НАУКОВІ ЗАПИСКИ ДЕРЖАВНОГО ПРИРОДОЗНАВЧОГО МУЗЕЮ
Том 19 Львів, 2004 С. 51-56

УДК 581.9

E.Julius, B.H. Prots

MODELLING OF PLANT SPECIES DATABASE (GIS) FOR THE TRANSCARPATHIAN FLOODPLAIN ECOSYSTEMS (UKRAINE)

Юліус А., Проць Б.Г. **Моделювання бази даних (ГІС) видів рослин Закарпатських заплавних екосистем (Україна)** // Наук. зап. Держ. природознавч. музею. – Львів, 2004. – **19**. – С. 51-56.

Визначено структуру бази даних видів рослин для аналізу флористичної різноманітності у заплавних екосистемах Закарпаття. Розглянуто функціональні взаємозв'язки структурних частин бази даних та її кінцевий результат. Визначено переваги ГІС (Географічна Інформаційна Система) бази даних у порівнянні із конспектом інформації. Зазначено її функціональну гнучкість та можливість взаємозв'язку із фауністичними базами даних.

Юлиус А., Проць Б.Г. **Моделирование базы данных (ГИС) видов растений Закарпатских пойменных экосистем (Украина)** // Науч. зап. Гос. природоведч. музея. – Львов, 2004. – **19**. – С. 51-56.

Определена структура базы данных видов растений для анализа флористического разнообразия в пойменных экосистемах Закарпатья. Рассмотрены функциональные связи структурных частей базы данных и ее конечный результат. Определены преимущества ГИС (Географическая Информационная Система) базы данных в сравнении с конспектом информации. Указана ее функциональная гибкость и возможность взаимосвязи с фаунистическими базами данных.

Problem description

The Transcarpathian flooded ecosystems represent the unique and one among the largest surviving refugium of the ancient riverine habitats of Central/Eastern Europe and largely overlooked highly important biodiversity hot-spots.

The high number of nationally and internationally threatened plant species (>150 species) and communities (>20 types) known to be present here, together with the imminent threat of destruction requires urgent establishment (or in some cases substantial enlargement) of national and international transboundary protected areas. The ecological importance of the riverine forests was not appreciated till recently. There exists also no effective conservation at the present moment. Furthermore, the present Ukrainian economic stagnation adds to the pressure upon these forests. Illegal logging, mismanagement, poverty, even political lawlessness remain a serious economic and environmental problem for the region. In these circumstances the country risks losing these unique monuments of nature [3, 4, 5].

The ideology of the present study has been strongly linked to the current WWF-UK funded project ("Transcarpathian Riverine Forests, Ukraine" Project, UK006702P). The aim of which is to provide the baseline ecological and environmental information, and thus develop databases and models for biodiversity conservation, restoration and sustainability of the habitats.

The importance of plant species database for the region has been recognised due to (1) permanent loose of the field data over the last decades (poor publishing) or limited access for the data (bureaucratic system, which needs a special permition); (2) neglected and understudied type of the habitat (floodplains have been not important areas for the research); (3) absence of herbarium materials and original information for many taxa (especially for the last decades). There is no certain information system available at the moment for the Transcarpathia, neither on biodiversity nor on parts of it.

The study focuses on modelling of plant species database (as a first database example) for the Transcarpathian floodplain ecosystems.

Methods

Database structure

The modelling theory approach for relational databases (Entity-Relation-Models - ER models, which is based on Hoffer-Prescott-McFadden notation) is used to create the plant species database for the region [6, 8, 10, 15]. The main feature of a relational database is the fact that all relations between data have to be presented by values [10]. The well structured relations contain minimal redundancy and allow the user to insert, modify or delete the rows in a table without errors and inconsistencies [8].

The three rules of normalization can be applied to reduce anomaly and redundancy in the database. Normalization is a formal process for deciding which attributes should be grouped into a relation. The first rule of normalization prohibits multi values as attributes and requires a value for every cell in the table. The primary keys of database have to be incorporated in the tables for clear identification of datasets. The existence of a primary key for every table is the second rule of normalization. The third rule of the normalization process contains that also all attributes only depend on the primary key [6, 8, 10, 11, 15].

