AN INTEGRATIVE DATABASE SYSTEM OF AGRO-ECOLOGY FOR THE BLACK SOIL REGION OF CHINA

Cuiping Ge[1,2], Jun Zhao[1*], Shaoliang Zhang[3], Lei Shang[3] and Sheng Yin[4]

[1*] *Hailun Station of the National Field Station of Scientific Research and Observation, Heilongjiang Key Laboratory of Black Soil Ecology, Northeast Institute of Geography and Agro-ecology, Chinese Academy of Sciences, Harbin, 150081*
Email: jzhao@mail.hrb.hl.cninfo.net
[2]*Graduate University of the Chinese Academy of Sciences, Beijing, 100049, China*
[3]*HLJ August First Land Reclamation University, DaQing, 163319, China*
[4]*Northest Agricultural University, Harbin, 150030, China*

ABSTRACT

*The comprehensive database system of the Northeast agro-ecology of black soil (CSDB_BL) is user-friendly software designed to store and manage large amounts of data on agriculture. The data was collected in an efficient and systematic way by long-term experiments and observations of black land and statistics information. It is based on the ORACLE database management system and the interface is written in PB language. The database has the following main facilities:(1) runs on Windows platforms; (2) facilitates data entry from *.dbf to ORACLE or creates ORACLE tables directly; (3)has a metadata facility that describes the methods used in the laboratory or in the observations; (4) data can be transferred to an expert system for simulation analysis and estimates made by Visual C++ and Visual Basic; (5) can be connected with GIS, so it is easy to analyze changes in land use ; and (6) allows metadata and data entity to be shared on the internet. The following datasets are included in CSDB_BL: long-term experiments and observations of water, soil, climate, biology, special research projects, and a natural resource survey of Hailun County in the 1980s; images from remote sensing, graphs of vectors and grids, and statistics from Northeast of China. CSDB_BL can be used in the research and evaluation of agricultural sustainability nationally, regionally, or locally. Also, it can be used as a tool to assist the government in planning for agricultural development. Expert systems connected with CSDB_BL can give farmers directions for farm planting management.*

Keywords: Database management system; ORACLE; Agricultural ecology; Model; Expert system.

1 INTRODUCTION

The black soil in Northeast China is one of only three such areas found in the world. It is world-famous for its fertility and high-yield grain production and is an important base for commodity grain production in China. Nowadays this region is being faced with serious environmental problems, such as loss of arable land, land degradation and erosion, and more and more fertilizer use while grain quality goes down. Therefore, many scientists plan to do research on these problems. For this research, serial data is very important, and in order to develop sustainable agriculture in black land, it is imperative to make good use of agro-ecological resources. The CSDB_BL database is a very important tool in scientific research. Furthermore, sharing scientific data on the Internet will enhance the development of scientific activities and discovery. In recent years, numerous databases in agro-ecological research have been described, such as Xia, et al. (2003), Barrington (2002), Bodner & Perloff (2003), Clements, et al. (2002), Filis (2003), de la Rosa (2002), Beck (2001), and Kobrich, et al. (2003). Data sharing on the Internet also has occurred, such as with the USDA Research database, the ARS

Database, the FAOSTAT Statistics Database, the USDA National Plants Database, the Projects, Grain Genes, Northeast Food System Partnership, the USDA Forest Service Global Positioning System (GPS) Page, the Soybase, GardenGuides, and Ganaraska systems, and AgDB, AgDirect (CSDB, 2000).

In China, also many databases have been built in recent years. Each one has specific characteristics, such as the CSDB (Chinese Scientific Database) (Zhang et al., 1997). However, a completed agro-database for the black soil of Northeast China has still not been made.

CSDB_BL is a sub-database of the Chinese scientific database system. The purposes of setting up this database are to store and manage a large amount of data that have accumulated in long-term research, observations, and statistics concerning the black soil of Northeast China, and also to make those data available for analysis in strategic and tactical research in agriculture and provide stronger data support and expert suggestions to researchers, officers, and farmers. Through numerous data analyses, we may discover new theories and important patterns through agro-ecological research.

