

# Regulatory Mechanisms in Biosystems

ISSN 2519-8521 (Print)  
ISSN 2520-2588 (Online)  
Regul. Mech. Biosyst.,  
2023, 14(3), 492–496  
doi: 10.15421/10.15421/022370

## Structure of data in cell biology research

V. Langraf\*, A. Svoradová\*\*, \*\*\* , K. Petrovičová\*\*\*\*, V. Brygadyrenko\*\*\*\*\* \* \* \* \* \*

\*Constantine the Philosopher University in Nitra, Nitra, Slovakia

\*\*Institute of Farm Animal Genetics and Reproduction, NPPC, Research Institute for Animal Production in Nitra, Nitra, Slovakia

\*\*\*Mendel University in Brno, Brno-sever, Czech Republic

\*\*\*\*University of Agriculture in Nitra, Nitra, Slovakia

\*\*\*\*\*Oles Honchar Dnipro National University, Dnipro, Ukraine

\*\*\*\*\*Dnipro State Agrarian and Economic University, Dnipro, Ukraine

### Article info

Received 02.07.2023

Received in revised form 09.08.2023

Accepted 20.08.2023

Constantine the Philosopher University  
in Nitra, Tr. A. Hlinka, 1, Nitra, 94901,  
Slovakia. E-mail: langrafvladimir@gmail.com

Institute of Farm Animal Genetics and  
Reproduction, NPPC, Research Institute  
for Animal Production in Nitra, Nitra, Slovakia.  
E-mail: savoradovaandrea1@gmail.com

Mendel University in Brno, I, Zemědělská 1665,  
Černá Pole, 61300, Brno-sever, Czech Republic.  
E-mail: savoradovaandrea1@gmail.com

University of Agriculture in Nitra,  
Institute of Plant and Environmental Sciences,  
Tr. A. Hlinka, 2, Nitra, 94901, Slovakia.  
E-mail: kornelia.petrovicova@gmail.com

Oles Honchar Dnipro National University,  
Gagarin av., 72, Dnipro, 49010, Ukraine.  
E-mail: brigad@ua.fm

Dnipro State Agrarian and Economic  
University, Serhii Efremov st., 25, Dnipro,  
49600, Ukraine. E-mail: brigad@ua.fm

**Langraf, V., Svoradová, A., Petrovičová, K., & Brygadyrenko, V. (2023). Structure of data in cell biology research. Regulatory Mechanisms in Biosystems, 14(3), 492–496. doi:10.15421/10.15421/022370**

Bioinformatics is a scientific field on the border between informatics and biology where problems in the field of biology are solved using statistical methods. Another part of it are database systems which serve to store data necessary for meta-analysis. In recent years, there has been a boom mainly thanks to enabling technologies that make it possible to obtain big data about the functioning of living cells of organisms. Bioinformatics tools are necessary to process these data and form an integral part of research in modern biological and medical sciences. Scientific research focused on molecular biology, as well as medicine, is increasingly focusing on data storage. It is understood that the correct structure of the database is important for the correct interpretation of the results of their research activities. For communication between tables in the database, it is essential to set the data type, assign Primary key and Foreign key, ensure data integrity, remove data plurality and understand the research logic. Based on these needs, we created a relational database using SQL Server 2017 and Microsoft SQL Server Management Studio 2017 (SSMS). We created the source code for programming the database and filling it with data in Structured Query Language (SQL) and T-SQL on the Microsoft platform. Of the data types, we used float for numbers with a floating decimal line, integer values were assigned an integer (int), date had a date data type, and text strings had a defined nvarchar data type. Our results bring new information in the field of bioinformatics about the creation of a database structure for data storage in cell biology research. These new insights will help big data in meta-analyses of data and applying scientific results to medical and scientific practice. The database will store data obtained in real time, which will ensure relevance in pointing out biological trends, regularities, relationships and links between cellular structures. All these aspects are very important for the spatial modeling of data and the creation of models of interactions of cell structures with use for applications in medical and biological practice.