Integrity constraints as important features of the database are used to tighten relationships within it. Integrity has two aspects: entity integrity and referential integrity. The entity integrity is a basic demand within the structure of a table. It prohibits that any primary key has the value null [6, 15]. The referential integrity demands certain relations between primary and foreign keys of two or more tables. A foreign key is a value field which has a relation to a primary key of other table. An example for a referential integrity constraint is that each foreign key value must match a primary key value [6].

The Entity-Relation-Models for all attributes, entities and their relations are displayed on Figure 1.

The “Microsoft Access 2000®” software is used for the practical realisation of the relational database. The reason to apply “Access 2000” is its possibility of adaptation to different sets of characters (Latin/Cyrillic). The “ArcView 2.3” software is approached as GISystem (Geographical Information System) software. The square codes are arranged according to the “Mapping of the Flora of Central Europe” [12]. The nomenclature is based on Tutin et al. [13, 14].

Demand for the data quality

The results of field inventories, analysis of herbaria data (UU, KW, LW, LWS etc), private collections and literature citations are used to accumulate the content of the database. All these data were saved on lists in the past.

These lists are, on the one hand, a huge resource of information, but, on the other hand, a problematic instrument to handle without the knowledge and manual interpretation of involved scientists. The data in these lists is strongly user dependent and can support only few standards. Also it is not possible to use computer aided analysis, like automatic preparation of distribution maps. These lists have a high data redundancy which causes inconsistency. Assuming changes of taxonomy all entries in the lists have to be changed manually. The advantages of an information system in form of a relational database have been proven by examples like Bioflor, British Flora and Flora Helvetica [1, 7, 9]. The need for an information system, in form of a database, has been recognised during the establishment of the project [4]. It is hugely needed to improve the quality of the data as well.

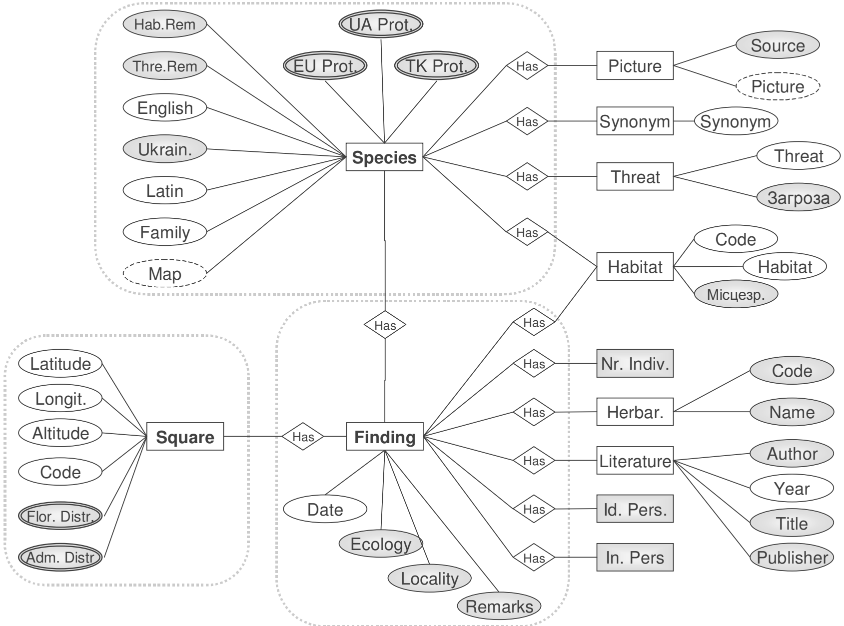

Fig 1. The entity-relation (ER) model of the plant species database for the Transcarpathian floodplain ecosystems: the square symbols symbolise entities, the oval symbols – attributes and the rhombus symbols – relations.

Results

The entity-relation (ER) model of the plant species database for the Transcarpathian floodplain ecosystems is presented on Figure 1. It contains of three major entities with several attributes. The most basic entity is 'Square'. The information related to the GISystem is located here.