Research on agro-ecology is very complex as it must take into account various factors concerning not only agro-ecology but also economics and decision-making. Therefore, the data being collected and researched over long periods of time play a valuable role in agro-ecological research and lay down a framework for a macro-strategy of agro-development.

The *CSDB_BL* (Zhao et al., 2003) is a comprehensive database system; it stores a great deal of regional agro-ecological information, including data on water, soil, climate, biology, and other special projects. It also stores social, economic, environmental, and other data. Figure 1 illustrates the data model design. It has friendly user interface and useful facilities. The database can be connected with expert systems and GIS, so that the data can be selected and transferred to the model as input values. Attributive data also can be transferred to the GIS. Powerbuilder 8.0 and Oracle are the main programming languages of *CSDB_BL*; VC and VB are being used for expert system development; Java, FrontPage 2000, and ASP language are being used for web database development. The database aims are to a) classify systematically the data on agro-ecological research on black soil; b) manage and make use of the data through standardization; c) integrate with GIS and support modeling by an expert system; and d) share data on the Internet. Also, the database can provide a scientific basis for precautionary measures to ensure agricultural sustainability in Northeast China. It can also be expected that the database will play an important role in regional grain safety, economic security, and environmental conservation.

This paper describes the scientific concept and software development of the database and its connection with expert system modeling and GIS. Metadata and data sharing on the Web are also discussed.

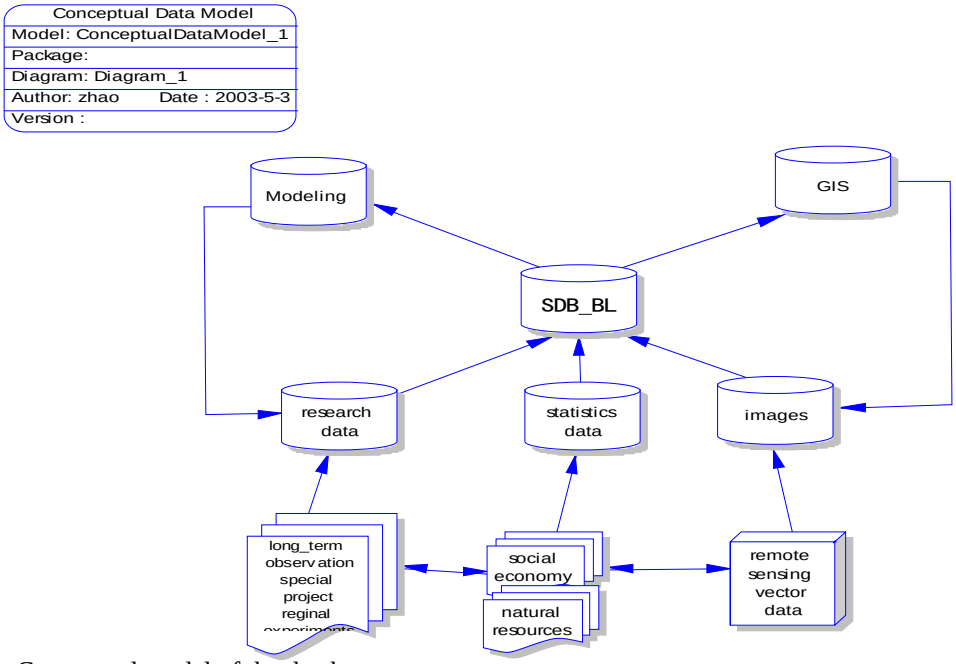

Figure 1. Conceptual model of the database

2 DATA STRUCTURE

2.1 Variable types

There are two types of variables: stored variables and driving variables. Stored variables in the database are classified into three sections: *research variables*, which include research on water, soil, meteorology, biology, and special projects; *statistical variables*, which included natural resources and social economy; and *remote sensing variables*, which included vectors and the grid's image data. The assistance menu aids in putting the data into the Oracle database or transferring it from *.dbf to Oracle. A conceptual data model of a soil dataset as an example is described in Figure 2 below. Driving variables are selected and calculated from the database according to the request of an expert system for macro-strategy and the need for tactical modeling. The data can be selected by the program and saved as *.dat for model use.