**Keywords:** molecular biology; data quality; SQL; database; SSMS.

### Introduction

The data structure of scientific research is based on connecting entities located in tables and creating relationships between them. Such a way of connecting data forms the basis of a data model, represented in a data schema. After the data model is designed and implemented, knowledge about the data structure is part of the technical structure (Halpin & Morgan, 2008). The architecture and design of the data model solves the problems of understanding the data and the connections between them, based on a technical level. Thus, this way of solving data storage is technically driven and the analyst must understand the schemas of the entire data model, data types and data dimensions (Coronel & Morris, 2014). For the subsequent proper analysis of scientific data, it is necessary to understand the data that the database will store and present. It is also necessary to convert them into a rigid framework that we can implement using software tools. Such a plan is most often complicated by ensuring data integrity from different sources, data quality, data plurality and software communication (Bimey, 2004; Canali, 2019).

The most frequently used form of data storage is using a relational database, which is the best choice of scientific institutions for storing,

processing and analyzing data generated by research. They store structured data and use structured query language (SQL) or T-SQL to access the database (Ouanouki et al., 2017; Ramzan et al., 2019). The theory of this database was put forward by Dr. E.F. Codd in the 1930s. The data in these databases is stored in two-dimensional tables composed of values located at the intersection of rows and columns. Each line, i.e. record, has its own identifier. Each table contains data that is focused on one topic. A relational database is therefore made up of different tables connected to each other based on relationships (common data) defined by Primary and Foreign key (Raj, 2018; Dalmaris et al., 2020).

Before starting to fill the tables with data from cell biology research, it is a good idea to set the data types (data type), which determine the format of the stored data. We always choose a data type that will take up as little space as possible. It is necessary to plan what data will be stored in the column of the table so that information is not lost based on a poorly set data type. It is also necessary to prevent a situation where we would have to change the set data types in the entire data model in all columns that follow each other. Choosing the right data type and its size is the basis of a well-designed database (Hanine et al., 2015; Samanta & Chaki, 2023). Currently, it is necessary to use mathematical-statistical models for the

correct interpretation of biological, ecological and medical research. Properly designed databases with stored big data are needed to perform such analyses. They have one of the most important tasks, their algorithms and technologies make it possible to search for new scientific knowledge. Based on these facts, scientists can further determine the relationships and influences between the elements of cell biology. The use of this knowledge, which is hidden in big data, makes it possible to apply it not only in the field of scientific research but also in medicine, and subsequently be used in the treatment of various diseases (McAfee, 2012; Akoka et al., 2017). The general definition of big data is related to a large amount of structured and unstructured data collected in real time (Sagiroglu & Sinanc, 2013). We can interpret the complexity of big data into the 3V model (Volume, Variety and Velocity), which is characteristic of them. Volume represents the data size of the examined sample measured in terabytes. Variety is the variety of data sources and structures. Velocity handles the generation and collection of data in streams. Thus, machine learning and deep learning is an essential part for understanding data systems, correct interpretation of scientific results and the ability to learn from these data (Gandomi & Haider, 2015; Tiwari et al., 2018). The advantages of big data and machine learning have been successfully applied in various private (banking, insurance, customer and market information, marketing) and scientific spheres (chemistry, genetics, energy) (Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). Artificial intelligence in business gets real: Pioneering enterprises aim for AI at scale. <http://sloanreview.mit.edu/projects/artificial-intelligence-in-business-gets-real>).

Our results will contribute new information about the structure of data in the data model, data quality and integrity of data obtained from cell biology research. The knowledge of the research will thus expand the knowledge of bioinformatics in the area dealing with big data and the creation of databases.

