



# Status of the electronic documentation system in the Herbarium of Agricultural University – Plovdiv (SOA)

Kiril Stoyanov

Department Botany - Agricultural University of Plovdiv

Department Botany, Agricultural University

12 Mendeleev Str., 4000 Plovdiv, Bulgaria,

tel. +359 32 654332

e-mail: orobanche@abv.bg

http://www.botanica.hit.bg

SOA herbarium contains more than 100000 herbarium samples. Analyzing the existing systems and standards, an electronic herbarium documentation system is created. The presented article represents the structure of the database and the state of the entered data till now.

## Software description

The program module (dSOA) of the database is compiled in the Botanical department of Agricultural University. It requires low system resources – Windows 95 on Pentium 100 with 16 MB RAM, 2 MB free disk space and VGA. This program does not require any external data software. This software allows to create unlimited number of thematic relational databases. The user interface of the program module is realized in Bulgarian language. The main data windows offer the both table and form view of the records. To reach compatibility with other systems is predicted that dSOA system could export and import CSV- tables, text reports and user-defined formats. In example some data could be exported as a Java-midlet and to be read on the field using any mobile device. dSOA could prepare complete report by number of stored records as well their relation to users, families, genera, authors and type specimens. The program module is able to display the distribution of selected specimens as UTM (MGRS) map (Fig.1).

## Database structure and data description (fig. 2)

The structure of the documentation system is shown in fig. 2. The database is multi-file realized.

**User access.** The user access is realized in 3 levels – administrator ("root"), registered users and unregistered users ("guest").

**Taxonomy.** The taxonomic description is divided in 3 indexed tables as follow: family, genus and species. These tables are available in the user interface as menus. The real data in the fields is the index code. This division is realized because to avoid typing errors which are indispensable if the data is stored manually and to make the main data file more compact. The field of intraspecific taxa remains manually entered.