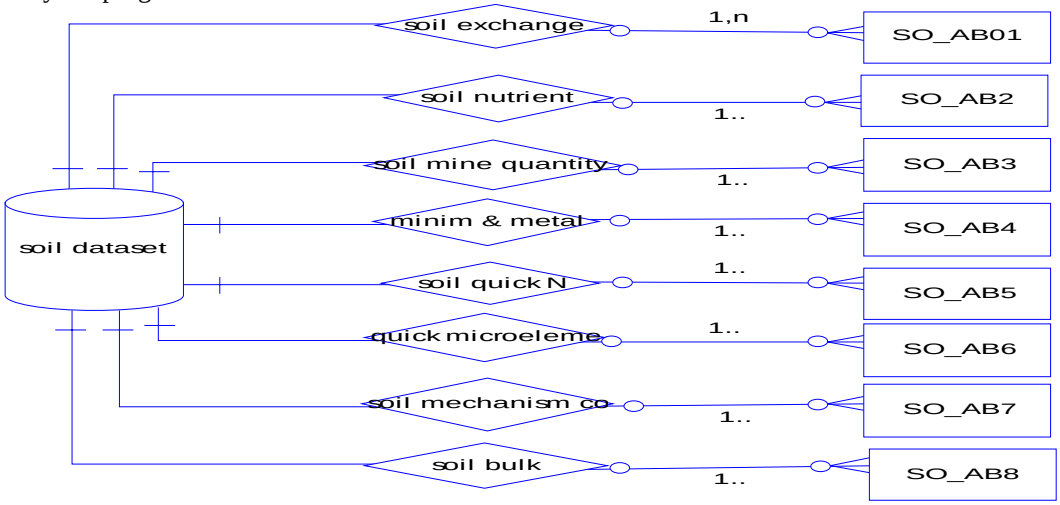

Figure 2. Conceptual data model of a soil dataset

2.2 Data types

Data stored in the database are as follows: Formats of grids, images, and pictures are stored as Arc/info coverage or bmp. Those of numbers and texts are saved as tables in Oracle or DBF. The codes of each type are defined as follows (Table 1).

Table 1. Codes of variables in the database

Code	Note	Code	Note
Wo	WATER DATASET	Tj	SOCIAL AND ECONOMY DATASET
Mo	METEOROLOGICAL DATASET	Pro	SPECIAL DATASET
So	SOIL DATASET	Vec	VECTOR DATASET
Bo	BIOLOGICAL DATASET	Dri	DRIVING DATASET

2.3 Data organization

Data tables derived from water observations and experiments are encoded as Wo, soil as So, biology as Bo, statistics as Tj, special projects as Pro, vector images as Vec, and driving valuables as Dri. These defined functions make data query easy. Meanwhile, the metadata are in one to one correspondence with the data entities, providing good clarity.

3 SYSTEM FUNCTIONS

3.1 System requirements

The software and hardware configurations required for installing CSDB_BL are: a PC computer with a WIN2000 platform, more than 256 MB RAM, 500 MB free disk space, a color monitor, a graphics adapter supporting 800640 pixels screen resolution, and an MS mouse or other compatible pointing device.

3.2 main menus

The *CSDB_BL* database is a standard Windows application using regular control functions. The software is written in Powerbuilder 8.0 and ORACLE. The expert system is written in VB and VC. Upon entering the database, a start-up window with a main menu appears at the top of the screen with the following main options: file, research data, nature resources, agricultural ecology, animal husbandry, expert system, GIS analysis, tools, windows, and help (Figure 3).

