

# Flexible data processing system for field survey on domestic animal diversity

Avnish K Bhatia , Dinesh K Yadav and Pardeep K Vij

National Bureau of Animal Genetic Resources, Karnal, Haryana - 132001 India

## ABSTRACT

Conservation of domestic animal diversity requires identification and current status of breeds through field surveys in their native tracts followed by data analysis and development of information systems for animal genetic resources. Various animal species possess distinct recording parameters requiring dynamic questionnaire formats. This warrants carefully designed database for development of a flexible computer package for management and analysis of the survey data. This paper describes such a generalized and flexible data processing system.

**Key words:** Domestic animal diversity, animal genetic resources, field survey, conservation, data processing system.

## INTRODUCTION

Domestic animal diversity is a vital resource for the present and the future well being of humanity and their conservation is essential for sustainable management of agricultural systems. Knowledge about breeds and their status in native tracts is required for their conservation. Livestock population censuses are conducted to estimate the number of animals by species. These censuses don't record breed of animals. Physical characteristics of animals of many breeds are unknown under field conditions. It becomes necessary to undertake breed surveys to determine the status of breeds in a country and then use the information to develop improvement and conservation strategies (Mpfu, Ojango and Anderson-Eklund 2006).

Field surveys are taken up to obtain basic information and population dynamics of breeds existing in a region / country. It includes designing of questionnaires to collect data pertaining to geographical and climatic features of breeding tracts, management practices followed by farmers, and morphological, production and reproduction characteristics of animals. Survey data is analyzed to get information on native tracts, socio-economic status of animal rearing communities, management practices of livestock in terms of housing, feeding, animal trade, migration, etc. Breed descriptor and other information emanating from the analyzed survey data becomes a source of data for information systems on animal genetic resources (Ayalew et al. 2003, FAO 2005).

Computerization of survey questionnaires is an important step in management and analysis of survey data. The following constraints make the task of computerizing survey data difficult.

- Inter-species variation in morphological, production and reproduction traits that makes it compulsory to design separate set of questionnaires for each species such as cattle, buffalo, sheep, goat, poultry, etc.
- Researchers undertaking surveys prepare different questionnaires. Although attempts have been made to prepare uniform species-wise questionnaires at global (FAO) and regional levels, researchers add data elements according to requirements of survey. Accordingly, the uniform survey questionnaires designed at global and regional levels may change with time and geography of survey.
- Field surveys on breeds are taken up on small scale (pilot survey) to assess first-hand status of breeds or it can be a complete survey on breeds (Nivsarkar and Gupta 1999). Such surveys are conducted with different set of species-wise questionnaires for breeds of livestock and poultry species.

These situations necessitate requirement of a data processing system that is flexible enough to accept data from different type of questionnaires. This paper describes such a flexible data processing system for animal genetic resources (DPSAnGR) for data storage and analysis.

In the remaining paper, we analyze requirements of the data processing system on animal genetic resources; explain system data modeling, and construction of the system. The data processing system is described from user's point of view. Future prospects of the system are also explored.

## MATERIALS AND METHODS

*Requirement Analysis:* A field survey on animal genetic resources is conducted in the native tract of a breed. Field data is recorded at three levels viz. at the level of geographical area such as native tract, at farmers' level and at animal's level (Wolf 1990). At the level of geographical area, parameters such as longitude, latitude,

1. Senior Scientist, Email: avnish@lycos.com
2. Senior Scientist, Email: dkyadav66@gmail.com
3. Principal Scientist, Email: pkvij@yahoo.com

altitude, average temperatures, average humidity, etc. are recorded once. At the farmers' level, management practices such as housing, migration, feed and fodders, watering, etc. are recorded for each farmer. Data on traits of animals is recorded for each animal with the farmers under survey.

There is general practice of formulating questionnaire formats into hierarchical way that makes data handling easy. Survey parameters are classified into different headings and sub headings like -General Information, Physical Characteristics of Animals, Production Parameters, etc. Data elements in questionnaires are either numeric or textual. A data element may be entered only once for a farmer or for an animal (e.g. horn-length) or repeatedly (e.g. weekly milk-recording).