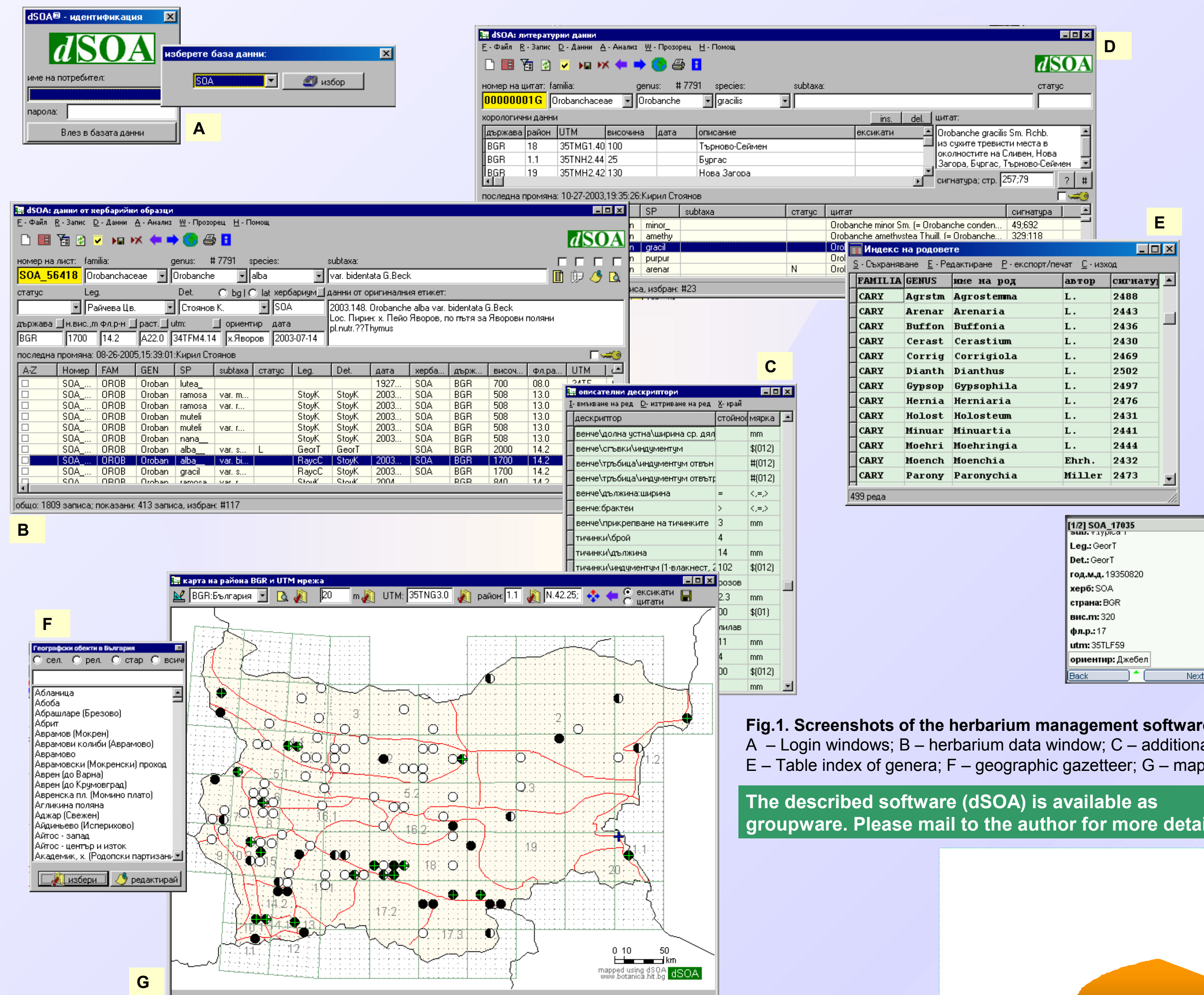


Fig. 1. Screenshots of the herbarium management software: A – Login windows; B – herbarium data window; C – additional description data; D – bibliographic data window; E – Table index of genera; F – geographic gazetteer; G – map drawing window; H – data exported in Java midlet.

The described software (dSOA) is available as groupware. Please mail to the author for more details.