The second basic entity is 'Species'. The information relying on the species is situated here. For each species we can find information on the protection status, threats, habitats, family names, common names and more important all known synonyms. The possibility of entering synonyms assists in solving problems of changing taxonomy. If a new name is assigned to a species only one entry in the entity 'Species' has to be changed. All other applications in the database are connected to the species key, which will be not changed.

The heart of the database is the entity ‘Finding’. Every location of a species has to be entered here. The square number and the species name need to be chosen to link the square and species information with the locality. The database faces the problem, which is connected to the use of two sets of characters. The grey colour is applied to visualise attributes which contain the Cyrillic characters.

Queries are used to analyze the data in the database. The queries once defined will be saved and are continuously available. The example for the distribution mapping query looks like following: “In which squares has the species [species name] been identified since 1950?” The query results are located in a table, which contains the square code and the desired information. This table can be exported from the database and imported to the GISystem.

The key to the GISystem is the square code. The queried data can be connected to the map with this code of the GISystem. As an example of distribution map is presented on Fig.2. The size of the square is 10 x 6 minutes according to the “Mapping of the Flora of Central Europe” [12]. Each square is represented by four points.

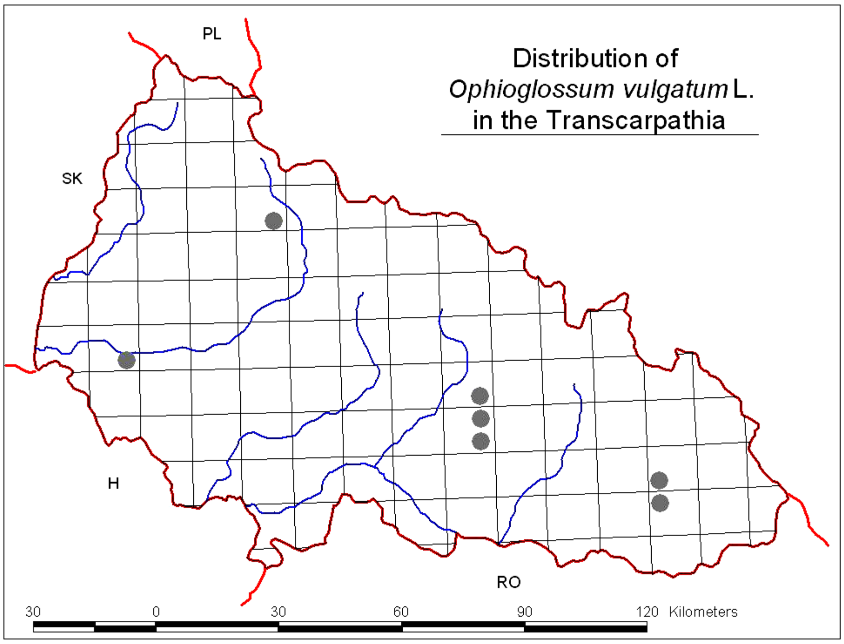

Fig.2. The distribution map example for plant species database.

Biodiversity analysis will be carried out on the species diversity level. The first step of the biodiversity analysis will be the calculation of the species richness. The species richness measure is based on species richness and sample size [2]. A distribution map of this species richness measure will be prepared. This map will be used to identify the diversity hotspots. Also a diversity index can be introduced to raise the quality of the biodiversity analysis. The diversity indices are based on species richness and their relative abundance [2].

The diversity indexes for special single group of species (like plant species or mammals) and totally for several groups might be a useful instrument to identify the biodiversity hotspots and could assist to solve the conservation management problems.