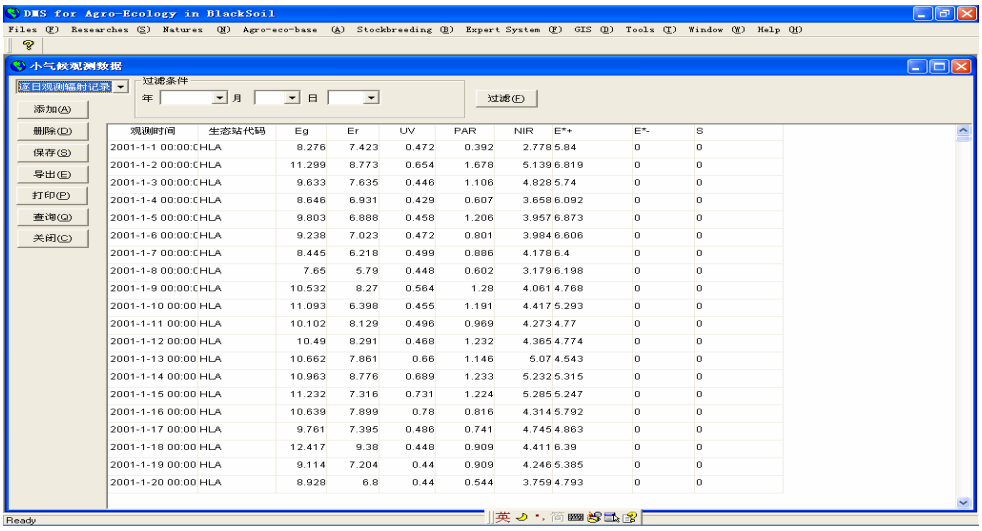

Figure 3. The main menu of the database

3.3 Main menu: file

This *CSDB_BL* option provides entry into the database and exit from it.

3.3.1 Import / export / edit/ delete/save

These sub-options are used to transfer data from VFpro to Oracle, to create new tables, and edit and delete tables. They also allow users to make inquiries on all the data tables and transfer or save data to other table formats, such as *.dbf, *.xls, and more than ten other types.

3.3.2 Management for county name and code

General information about the Northeast region is largely stored in a coded format system according to the criterion of Chinese regions, such as Hailun County coded as 232304. This code system is a part of the database, which makes querying easy.

3.3.3 Management of user priority

Three types of user priorities can be configured. The first type user is a DBA who has priority and full authority to manage the database. The second has the ability to enter into the database, modify, edit, and delete. The third is an ordinary user, who can only browse the data and run the expert model.

3.4 Main menu: research data

In this dataset, all the research data since 1980 from the Hailun long-term agricultural ecological experimental station, Songnen Plain, and Northeast of China have been stored. These data include water, soil, meteorology, biology, and also some special projects unique in China. The detailed descriptions of data are as follows:(1) moisture data: long-term experiments (begun in 1985) of soil water content with different treatments in the field layer, which begins 10 cm from the soil surface to a depth of 270 cm, and observations of the groundwater table in different crop growth periods, a total of 150 variables; (2) soil data: information on physical and chemical

characteristics of soil such as soil nutrients, fertility microelements, and so on, a total of 197 variables; (3) meteorology data: long-term observations of climate, radiation, normal meteorology for the Northeast since 1953, a total of 79 variables; (4) biological data: long-term observations of biology concerned with crop yield, energy accumulation, consumption, and so on, a total of 125 variables; and (5) special project data: including a survey of soil quality, experiments with different tillage, and so on, a total of 42 variables.

3.5 Main menu: natural resources data

In this dataset, information about animals and plants, meteorology, water resources, land resources, forestry, and social-economic resources, which are in severe danger in the region of black soil, have been stored since 1980. Data from more than 100 counties have been collected, which will be very useful for large-scale agro-research and macro-strategic planning, with a total of 157 variables.

3.6 Main menu: basic information of agricultural ecology

This section stores basic information concerned with regional agricultural ecology, such as area of arable land, fertilizer use for crop per hectare, irrigation information, farmers' income, cost of planting per hectare, economic benefits, and a survey of soybean seeds resources.

3.7 Main menu: the basic information of stockbreeding

This dataset stores information about regional stockbreeding, including husbandry information for cattle, pigs, and sheep. Production rates, costs and sale price and even information on poultry farming in Heilongjiang province are also included.