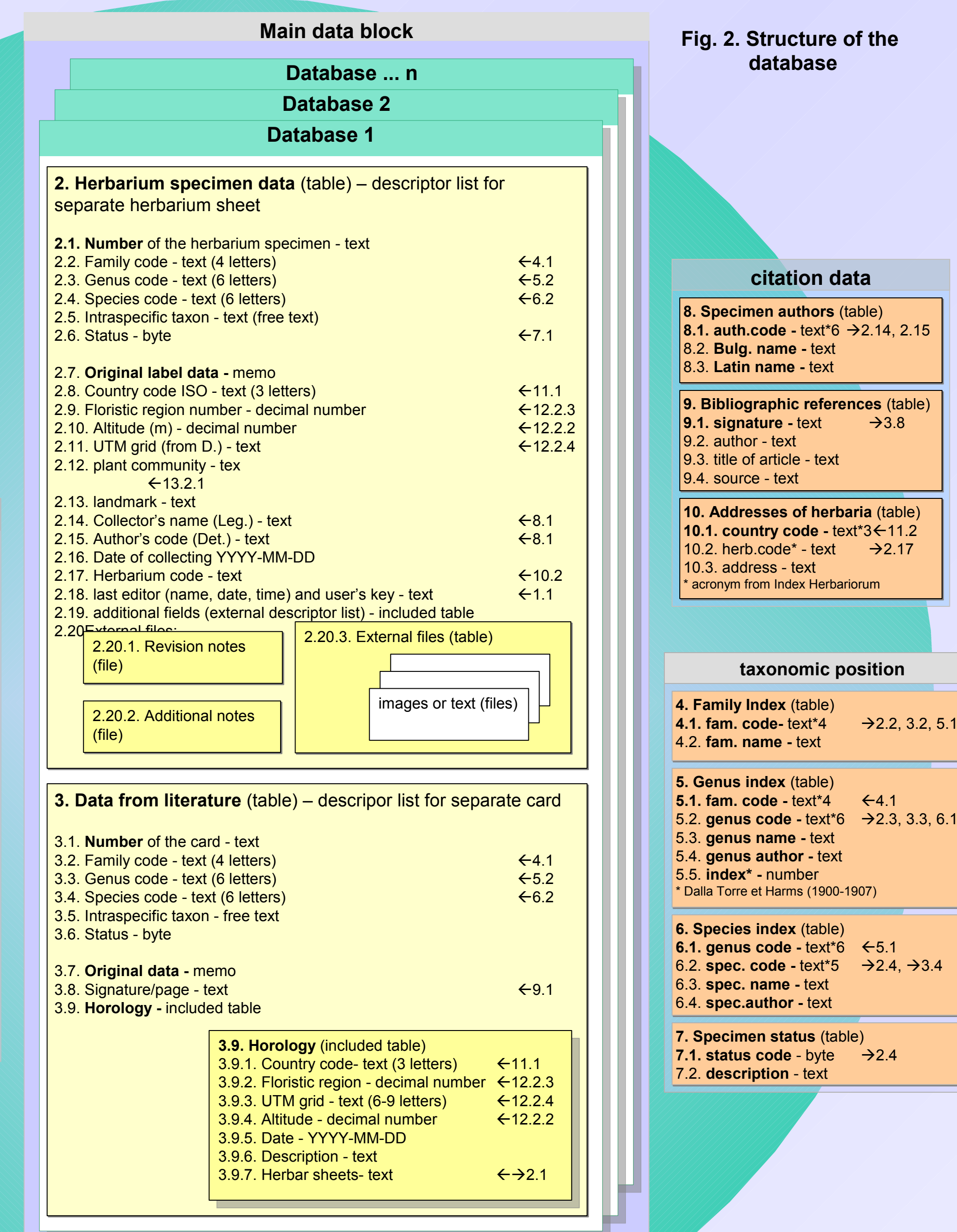


Fig. 2. Structure of the database

**Citation data.** Three index tables are created to enter citation data. The first one describes the authors of the herbar sheets. The index code allows to view the name of the authors with the both alphabets. The second index describes the literature sources and allows to search floristic data from the literature. The Index Herbarium acronyms are stored in unlinked file available as a list.

**Geographic reference data.** The names of the countries are indexed in a table as 3-letter ISO standard codes. The country index is used as a trigger for the other two stacks of tables in this group – gazetteer and plant communities. The first one is physically related to the main data set and consists toponyms, altitudes, floristic regions, UTM (MGRS) grid codes and geographic coordinates. It is related to map definition sheet. The plant communities are stored as mapping units.

**Main data block.** The main data block is realized as a stack of thematic databases. All the databases are built on the basis of two tables – data from herbarium sheets and data from publications. The user can switch between different thematic databases. The system keeps in each record information about the last data change. The data fields as family, genus, species, status (type specimens), geographic data, authors' names and last editor data are linked to the tables described above. Coding and linking the fields with uniform values to external tables we achieve not only better compactity but we avoid technical errors on data input.

**Data from herbarium specimens.** We accept the number of herbarium sheet as an unique number or primary key index in database. In this way we answer to the both requirements: searching by serial number and limiting the record/number duplication. The date format is "year-month-day" in case to simplify searching. The fields of the intraspecific taxa, data from the original label and landmark have to be stored manually from the keyboard. Because of the diversity of the data, every record can be linked to external files and additional data fields. The additional data fields are defined in a descriptor list. The external files consist one text file for the revision notes, one file for additional notes for the specimen and one table of other external files, incl. images. This decision allows to expand the data structure according to the features of the stored taxa. The stored data is searchable by the main fields and by existence of additional files.