## Materials and methods

We used SQL server 2017 and its superstructure, the Microsoft SQL Server Management Studio 2017 (SSMS) program ((RTM)-14.0.1000.169 (X64) Microsoft Corporation Express Edition (64-bit) on Windows 10 Home 10.0 [X64](Build 18362)), to program the source codes. We created the source codes for programming the database, tables and filling them with data in SQL (Structured Query Language) and T-SQL. The Microsoft platform was used.

## Results

For the needs of cell biology research, we designed a relational database consisting of 12 tables. Tables that serve as code lists are marked with the abbreviation cl (Code list) before the name of the table. These include cl\_class with stored class names, cl\_species tables storing species names, cl\_cellType contain cell types or cell divisions based on their origin, potency, etc., cl\_cell stores the names of specific cells under investigation. Dimensional tables are marked with the letter d before the table name. Single record entity data is stored here without duplication. This category includes tables d\_freezing with data on the method of freezing cells and cryoprotectants used, d\_pageRole containing data on scientists who participated in the research, d\_methodAssessment with information on cell evaluation methods and parameters that are evaluated, d\_dyes has summarized data on the dyes used for cell analyses, d\_markers contains information about markers used in cell analyses, d\_extenders contains data about diluents used in cell analyses, d\_collectionMethod with stored data about cell sampling methods. Frequency tables are marked with the letter f before the name of the table. Data are stored in these tables at repeated intervals according to the frequency of cell analyses. This includes the f\_dataEntry table with records from cell analysis.

Each column in the table has an associated data type. Columns where integer data are stored have data type integer (int), this includes columns ID, parserID, classID, speciesID. Numeric data with a floating point has a data type of float. The column with this data type is value. The date is set to data type date. Text data has the data type nvarchar, after which the numerical value for the length of the stored data is indicated. Columns

with such data type are species, class, classLj, cellType, cell, freezing, Cryoprotectants, surname, lastName, title, institution, workplace, street, town, PSC, workPosition, email, phone, typ, methodName, parameters, dyes, cellStatus, localization, marker, methodName, methodNameAssessment, methodNameCollection, extenders, cellStatus, damagedStructure, damage.