3.8 Main menu: Agro-expert system

This section plays a very important role in the database. Specific data can be selected from the database and shared for expert systems. There are two main sub-expert systems, one being a macroscopic strategy for evaluating changes in agricultural resources and environment (Liu et al., 2002) (Figure 4). Its purpose is to evaluate and analyze large-scale environmental changes and study the overall aspects and workings of local and national agricultural ecological and economic developments. Therefore, it provides the data and analysis necessary for the government to make optimum development plans for regional or national agriculture. It will be a very useful tool for sustainable agriculture and ecological and social development. The other section is tactical. It deals with problems that occur during the growing season because of disease, insects, and water and nutrition deficiency (El-Beltagy et al., 1995). One example is the DSS for Soybean Growth (Figure 5). Other examples include topics in meteorology, social economy, production yield, and the ecological background. The tactical part includes data on meteorology, soil property, planting, irrigation, and fertilization and harvest information.

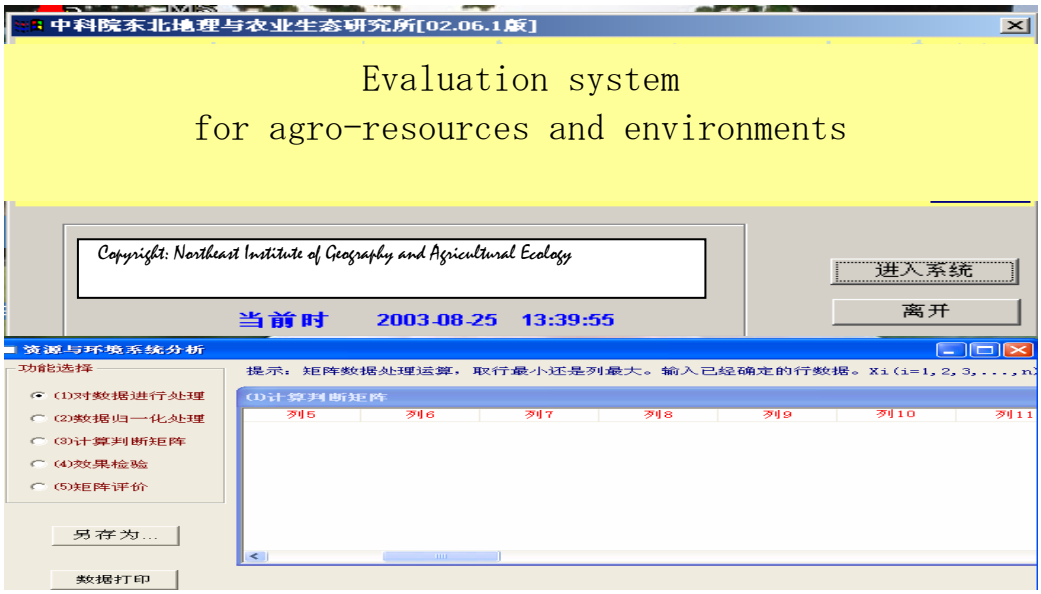

Figure 4. Data input window of evaluation expert system for agro-resources and environments

.

Figure 5. Soybean growth simulation expert system

3.9 Spatial dataset

This dataset includes two basic components: a geometrical database and an attributive database. The geometrical or spatial database comprises records of the location and extent of an object represented by a point, line, or surface and is handled by geographic information system (GIS) software, such as ARC/INFO, ARCVIEW, and SUPERMAP(Yu, 2001). Stored images data of remote sensing are included from 1980 to the 1990s and also in the year 2000, a land use map of 1:100000. The attributive database is geographically referenced and describes the characteristics and uses of the objects.

3.10 Tools

The *tools* have component functions that can be configured in the system by the DBA, such as the connection between databases and programs or the connection with other window interfaces.

3.11 Windows

The DBA can design *CSDB_BL*'s starting style, change the color and the picture. It can also set up an interface between the dataset and model.

3.12 Help

This function can help users understand how to use the database system and modeling. It also will be included in a user manual in HTML format.

3.13 Function of query data

Query data can be conducted by the inquire menu; therefore it can find the data according to location, time, key words, and synthesis inquiries. It provides not only written data but also pictures of seeds or plant diseases and insect pests. Query also supports the function of simply statistics.