**Data from literature.** The chorological data in the literature is structured in another order. Each floristic record for a taxon contains a list of localities sometimes with citation of herbar sheets. It means, each record in the database must have included table with similar structure. The other fields are linked as the fields in the herbar records. The field "signature/page" is linked to the bibliographic reference table. This organization of the database allows comparison between data from literature and data from herbar sheets. This makes possible to create comparative distribution maps of any taxon.

## Overview of the herbarium data stored till now (fig. 3-6)

The stored data till now contains 7004 records representing information about 75 families, 275 genera and 845 species, collected by 40 authors. The herbarium contains 210 types. The stored data cover the 20 floristic regions, mainly from the Rhodopes, Tracian plain, Black Sea coast and Stara-planina. The territory of Bulgaria is presented by 399 UTM (MGRS) squares 10x10 km. The bigger part of the specimens is collected from the lowland's belt. The most intensive collecting season is between may and july, followed by august and september. The stored records show that the bigger part of the stored specimens is collected between the 40s and 70s of the 20<sup>th</sup> century.

## Състояние на електронната система за документация в хербариума на Аграрен университет - Пловдив

Кирил Стоянов - Аграрен университет - Пловдив  
orobanche@abv.bg - www.BOTANICA.hit.bg

Представена е електронната система за документация на хербариума SOA. Обсъждат се структурата на базата данни и състоянието на въведените данни до сега. Програмният модул (dSOA) има ниски системни изисквания и не се нуждае от външен софтуер за управление на данни. С негова помощ могат да се създават неограничен брой тематични релационни бази-данни. Потребителският интерфейс е на български език

Главните прозорци за данни предлагат едновременно изглед на записите в таблица и формуляр. За постигане на възвместимост с други системи, е предвиден експорт/импорт на таблици, текстове, или друг дефиниран от потребителя формат. dSOA може да подготвя подробни отчети по въведените показатели и да построява UTM-карти на образците (Fig. 1). Структурата на системата е показана на фиг. 2. Базата данни е многофайлова. Потребителският достъп е на три нива. Таксономичното описание е разделено в индексирани таблици, достъпни като потребителски менюта. Данните за цитирани са индексирани в три таблици - автори, потребителски източници и кодове на хербариумите.

Имената на държавите са индексирани като трибуквени кодове по ISO стандарт, които се използват за превключване между таблици с географски данни - топонимни справочници (свързани с дефиницията на карта) и растителни съобщества (картируеми единици). Главният блок с данни е реализиран като стек от тематични бази данни, които могат да се прелистват по време на работа. Всички бази данни са построени на основата на две главни таблици от данни - от хербарни образци и от публикации. Системата запазва във всеки запис информация за датата и последния потребител. Хербарните образци се записват под номерата, под които са заведени в хербариума за да могат да се търсят по серийен номер и да се избегне дублиране на номера. Индексираният полет, описани по-горе, са свързани към главните таблици. Датите се въвеждат във формат "година-месец-дата" с цел опростяване на заявките при търсене. Тъй като информацията е разнообразна, е предвидена възможност за добавяне на външни полета с данни и външни файлове (бележки, ревизии, произволен брой допълнителни файлове) към всеки запис.

Така въведените данни могат да се филтрират по всяко главно поле и по наличието или липсата на външни полета и файлове. Литературните данни са структурирани в записи с вложени таблици. Съхранените до сега данни съдържат 7004 записа, представящи информация за 75 семейства, 275 рода и 845 вида, събрани от 40 автори. Хербариумът съдържа 210 типови образци. Съхранените данни покриват всички 20 флористични района, главно от Родопите, Тракийската низина, Черноморското крайбрежие и Стара планина. Територията на България е представена от 399 UTM (MGRS) квадрата 10x10 km. Най-много образци са събрани от низинния пояс. Най-интензивен сезон за събиране на образци е между май и юли, следвани от август и септември. Съхранените записи показват, че голямата част от образците е събрана между 40-те и 70-те на XX век.