Columns where data must be entered are marked NOT NULL, those where data does not need to be stored have NULL. Columns with an assigned primary key are labeled as PRIMARY KEY and with a foreign key as FOREIGN KEY. The programmed source code for creating the database is given below:

```
BEGIN
Create database CELL_BIOLOGY
END
GO
BEGIN
Create table [CELL_BIOLOGY].[dbo].[cl_species] (
[ID] [int] NOT NULL,
[species] [nvarchar](50) NOT NULL,
PRIMARY KEY (ID))
Create table [CELL_BIOLOGY].[dbo].[cl_class] (
[ID] [int] NOT NULL,
[class] [nvarchar](20) NOT NULL,
[classLj] [nvarchar](20) NOT NULL,
PRIMARY KEY (ID))
Create table [CELL_BIOLOGY].[dbo].[cl_cellType] (
[ID] [int] NOT NULL,
[cellType] [nvarchar](50) NOT NULL)
Create table [CELL_BIOLOGY].[dbo].[cl_cell] (
[ID] [int] NOT NULL,
[cell] [nvarchar](50) NOT NULL)
Create table [CELL_BIOLOGY].[dbo].[d_freezing] (
[ID] [int] NOT NULL,
[freezing] [nvarchar](50) NULL,
[cryoprotectants] [nvarchar](50) NULL)
Create table [CELL_BIOLOGY].[dbo].[d_pageRole] (
[ID] [int] NOT NULL,
[surname] [nvarchar](50) NOT NULL,
[lastName] [nvarchar](50) NOT NULL,
[title] [nvarchar](20) NULL,
[institution] [nvarchar](20) NULL,
[workplace] [nvarchar](100) NULL,
[street] [nvarchar](50) NULL,
[town] [nvarchar](50) NULL,
[PSC] [nvarchar](20) NULL,
[workPosition] [nvarchar](50) NULL,
[email] [nvarchar](50) NOT NULL,
[phone] [nvarchar](20) NULL,
PRIMARY KEY (ID))
Create table [CELL_BIOLOGY].[dbo].[d_methodAssessment] (
[ID] [int] NOT NULL,
[cell] [nvarchar](50) NULL,
[typ] [nvarchar](50) NULL,
[methodName] [nvarchar](50) NULL,
[parameters] [nvarchar](100) NULL)
Create table [CELL_BIOLOGY].[dbo].[d_dyes] (
[ID] [int] NOT NULL,
[dyes] [nvarchar](50) NOT NULL,
[cellStatus] [nvarchar](100) NOT NULL)
Create table [CELL_BIOLOGY].[dbo].[d_markers] (
[ID] [int] NOT NULL,
[localization] [nvarchar](50) NOT NULL,
[typ] [nvarchar](50) NOT NULL,
[marker] [nvarchar](50) NOT NULL)
```

```

Create table [CELL_BIOLOGY].[dbo].[d_extenders] (
    [ID] [int] NOT NULL,
    [name] [nvarchar](50) NOT NULL,
    [species] [nvarchar](50) NULL)

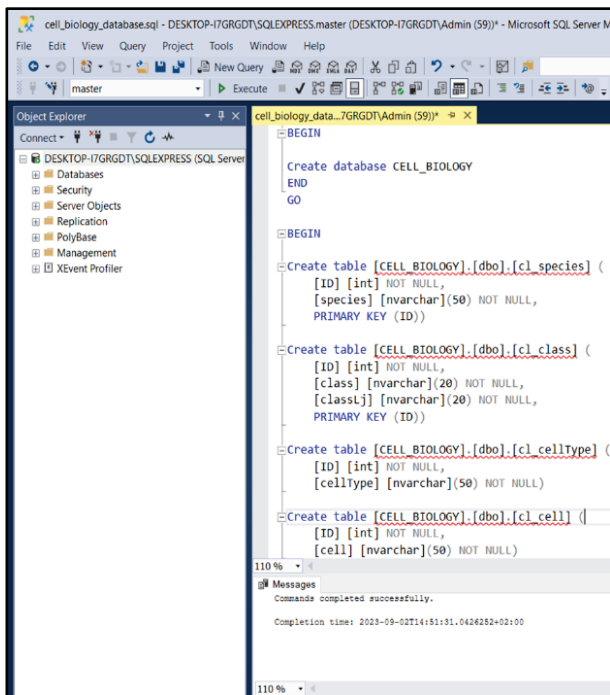
Create table [CELL_BIOLOGY].[dbo].[d_collectionMethod] (
    [ID] [int] NOT NULL,
    [species] [nvarchar](50) NULL,
    [methodName] [nvarchar](50) NULL)

Create table [CELL_BIOLOGY].[dbo].[f_dataEntry] (
    [ID] [int] NOT NULL,
    [parserID] [int] NOT NULL,
    [classID] [int] NOT NULL,
    [speciesID] [int] NOT NULL,
    [date] [date] NOT NULL,
    [methodNameAssessment] [nvarchar](50) NOT NULL,
    [parameters] [nvarchar](50) NULL,
    [methodNameCollection] [nvarchar](50) NULL,
    [freezing] [nvarchar](50) NULL,
    [cryoprotectants] [nvarchar](50) NULL,
    [cellType] [nvarchar](50) NULL,
    [cell] [nvarchar](50) NULL,
    [marker] [nvarchar](50) NULL,
    [dyes] [nvarchar](50) NULL,
    [extenders] [nvarchar](50) NULL,
    [cellStatus] [nvarchar](50) NULL,
    [damagedStructure] [nvarchar](50) NULL,
    [damage] [nvarchar](50) NULL,
    [value] [float](8) NOT NULL,
    FOREIGN KEY (parserID) REFERENCES d_pageRole(ID),
    FOREIGN KEY (classID) REFERENCES cl_class(ID),
    FOREIGN KEY (speciesID) REFERENCES cl_species(ID))

END
GO

```

After the source code has been run, the Microsoft SQL Server Management Studio program displays a message about the correct completion of the command "commands completed successfully" (Fig. 1).