Accordingly, data is analyzed at the three levels: for a district, on the basis of farmers and on the basis of animals of the breed under survey. Data is analyzed in terms of averages in case of quantitative data and frequencies in case of qualitative data. Lists also form output of analysis for the text data. Analysis of animal data is performed for different classes of animals such as young-ones, adult females, adult males, draft animals, etc. Analysis of data is also performed for a district, category of farmers such as low-income and high-income groups.

*Data Modeling* : Entity-relationship approach has been utilized for conceptual data modeling (Chen 1976, Lescourret et al. 1993) for the database in DPSAnGR. The approach considers real-world objects as entities having certain attributes. Relationships are connections between two or more entities, which may be defined by

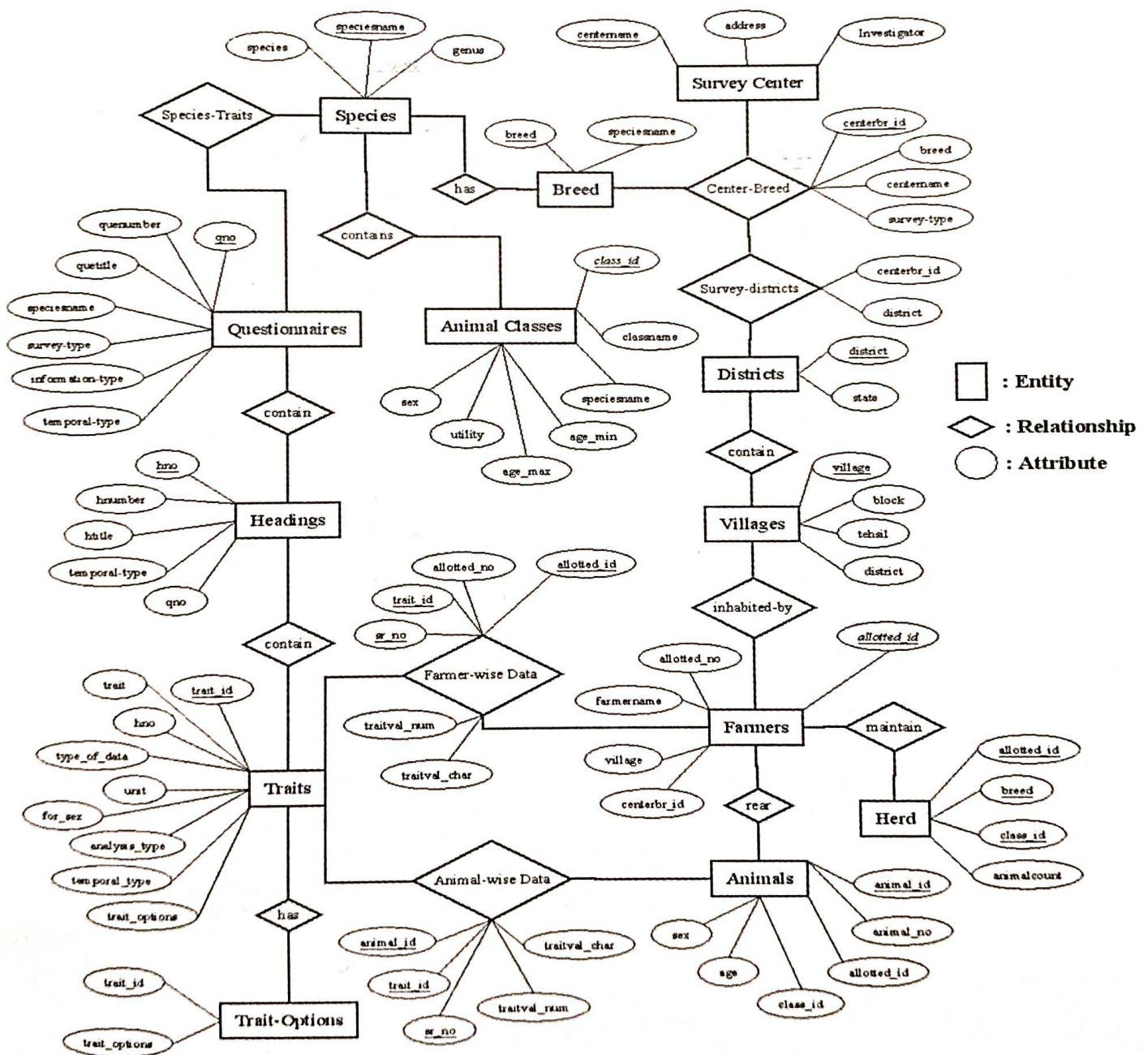


Figure 1: Entity-Relationship diagram showing conceptual model of underlying database in DPSAnGR

a set of attributes. Conceptual data modeling consists of identifying entities and relationships in the database theme and allocating attributes (basic data elements) to entities and relationships. Figure-1 shows the entity-relationship diagram for the underlying database, which has been translated into database structure. Entities, relationships and their attributes are displayed in the diagram.