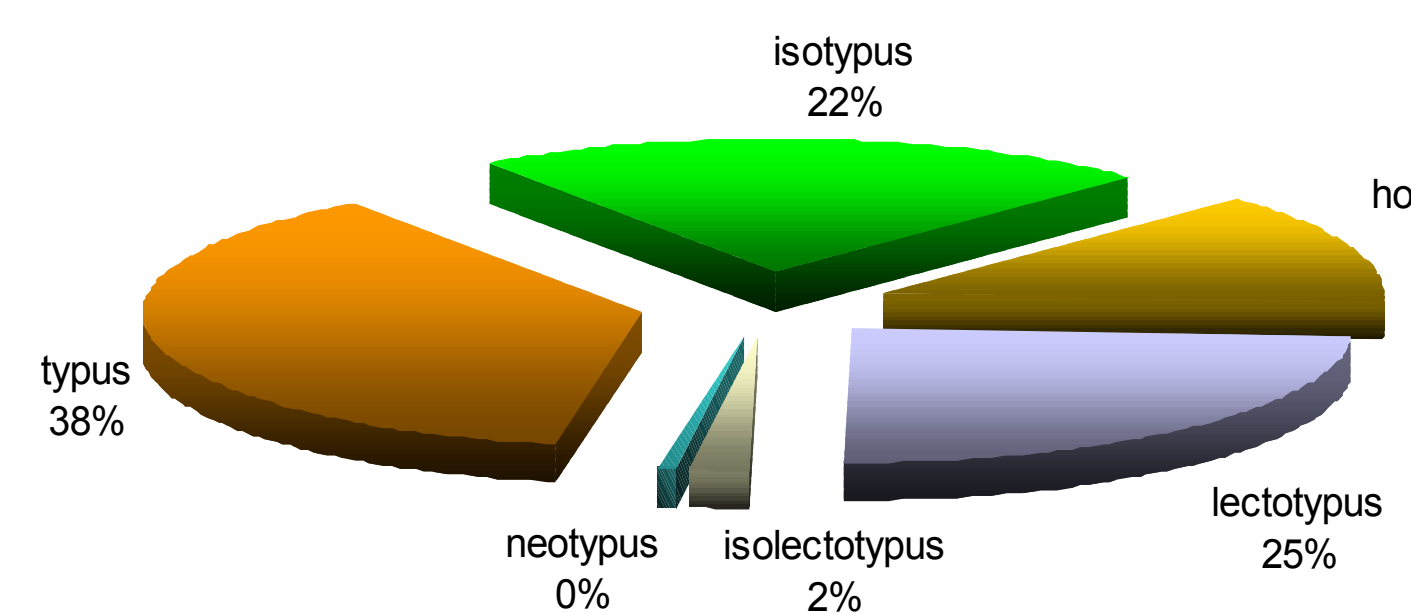


Fig. 4. Type collection.