**Fig. 1.** Source code with notification of successful execution of the command "commands completed successfully"

To upload 2 records to each table in the database, we have created the command source code which is given below:

```

BEGIN
INSERT INTO [CELL_BIOLOGY].[dbo].[cl_species]
VALUES (1, 'rabbit',
        (2, 'aries')

INSERT INTO [CELL_BIOLOGY].[dbo].[cl_class]
VALUES (1, 'birds', 'Aves'),
        (2, 'mammals', 'Mammalia')

INSERT INTO [CELL_BIOLOGY].[dbo].[cl_cellType]
VALUES (1, 'embryonic'),
        (2, 'sexual')

INSERT INTO [CELL_BIOLOGY].[dbo].[cl_cell]
VALUES (1, 'sperm'),
        (2, 'adipocytes')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_Freezing]
VALUES (1, 'permeable', 'DMSO'),
        (2, 'impermeable', 'Ficoll')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_methodAssessment]
VALUES (1, 'sperm', 'microscopic', 'CASA', 'motility'),
        (2, 'stem cells', 'microscopic', 'Fluorescence microscopy', 'cell
phenotype')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_dyes]
VALUES (1, '7-AAD', 'proportion of dead cells'),
        (2, 'SYBR-14', 'proportion of living cells')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_pageRole]
VALUES (1, 'Svoradová', 'Andrea', 'RNDr. (PhD)', 'VUŽV', 'Nitra
Animal Production Research Institute', 'Hlohovecká 2', 'Lužianky', '951 41',
'scientific worker', 'XXXXXX@gmail.com', '+421900000000'),
        (2, 'Langraf', 'Vladimír', 'RNDr. (PhD)', 'UKF', 'Constantine the Philo-
sopher University in Nitra', 'Tr. A. Hlinku 1', 'Nitra', '94901', 'scientific
worker', 'XXXXXX@gmail.com', '+421900000000')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_markers]
VALUES (1, 'surface markers', 'Mesenchymal', 'CD 90'),
        (2, 'Intracellular markers', 'Hematopoietic', 'Vimentin')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_Extenders]
VALUES (1, 'Cortalap®', 'rabbit'),
        (2, 'Trilady!', 'aries')

INSERT INTO [CELL_BIOLOGY].[dbo].[d_collectionMethod]
VALUES (1, 'artificial vagina', 'rabbit'),
        (2, 'dummy', 'bull')

INSERT INTO [CELL_BIOLOGY].[dbo].[f_dataEntry]
VALUES (1, 1, 2, 1, '01-05-2023', 'Fluorescence microscopy', 'cell
phenotype', '', 'permeable', 'glycerol', 'pluripotent', 'blastocyst cells', 'CD
90', 'Propidium iodide', '', 'proportion of dead cells', 'mitochondria', 'ripple
mitochondria', 10),
        (2, 1, 2, 1, '01-05-2023', 'Trypan blue', 'cell viability', 'artificial va-
gina', 'permeable', 'DMSO', 'sexual', 'sperm', ' ', '7-AAD', 'Cortalap®',
'proportion of dead cells', 'header', 'deformation of the head', 17.95)

END
GO

```

After running the source code, the Microsoft SQL Server Management Studio program displays a message about uploading records "2 rows affected" (Fig. 2).