For the spatial database, the IS of SUPERMAP software are used to develop the functions of WEBGIS, and SDE are used in the query function. Therefore, maps of land-use and other images can be viewed on the website with celerity. It also can connect with models. The query functions of the spatial database are as follows: (1) object to attribute, (2) attribute to object, (3) query for a certain scale geographical entity through views of a single spatial point, (4) whole query of joining attributes and spatial conditions.

4 METADATA

4.1 Design for metadata concept

Metadata are the data concerned with the content, quality, condition, and other properties of data. It is also called described data or annotation data. There is no different between metadata and other data which can be stored in any format.

While information technology has made some major technological advances, the large amounts of data being stored put forward a problem for data managements and users. However, the main problem is how to find and use the data quickly that are needed. Therefore, the concept of metadata has been introduced into database management.

This database uses the criterion of metadata version 1.0* for Chinese ecological research. Its content includes: dataset information, research project information, method information, dataset structure, dataset statement, Web information, scale information, cited information, and personal information. The conceptual model of soil NPK treatment and yield measurements at Hailun station has been used as an example of metadata (Figure 6).

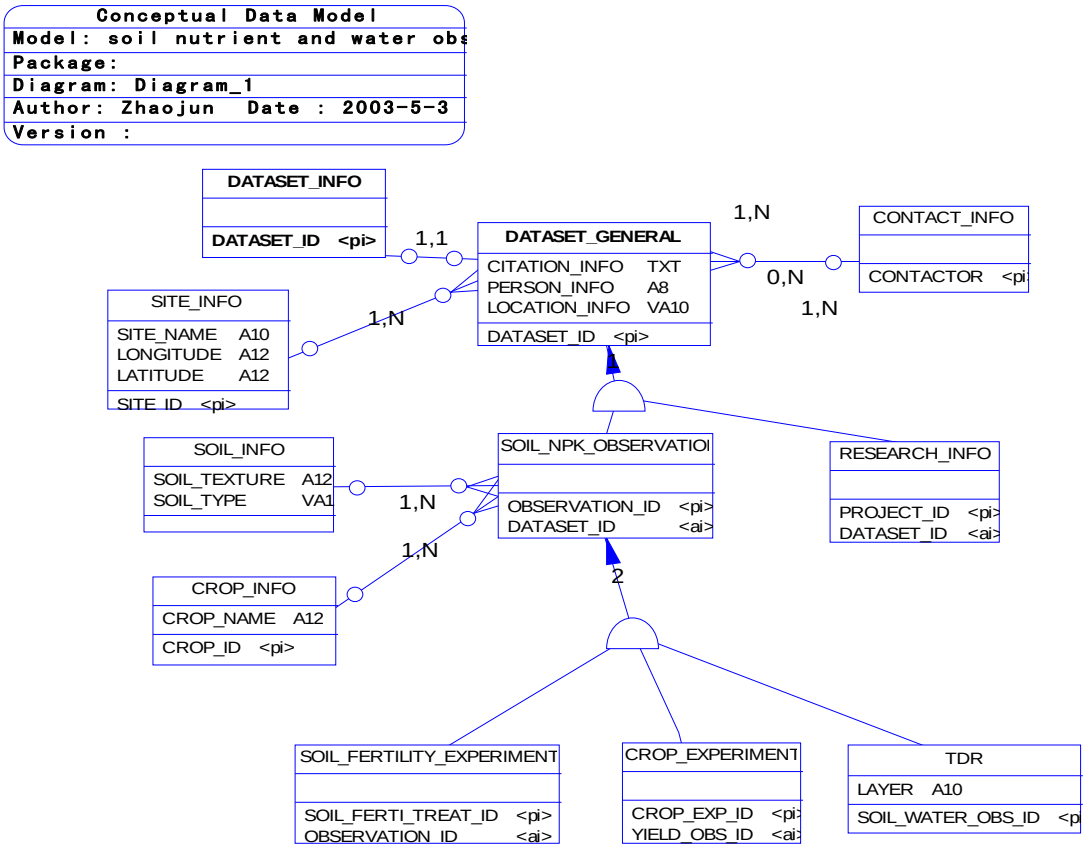

Figure 6. Conceptual model of soil NPK treatment and yield measurements at Hailun station

4.2 Metadata management

The management of metadata is controlled by the metadata server. It includes data lists, catalog establishments, and updates. It is developed in XML language.