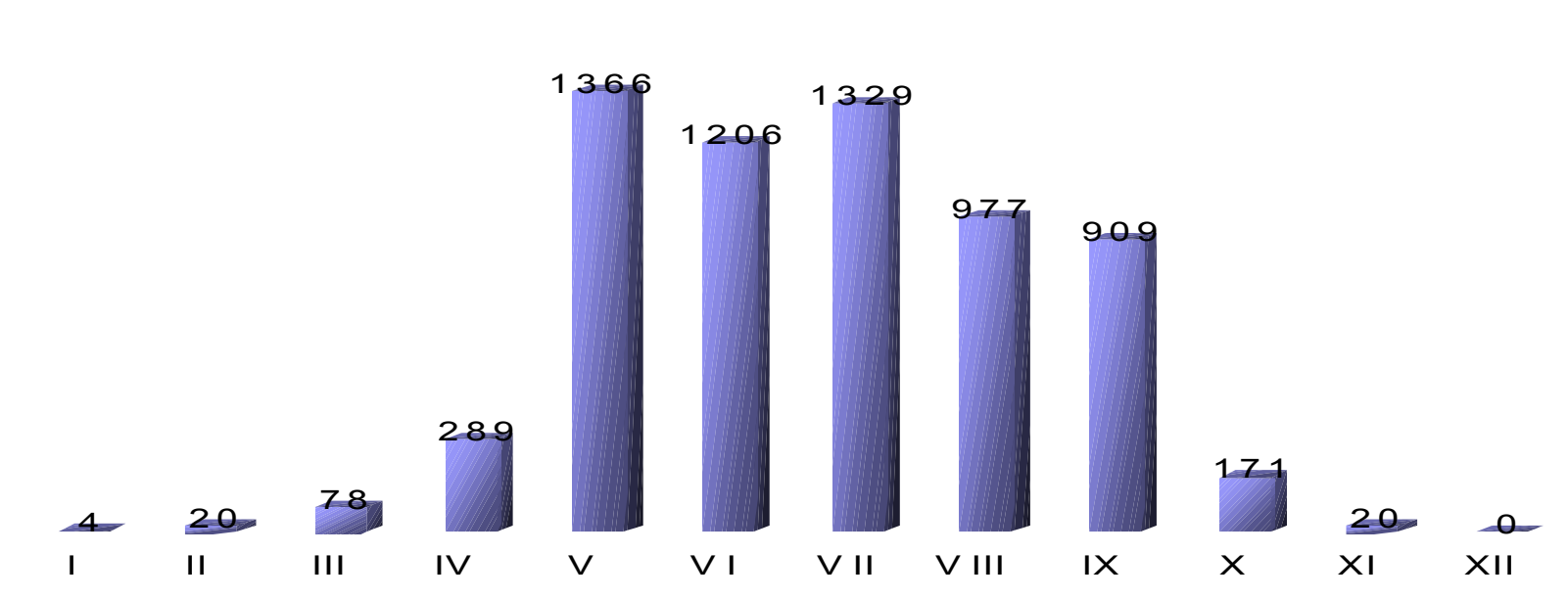


Fig. 3. Records divided by months.