Database tables have been defined using relational data model. Entities and relationships with attributes are defined as tables. First set of tables store geographical information such as "Districts" and "Villages" selected for survey. Animal species and breeds are stored in separate tables named "Species" and "Breed", respectively. A breed-survey-center is defined as a relationship among survey-center, breed and survey-districts, which has been included as table "Center-Breed" with data fields shown in figure-1.

Information on data-elements in questionnaires has been broken up into three tables: "Questionnaires", "Headings" and "Traits" to store questionnaires, headings in questionnaires and traits under a heading, respectively.

Some of qualitative traits may contain options to choose from during data entry. For example, the trait 'Body Color' has the options: black, white, etc. A table called "Trait-Options" has been defined to store these options. Trait options are used as a list in user-interface of the data processing system.

Information on farmers has been stored in a table "Farmers". Animals with a farmer are stored in the table named "Animals". Actual survey data is stored in two tables. The first table named "Farmer-wise-Data" is a relationship between farmers and traits, and stores data on farmers. The other table named "Animal-wise-Data" is a relationship between animals and traits, and stores field data on animals. The field 'srno' in these tables facilitates entry of repeated data for a trait of farmers and animals.

Animal classes for a species are defined on the basis of age, sex and utility. These classes are utilized for data entry on classes of animals such as calves, cows, bulls and bullocks (in cattle species) for animals. Data on herd / flock are also based on animal classes. A table named "Animal-Classes" has been defined for the purpose.