5 DATA SHARING

Because of the importance of using the Web for data sharing, a Web database is now being constructed. The test website is http://www.blackland.csdb.cn. With this website, users from anywhere are able to browse freely and query any of the metadata of *CSDB_BL*. They can also download selected data chosen by specifying metadata (Figure7).

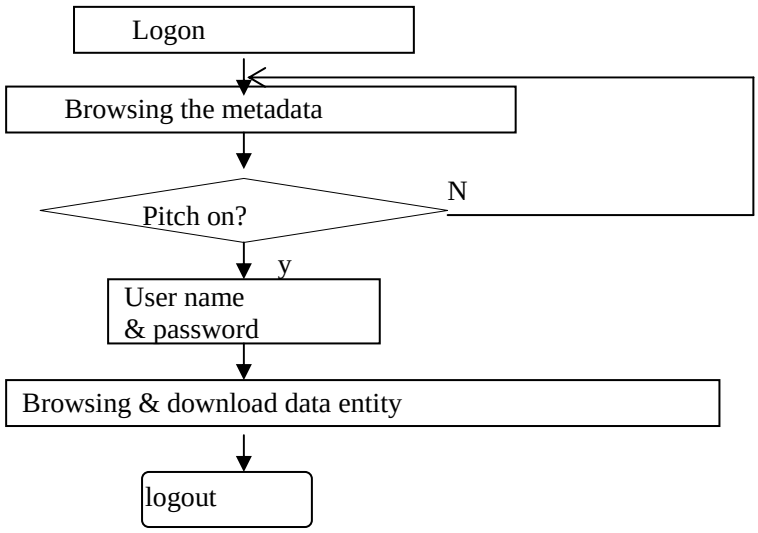

Figure 7. Design for the Web database application

6 CONCLUSION

The *CSDB_BL* database management system for Windows described in this paper represents a useful tool in the development of agricultural ecology. The software has been developed to enable the *CSDB_BL* to connect with GIS and several models. Each sub-dataset can be facilitated by a special research field on water, soil, climate, biology, etc. It also provides the design and implementation of environmental decision-making support systems. The *CSDB_BL* remains open for future development and improvement, including interfaces with innovative new programmers and connections with GIS and other expert systems. It will improve the data auto-transfer from dataset to the models and develop on-line Internet versions.

The application of *CSDB_BL* will help formulate new theories and monitor regular changes by analyzing the huge amount of agro-ecological data currently being collected. In the meantime, it will assist the government in developing a plan based on scientific data to improve China's agro-ecological and agro-economic position in the world, which in turn will hopefully lead to social economic advances for China.

7 ACKNOWLEDGEMENTS

This research was supported by the Chinese Academy of Sciences INF105-SDB-1-28 and National Science and Technology Project 2006DKA32300-04. The authors would like to thank the Heilongjiang Key Lab of Black Soils Ecology and Hailun agro-ecology station for their help and also Prof. Ben Price, who is working in Northeast Forestry University, for his help in manuscript preparation.

8 REFERENCES

Barrington, S., Nieuwenhof, P.,Vigneault, C., Choiniere, D., & Clarke, S. (2002) A computer program to size agricultural standby power units and electrical service entrances. *Computers and Electronics in Agriculture 36 (1)*, 79-86.

Beck, H. (2001) Agricultural enterprise information management using object databases, Java, and CORBA. *Computers and Electronics in Agriculture 32 (2)*, 119-147.

Bodner, J.E. & Perloff, B.P. (2003) Databases for analyzing dietary data-the latest word from What We Eat in America. *Journal of Food Composition and Analysis 16 (3),* 347-358.