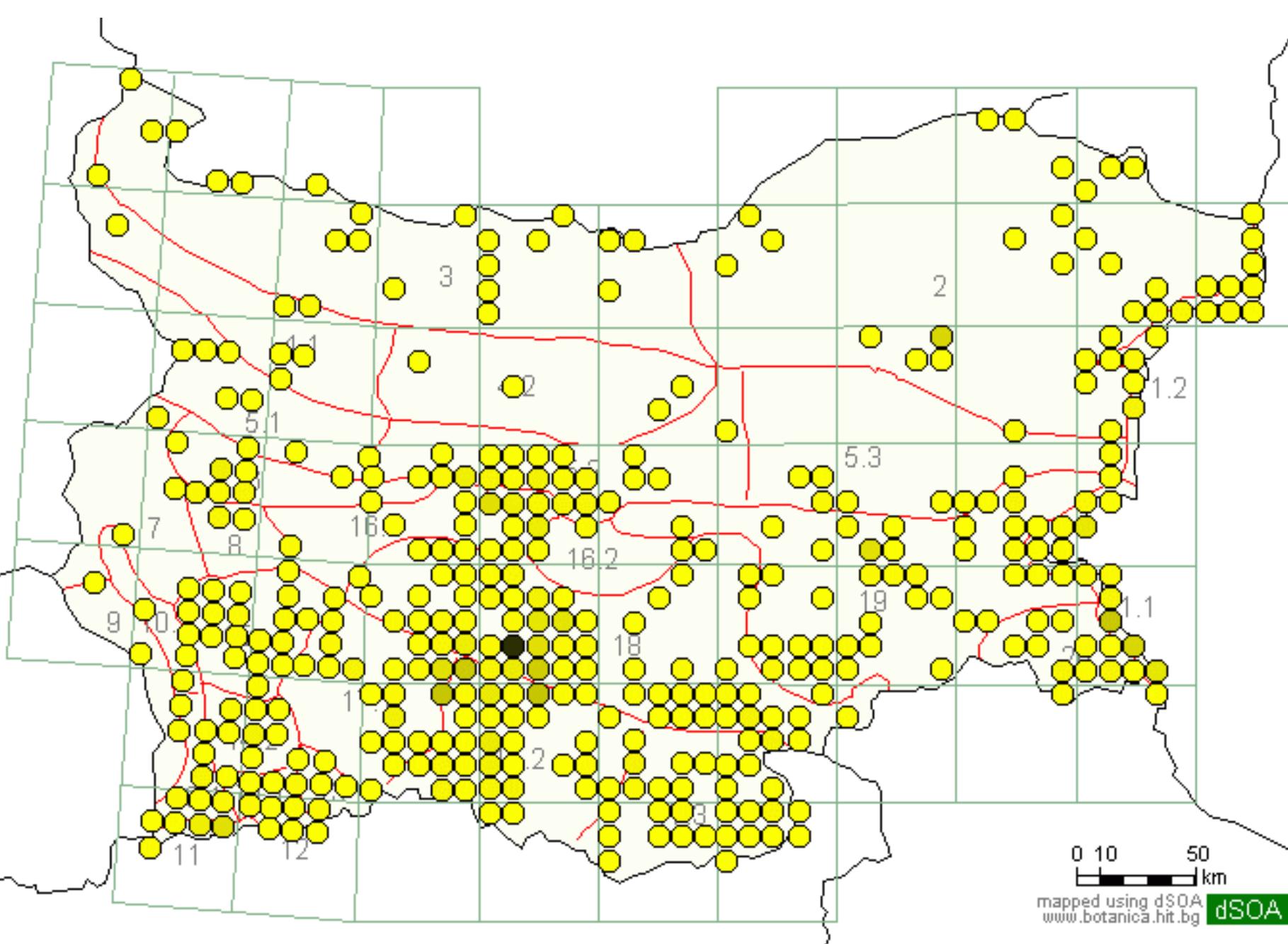


Fig. 5. Horizontal distribution according to the UTM grid 10x10 km.

