relacionar as entidades, com o intuito de viabilizar a recuperação dos dados.

Seguindo esses passos, para organização e controle do banco de DNA de bovinos de leite, foi definido o modelo de dados apresentado na Figura 2, onde os atributos em negrito representam as chaves primárias das tabelas, os atributos de natureza numérica são representados em amarelo, em cor de rosa são conjuntos de caracteres e em verde são do tipo data. Os relacionamentos entre as tabelas são representados por linhas contínuas, indicando os termos envolvidos.

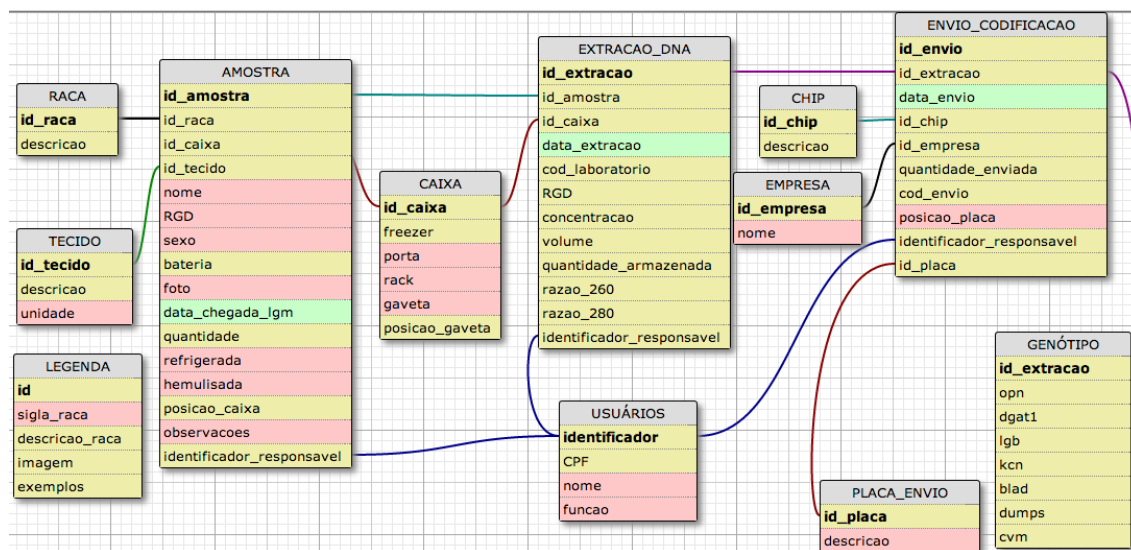


Figura 2: Modelo de dados relacional para organização e controle do banco de DNA de bovinos de leite

Conclusões

Como trabalho futuro, será implementada uma interface gráfica web para comunicação com o sistema de banco de dados.

Literatura citada

- ELMASRI, Ramez, NAVATHE, Shamkant. Sistemas de Banco de Dados. 4.ed. Rio de Janeiro: Addison Wesley, 2005.
- SAYAO, L. F.. Modelos teóricos em ciência da informação - abstração e método científico. Ci. Inf., Brasília, v. 30, n. 1, Abril, 2001. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100010&lng=en&nrm=iso>. Acesso em 03 de outubro de 2013. <http://dx.doi.org/10.1590/S0100-19652001000100010>.
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. Ciência da Informação, v. 29, n. 1, p. 93-102, jan./abr. 2000