Conclusion

The advantages of the use of a relational database in connection to GISystem compared to the lists are obvious. The redundancy will be minimized in the relational database by the technical point of view. This will reduce the needed disk space compared to the lists and will increase the speed for data processing. The mistakes, which occur during the data entering, will be reduced due to the use of "selection" menus. For example, the Latin names will be only entered once and will be later only chosen from the menus. Despite the technical aspects the database will accelerate the data processing for scientific studies compared to the lists. The access for the pre-processed data (e.g. distribution maps) will be much easier. It might assist not just the project studies but the other distribution research. Of course, there are also disadvantages of the database application. First of all extra work is needed to transfer old data to the new database system. However, this problem will disappear in the future. Even more, the locality of finding will be presented in geographical coordinates and not described in words. Not only a standard for locality but also other standards like habitats will be enforced. The incorporation of standards is crucial to use computer aided analysis, to decrease the user dependency of data and to support data sharing. Compared to the use of lists new problems of organizational nature can occur. The centralisation of the data is on the one hand a good for the sharing of data but on the other could be an obstacle for data collection.

Outlook

The described relational database has a flexible structure and offers the possibility to incorporate more fields of studies. At the moment it is planned to insert studies on mammals, birds, amphibians, reptiles, fishes, molluscs and a number of entomological research.

After the project "Biodiversity, Conservation and Sustainable Use of the Transcarpathian Riverine Forests" has come to an end, the database will be valid. It could be used for the administration of possible protection areas.

It exist a potential to increase the size of the area for which the database will be valid.

Acknowledgement

We thank the WWF-UK ("Transcarpathian Riverine Forests, Ukraine" Project, UK006702P) and anonymous author for improving of English.

1. British Flora The New Atlas // Oxford University Press. – 2002. – Software.
2. Cogalniceanu D. Biodiversity // Verlag Dr. Kessel. – 2003. – 105 p.
3. Drescher A., Prots B, Mountford O. The world of old oxbowlakes, ancient riverine forests and meliorated mires in the Tisza river basin // Fritschiana. – 2003b. – Vol.45. – P. 43-69.
4. Drescher, A., Prots B., Mountford, O. Project Transcarpathian riverine forests, Ukraine: at the start point // Benedek, J.,Schulz, E. (Hrsg.) Südosteuropa. – 2003a. – WGM 63. – P. 139-144.
5. Hamar, J.; Sárkány-Kiss, A. (eds.). The Upper Tisa Valley. Preparatory proposal for Ramsar site designation and an ecological background. Hungarian, Romanian, Slovakian and Ukrainian co-operation. – Szolnok, Tisza Klub for Environment and Nature, 1999 – 502 p.

6. Hoffer J.A., Prescott M.B., McFadden F.R. Modern database management // Pearson Education New Jersey. – 2002. – 638 p.
7. Klotz, S., Kühn, Ingolf, Durka, W. BIOFLOR - Eine Datenbank mit biologisch-ökologischen Merkmalen zur Flora von Deutschland, m. CD-ROM Hrsg. v. Bundesamt für Naturschutz 2003. XVI, 334 S.
8. Kroenke D.M. Database Processing // Pearson Education New Jersey. – 2004. – 702 p.
9. Laubner K., Wagner G. Flora Helvetica 1.0 // Paul Haupt Verlag. – 1997. – Software.
10. Matthiessen G., Unterstein M. Relationale Datenbanken und SQL // Addison-Wesley (Deutschland). – 2000. – 348 p.
11. Nicol N., Albrecht R. Microsoft Access 2000 – Das Handbuch // Microsoft Press (Deutschland). – 1999. – 838 p.
12. Niklfeld, H., 1971: Bericht uber die kartierung der flora Mitteleuropas. Taxon 20(4): 545-571.
13. Tutin T.G., Burges N.A., Chater A.O., Edmondson J.R., Heywood V.H., Moore D.M., Valentine D.H., Walters S.M. & Webb D.A. 1993. – Flora Europaea. Vol. 1, 2nd ed.
14. Tutin T.G., Heywood V.H., Burges N.A., Moore D.M., Valentine D.H., Walters S.M. & Webb D.A. 1968, 1972, 1976, 1980. – Flora Europaea. Vols. 2-5.
15. Vossen G. Datenbankmodelle, Datenbanksprachen und Datenbank-Management-Systeme // Addison-Wesley (Deutschland). – 1994. – 686 p.

University of Applied Science Eberswalde, Eberswalde, Germany
State Museum of Natural History of NAS of Ukraine, Lviv, Ukraine