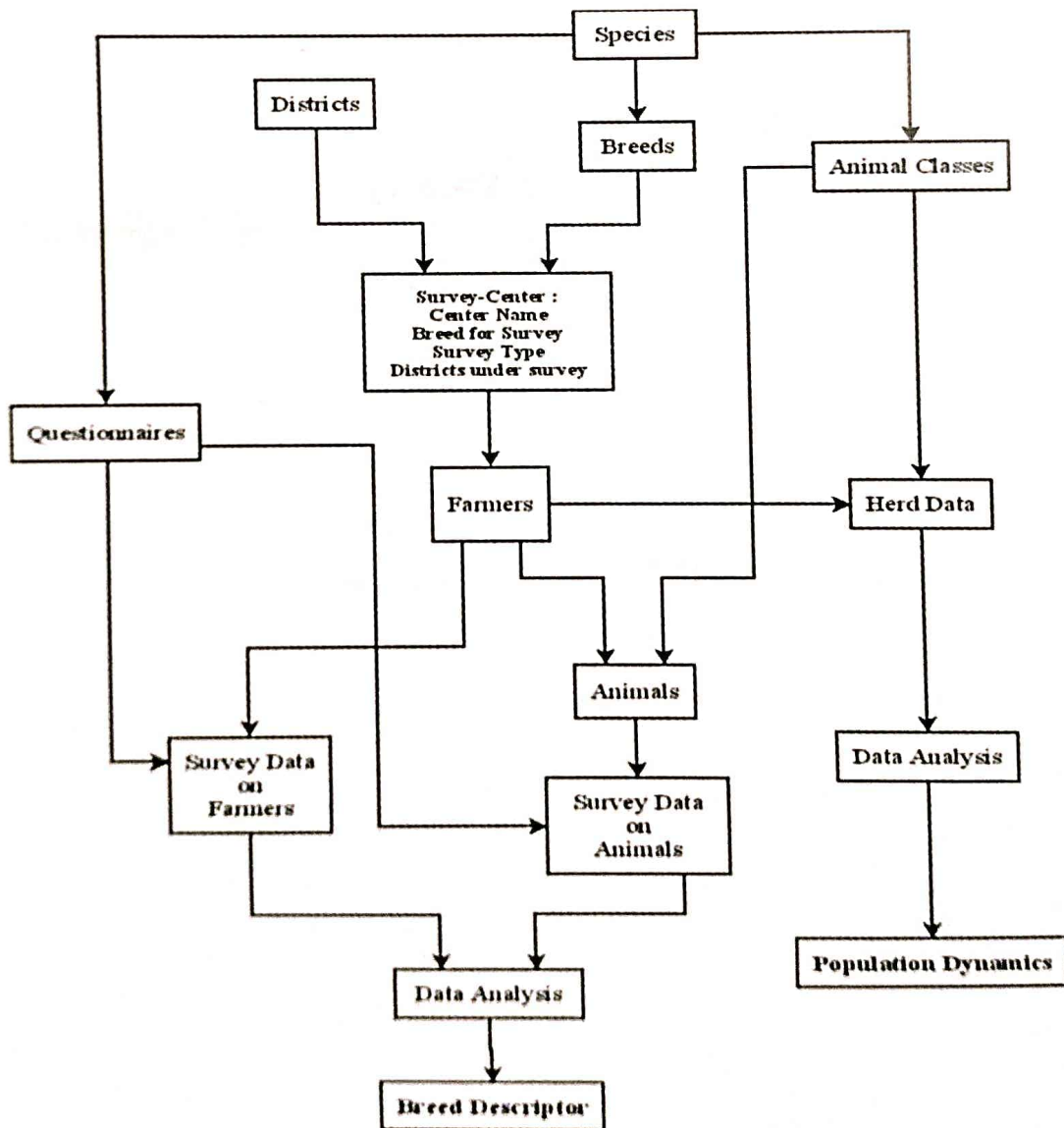


Figure 2: Flowchart showing sequence of operations while using DPSAnGR for survey data management and analysis.

One table named "Herd" is defined to store herd / flock of animals being maintained by farmers. This data is used in estimation of population status of the breed under survey.

**Data Processing System Development:** Data processing system has been developed using MS-Access 2000. Database tables for the database and forms for user interface are defined. The system contains a menu of options that leads to various forms for data entry and data analysis. It contains context-sensitive help facilities developed using MS-HTML Help Workshop. User manual has been provided as a MS-Word file along with the package.

## RESULTS AND DISCUSSION

**System Description:** Careful database design and development of user-interface has resulted into a flexible data processing system for animal genetic resources. The system doesn't have any limitation on number of breeds

under survey at a center, number of questionnaires and traits, and field survey data on a breed. It is a user-friendly package for management and analysis of field survey data on animal genetic resources. Figure-2 shows the sequence of data entry and analysis operations while using the Data Processing System.

DPSAnGR works by storing data for various entities in below-listed steps.

1. Name of species and the breed is entered into the database.
2. User is required to enter survey-districts in the breeding tract of the breed under survey.
3. Information about the survey center (name, investigator, address) is entered.
4. A survey center for a breed is defined as combination of center name, breed under survey, type of survey (pilot or complete) and districts in the breeding tract that are selected for taking up survey work.

Figure 3: Form for entry of field survey data on breeds.

Figure 4: Form for entering options for analysis of survey data.

5. Questionnaire format for the species is entered. It involves entry of the questionnaires, trait-headings in a questionnaire and traits under a heading. If the questionnaire format is already entered, users can check it for missing traits. Separate questionnaires are defined for pilot and complete (sample) types of survey.
6. Species-wise animal classes are entered, which facilitate entry of data for animals and herd / flock of animals being maintained by farmers.
7. Data on farmers (allotted no., name, village, community, etc.) rearing the animal of the breed is entered next.
8. Data on animals (animal no., class, age, sex) being maintained by farmers is entered.
9. Survey data on farmers is entered by selecting a farmer, questionnaire, heading, and trait from the lists displayed in the form for data-entry.
10. Survey data on animals is entered by selecting a farmer, an animal, questionnaire, heading, and trait.
11. Information on herd / flock maintained by farmers is entered. It includes data on animals of other breeds and species with farmers.

Although users are required to enter data in a systematic way, data elements on questionnaires, farmers, animal, survey data can be added at any stage by closing the current form and opening the respective form for data entry. Figure-4 shows a form for entry of survey data after entering questionnaires, farmers, animals and other information on various entities in the database.

Data analysis can be performed in the following ways.