After successfully creating the database and uploading 2 records to individual source code help tables, we can view the entire database in the Microsoft SQL Server Management Studio program in the Database section. The name of the created CELL\_BIOLOGY database will be displayed in this section. After clicking on it in the tables section, we can find the created tables. After running the command, we will see the selected table [CELL\_BIOLOGY].[dbo].[f\_dataEntry] also with naked data (Fig. 3). The command to bring up the table with the saved record is given below:

```

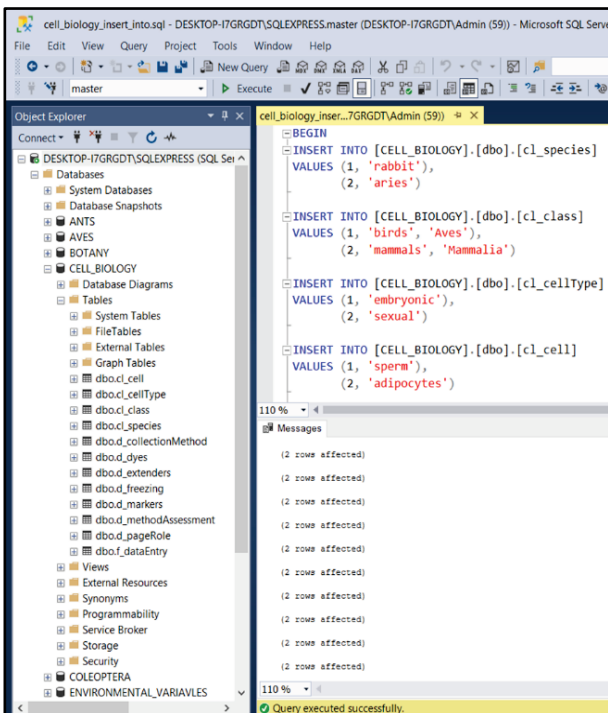
SELECT [ID]
       , [parserID]

```

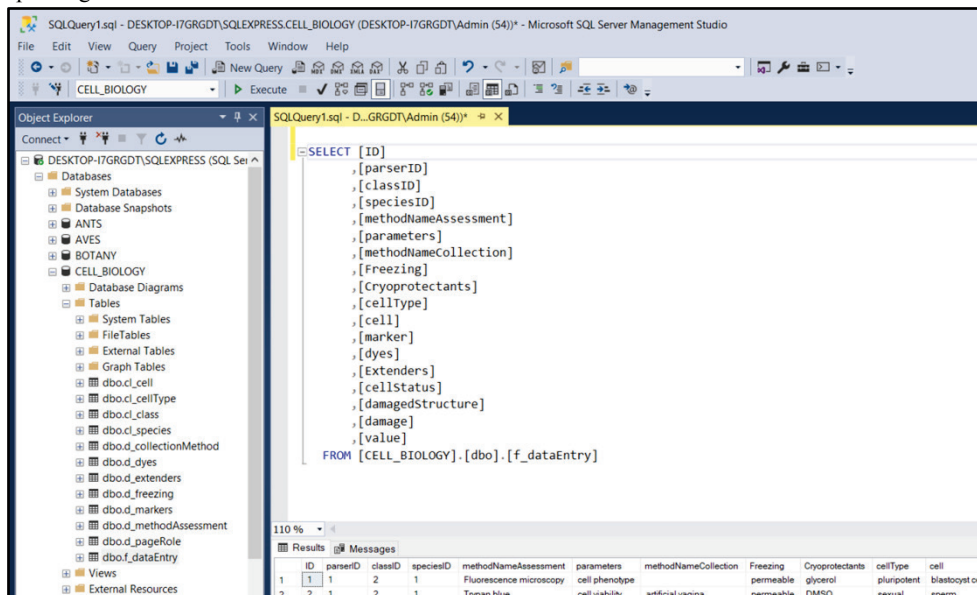
```

],[classID]
],[speciesID]
],[methodNameAssessment]
],[parameters]
],[methodNameCollection]
],[Freezing]
],[Cryoprotectants]
],[cellType]
],[cell]
],[marker]
],[dyes]
],[Extenders]
],[cellStatus]
],[damagedStructure]
],[damage]
],[value]
FROM [CELL_BIOLOGY].[dbo].[f_dataEntry]