Clements, A.C.A., Pfeiffer, D.U., Otte, M.J., Morteo, K., & Chen, L. (2002) A global livestock production and health atlas (GLiPHA) for interactive presentation, integration and analysis of livestock data. *Preventive Veterinary Medicine 56 (1),* 19-32.

CSDB (2000) Scientific Database of China. Retrieved from the WWW August 5, 2007: http://www.sdb.ac.cn

de la Rosa, D., Mayol, F., Moreno, F., Cabrera, F., Diaz-Pereira, E., & Antoine, J. (2002) A multilingual soil profile database (SDBm Plus) as an essential part of land resources information systems. *Environmental Modelling and Software 17 (1),* 721-730.

Doraiswamy, P.C., Pasteris, P.A., Jones, K.C., Motha, R.P., & Nejedlik, P. (2000) Techniques for methods of collection, database management and distribution of agrometeorological data. *Agricultural and Forest Meteorology 102 (1-2),* 83-97.

Franko, U., Schramm, G., Rodionova, V., Korschens, M., Smith, P., Coleman, K., Romanenkov, V., & Shevtsova, L. (2002) EuroSOMNET-a database for long-term experiments on soil organic matter in Europe. *Computers and Electronics in Agriculture 33 (3),* 233-239.

Filis, I.V., Sabrakos, M., Yialouris, C.P., Sideridis, A.B., & Mahaman, B. (2003) GEDAS: an integrated geographical expert database system. *Expert Systems with Applications 24 (1),* 25-34.

Honnay, O., Piessens, K., Van Landuyt, W., Hermy, M., & Gulinck, H. (2003) Satellite based land use and landscape complexity indices as predictors for regional plant species diversity. *Landscape and Urban Planning, 63 (4),* 241-250.

Kobrich, C.,Rehman, T., & Khan, M. (2003) Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in Chile and Pakistan.*Agricultural Systems* 76 (1), 141-157.

Krogh, L., Noergaard, A., Hermansen, M., Greve, M.H., Balstroem, T., & Breuning-Madsen, H. (2003) Preliminary estimates of contemporary soil organic carbon stocks in Denmark using multiple datasets and four scaling-up methods. *Agriculture, Ecosystems and Environment 96 (1-3),* 19-28.

Liu, Z.Q.,Wang, Y.R., Liu, J.D., et al. (2003) Identification of agricultural resource & environment determination of the principle grain production area in China. *System Sciences and Comprehensive Studies in Agriculture 19 (1),* 65-70. (in Chinese with English abstract).

Meng, K. & Zhang, X.Y. (1998) Mechanism and ecological recovery of blacksoil degradation at Songnen plain. C*hinese Journal of Soil Science 29 (3),* 100-102. (in Chinese with English abstract)

Nemes, A., Schaap, M.G., Leij, F.J., & Wosten, J.H.M (2001) Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology 251 (3-4),* 151-162.

Pehrsson, P.R., Haytowitz, D.B., & Holden, J.M. (2003) The USDA's National Food and Nutrient Analysis Program: update 2002. *Journal of Food Composition and Analysis 16 (3),* ***331-341.***

Xia, Y., Stinner, R.E., Brinkman, D., & Bennett, N. (2003) Agricultural chemicals use data access using coldfusion markup language and a relational database. *Computers and Electronics in Agriculture 38 (3),* 217-225.

Yu, C. (2001) Supermap: A new technology for precise agriculture. *Computer and Agriculture 9*, 27-28. (in Chinese)

Zhang, J.Z., Xu, Z.H., Hu, Y.R. (1997) Studying the development of scientific and technical databases in Canada and USA, Choosing our own way to scientific information development and services, - A report on the visit to the Scientific and Technical Data Centers in Canada and U.S.A. In Li,W.P. (ed.), *Thesis of Scientific Databases and IT.* Chinese Environment and Science Press: Beijing. (in Chinese with English abstract)

Zhao, J., Zhang, Y., Sui, Y.Y., & Zhang, Q.P. (2003) Primary establishment for database of Northeastern agricultural ecology based on WEB. *System sciences and comprehensive studies in agriculture 19(3),* 202-206. (in Chinese with English abstract)