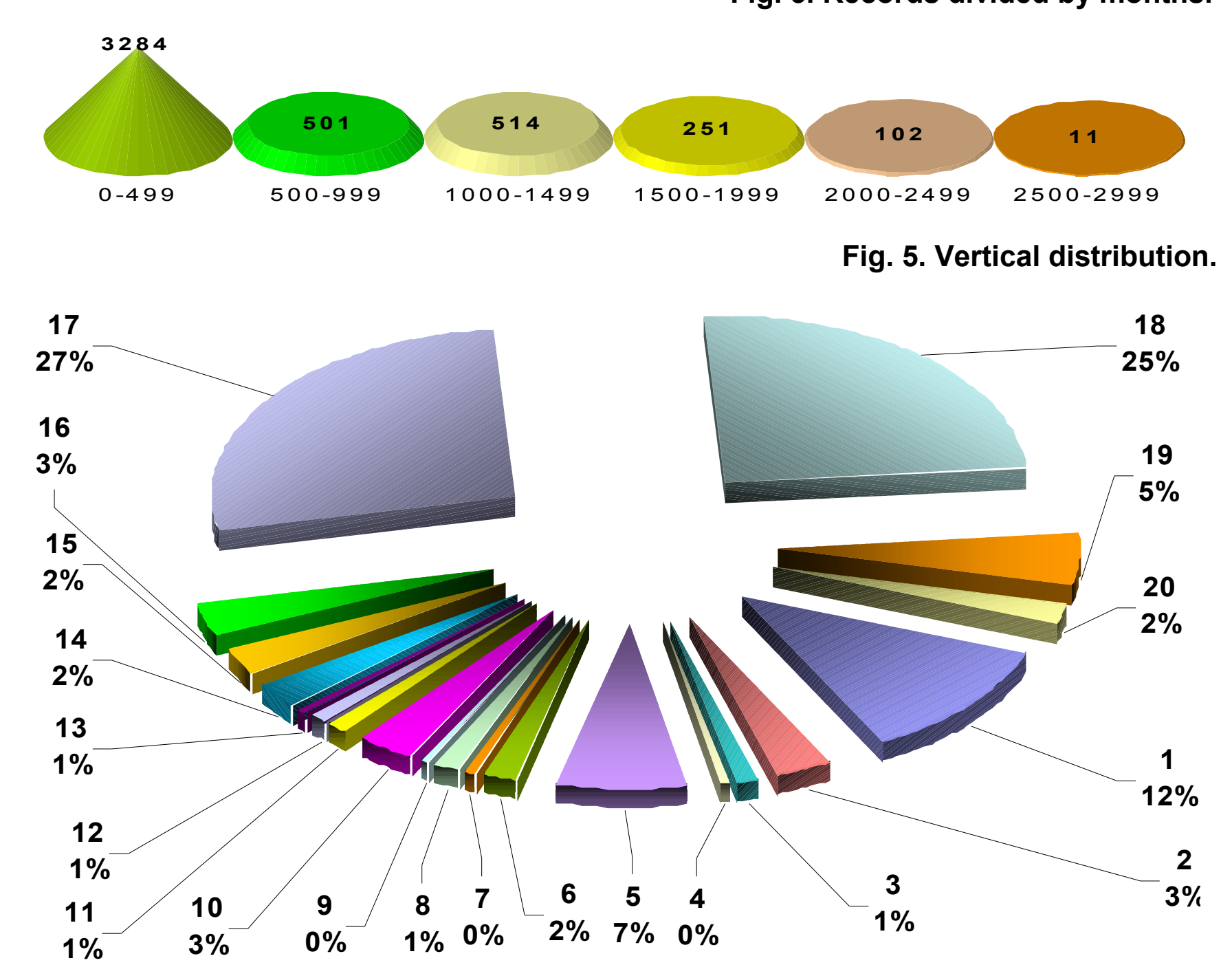


Fig. 6. Horizontal distribution by floristic regions (numbers from Fig.5.)

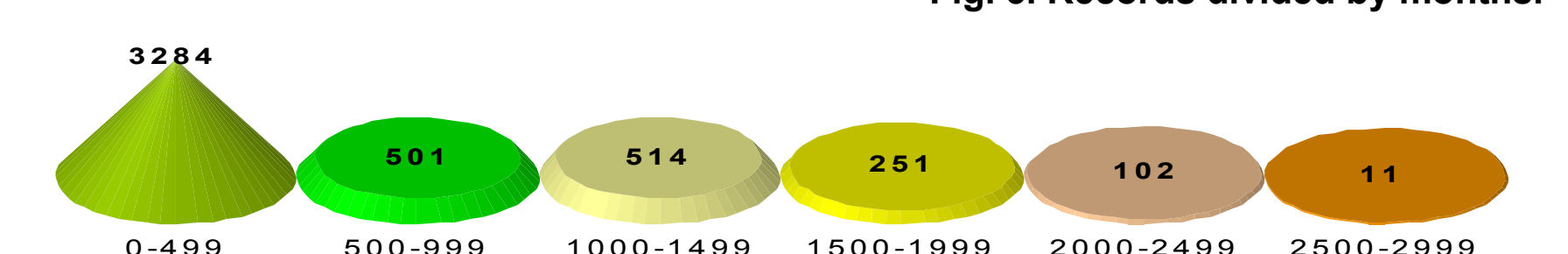


Fig. 5. Vertical distribution.