```



**Fig. 2.** Source code with notification about uploading 2 records "2 rows affected"



**Fig. 3.** Calling the f\_dataEntry table with uploaded records

## Discussion

Modern analyses of biological data use knowledge from the field of bioinformatics focused on databases, big data with a link to meta-analyses. The collected data contain very important information about the mechanisms of the functioning of relationships and interactions of various biological structures (Ramesh & Henderson, 2018). They are mainly focused on storing data for the areas of nucleic acids, DNA (GenBank), RDA (RNA-central), proteins: Proteins Database (PDB), Universal Protein Resource (UniProt), The Human Protein Atlas, cancer (Cancer Genome Atlas (TCGA), Cancer Genome Consortium (ICGC)), for industrial bioeconomy. The Universal Protein Resource consists of three databases: UniProt Knowledge-base (UniProtKB), UniProt Reference Clusters (UniRef) and UniProt Archive (UniParc). The amount of implemented data is large and sometimes there are problems with storage and transmission. What will be proven subsequently also during data processing (Ragunath et al., 2009; Burge et al., 2013; Nielsen & Keasling, 2016; Benson et al., 2018). In the paper, we proposed the source codes for creating a new relational database structure and filling it with data for the needs of cell biology research. The need for such a database increases linearly with the increasing amount of data obtained by scientific research. Currently, such a database is not available, and its creation will solve this deficiency and contribute new knowledge to the field of bioinformatics.

For proper data storage and fast communication between tables, the database must have an appropriately organized structure and a selected data model (Gharajeh 2018; Duigou et al., 2019). In our case, we chose the E-R model designed in the Microsoft SQL Server Management Studio database program on the Microsoft platform. The SQL server uses the Structured Query Language (SQL) and its extension T-SQL, which we also used when writing our source codes.

A relational database consists of a collection of tables storing specific sets of data. These are divided into subsets of table forms. The concept of a relational database is derived from the principles of relational algebra and was elaborated by its discoverer, E. F. Codd. Most of today's operating database systems use the principles of a relational system (Batra, 2018; Dalmaris et al., 2020). Our designed database also works on the principle of a relational database. The tables contain data focused on one topic, i.e. a subset. They are connected to each other using Primary and Foreign key relationships, which are defined on specific columns.

To ensure the integrity, consistency and validity of the data in the database, it is necessary to set the correct data types. If they are incorrectly set, the data transmission during data interrogation will be reduced, which should also affect the effectiveness of the results in meta-analyses (Siegel et al., 2022). In our database, we used data type float for floating-point numbers, integer for integer values, date for date and nvarchar for text strings.

Various programs dBASE III PLUS, dBASE IV, FoxBase, FoxPro, Paradox, Clipper, WinBase 602, MS Access, MS Visual FoxPro Access, MYSQL and Microsoft SQL Server Management Studio are used to create functional databases. Some work on the Microsoft or Oracle platform. In our case, we used the Microsoft platform. Thus, an appropriately designed relational database structure interpreting the relationships and interactions between cellular structures is necessary for objective meta-analysis. Our database focused on cell biology is designed for storing Big data and quick selection of data used in meta-analysis and subsequent application of results in medical and scientific practice.

## Conclusions

The relational database designed by us for the needs of cell biology research helped to solve the storage of big data, necessary for the following meta-analyses. We can apply the results of these analyses in science, for the purposes of establishing connections between the investigated structures. Also in the field of medicine with application to the treatment of various genetic diseases. Much attention was paid to data storage mainly in the field of banking, insurance, marketing and in various private spheres. Science has lagged behind in the use of technological knowledge of big data, machine learning and database architecture. That's why we created a new database structure for storing cell research data. For fast communication, we have set appropriate data types and connection between tables using Primary Key and Foreign Key. All the above-mentioned steps of setting up the database led to ensuring the integrity of the data and to applying it in practice.

This research was supported by the grants VEGA 1/0604/20 Environmental assessment of specific habitats in the Danube Plain. KEGA No. 002UKF-4/2022 Metaanalyses in Biology and Ecology (Databases and Statistical Data Analysis).