1. Data can be analyzed for data elements under a heading or a questionnaire.
2. Analysis can be performed for separate districts or as pooled.
3. Analysis of data on animals can be performed on the basis of animal classes.
4. Strata can be defined on numeric data-elements of farmers (income, land holding size) for conducting stratified analysis.
5. Herd data analysis can be performed to get number of animals of various classes for different breeds and species being maintained by farmers in the breeding tract.
6. Data can be analyzed by importing and saving it to a text file for use in standard software like MS-Excel, SPSS, etc.

Figure-4 shows a form for entering the options of data analysis. Results of data analysis are displayed in the form of reports, which can be printed or stored as files.

*Future Prospects:* The data processing system is

presently available on Compact Disc and can be installed on personal computer with MS-Windows XP and MS-Access 2000 or above. In future, it can be re-developed using web tools to make it available on the Internet through a website (Guralnick and Neufeld 2005). It will enable the users to enter and analyze field survey data on animal genetic resources through the Internet. With little changes in data files and terminologies, the data processing system can also be developed for field surveys on fish and wildlife genetic resources (FAO 1992, NBFGR 2007).

## CONCLUSION

Conceptual data modeling constitutes a basic step for design of database files. Such modeling is relevant for the general design of data processing and information systems. This paper presents a data processing system for animal genetic resources (DPSAnGR) using data modeling on the basis of entity-relationship approach for use by animal geneticist, breeders and developmental agencies. It is of general and flexible applicability for management of field survey data of any domestic animal species. Data can be analyzed in different ways such as on the basis of districts, animal classes, and user-defined strata. The system can help the user to formulate characterization, improvement and conservation strategies for domestic animal diversity. With little changes, the data processing system can also be used for field surveys on other biodiversity resources such as fishes and wildlife.

## ACKNOWLEDGEMENT

Availability of free software Dia 0.96.1 for drawing entity-relationship, flowcharts and other diagram is duly acknowledged.

## REFERENCES

- Ayalew W, Edward J, Rege O, Getahun E, Tibbo M and Mamo Y. 2003. Delivering systematic information on indigenous animal genetic resources – the development and prospects of DAGRIS. In: Proc. Conference on Tropical and Subtropical Agricultural and Natural Resource Management (TROPENTAG) October 8-10, 2003, University of Gottingen, Deutschar.
- Chen P P. 1976. The entity-relationship model – toward a unified view of data. *ACM Transactions on Database Systems* 1: 9-36.
- FAO. 1992. FAO Expert Consultation on the Management of Global Animal Genetic Resources: Recommendations. The Management of Global Animal Genetic Resources. J. Hodges (ed.) FAO Animal Production and Health Paper No. 104. FAO, Rome, Italy.
- FAO. 2005. Domestic Animals Diversity Information System (DAD-IS), Food and Agriculture Organization. <http://dad-is.fao.org/>.

- Guralnick R and Neufeld D. 2005. Challenges building online GIS services to support global biodiversity mapping and analysis: lessons from the mountain and plains database and informatics project. *Biodiversity Informatics* 2: 56-69.
- Lescourret F, Genest M, Barnouin J, Chassagne M and Faye B. 1993. Data modeling for database design in production and health monitoring systems for dairy herds. *Journal of Dairy Science* 76: 1053-1062.
- Mpofu N, Ojango JM and Andersson-Eklund L. 2006. Quantitative methods to improve the understanding and utilization of animal genetic resources. Ojango, J.M. (ed.); Malmfors, B. (ed.); Okeyo, A.M. (ed.). ILRI, Nairobi (Kenya); Swedish University of Agricultural Sciences, Uppsala (Sweden). Animal Genetics Training Resources, Version 2. Capacity Building for Sustainable Use of Animal Genetic Resources in Developing Countries. CD ROM, ILRI Nairobi (Kenya).
- NBFGR. 2007. Vision-2025: Perspective Plan, National Bureau of Fish Genetic Resources, Lucknow, Uttar Pradesh, India.
- Nivsarkar AE and Gupta SC. 1999. Essential elements for survey, identification, characterization and monitoring in the conservation and management of genetic resources of livestock. *Indian Journal of Animal Genetics and Breeding* 21: 64-71.
- Wolf C. 1990. Computer analysis of survey data - file organization for multi-level data. URL = <http://www.aec.msu.edu/agecon/fs2/survey/level.pdf>.