## References

- Akoka, J., Comyn-Wattiau, I., & Laoufi, N. (2017). Research on big data – A systematic mapping study. *Computer Standards and Interfaces*, 54, 105–115.
- Batra, R. (2018). A history of SQL and relational databases. In: *SQL Primer: An accelerated introduction to basic SQL*. Apress. Pp. 183–187.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2014). GenBank. *Nucleic Acids Research*, 42, 7–32.
- Bimey, E., & Clamp, M. (2004). Biological database design and implementation. *Briefings in Bioinformatics*, 5(1), 31–38.
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., Eddy, S. R., Gardner, P. P., & Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, 41, 226–232.
- Canali, S. (2019). Evaluating evidential pluralism in epidemiology: Mechanistic evidence in exposome research. *History and Philosophy of the Life Sciences*, 41(1), 4.
- Coronel, C., & Morris, S. (2014). *Database systems: Design, implementation and management*. 11th ed. Cengage Learning, Stamford.
- Dalmaris, E., Avramidou, E. V., Xanthopoulou, A., & Aravanopoulos, F. A. (2020). Dataset of targeted metabolite analysis for five taxanes of Hellenic *Taxus baccata* L. populations. *Data*, 5(1), 22.
- De Lorenzo, V., Prather, K. L., Chen, G. Q., O'Day, E., von Kameke, C., Oyarzún, D. A., Hosta-Rigau, L., Alsafar, H., Cao, C., Ji, W., Okano, H., Roberts, R. J., Ronaghi, M., Yeung, K., Zhang, F., & Lee, S. Y. (2018). The power of synthetic biology for bioproduction, remediation and pollution control: The UN's sustainable development goals will inevitably require the application of molecular biology and biotechnology on a global scale. *EMBO Reports*, 19(4), e45658.
- Duigou, T., du Lac, M., Carbonell, P., & Faulon, J. L. (2019). RetroRules: A database of reaction rules for engineering biology. *Nucleic Acids Research*, 47(D1), 1229–1235.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Gharajeh, M. S. (2018). Biological big data analytics. *Advances in Computers*, 109, 321–355.
- Halpin, T., & Morgan, T. (2008). *Information modeling and relational databases*. Elsevier Inc.
- Hanine, M., Bendarag, A., & Boutkhoum, O. (2015). Data migration methodology from relational to NoSQL databases. *International Journal of Computer and Information Engineering*, 9(12), 2566–2570.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90, 60–68.
- Nielsen, J., & Keasling, J. D. (2016). Engineering cellular metabolism. *Cell*, 164(6), 1185–1197.
- Ouanouki, R., April, A., Abran, A., Gomez, A., & Deshamais, J. M. (2017). Toward building RDB to Hbase conversion rules. *Journal of Big Data*, 4, 10.
- Ragunath, P. K., Venkatesan, P., & Ravimohan, R. (2009). New curriculum design model for bioinformatics postgraduate program using systems biology approach. *Journal of Computer Science and Systems Biology*, 2(6), 300–305.
- Raj, P. (2018). A detailed analysis of noSQL and newSQL databases for big data analytics and distributed computing. *Advances in Computers*, 109, 1–48.
- Ramesh, S., & Henderson, A. (2018). A semantic data model: Meaning making from data structures in the SQL server. *Journal of Information Systems Engineering and Business Intelligence*, 4(2), 106–115.
- Ramzan, S., Sarwar, I., Ramzan, B., & Anwar, W. (2019). Intelligent data engineering for migration to NoSQL based secure environments. *IEEE Access*, 7, 69042–69057.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In: *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego. Pp. 42–47.
- Samanta, A., & Chaki, N. (2023). A bi-directional engine for transformation of big data between NoSQL and SQL data.
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1), 7–33.
- Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers and Industrial Engineering*, 115, 319–330.