



faculty of social  
sciences, arts  
and humanities

42/2023

Research Journal  
Studies about Languages  
pp. 52-67  
ISSN 1648-2824 (print)  
ISSN 2029-7203 (online)  
DOI 10.5755/j01.sal.1.42.33107

## LINGUISTICS / KALBOTYRA

Experience of design and development of a new open access interactive multifunctional database management system for special lexis of biology

Received 12/2022

Accepted 05/2023

**HOW TO CITE:** Stalažs, A., Šķirmante, K., Sviķe, S., Jasmonts, G., & Ziediņš, R.E. (2023). Experience of design and development of a new open access interactive multifunctional database management system for special lexis of biology. *Studies about Languages / Kalbų studijos*, 42, 52–67. <https://doi.org/10.5755/j01.sal.1.42.33107>

# Experience of design and development of a new open access interactive multifunctional database management system for special lexis of biology

Naujos atviro prieigos interaktyvios daugiafunkcinės duomenų bazių valdymo sistemos, skirtos specialiajai biologijos leksikai, dizaino ir vystymo patirtis

ARTURS STALAŽS, Institute of Horticulture, Latvia

KARINA ŠĶIRMANTE, Ventspils University of Applied Sciences, Latvia

SILGA SVIĶE, Ventspils University of Applied Sciences, Latvia

GINTS JASMONTS, Ventspils University of Applied Sciences, Latvia

ROBERTS ERVĪNS ZIEDIŅŠ, Ventspils University of Applied Sciences, Latvia

## Abstract

Terminology work requires a lot of data processing in the practical research of comparative linguistics, especially when studying linguistic and terminological changes over time. The lack of appropriate wide-usage electronic tools and databases can slow down such research. Thus, an efficient information management solution is necessary to improve the efficiency of terminology and comparative linguistic research.

Since 2021, a team of terminologists, translators, researchers, and information system developers worked together for two years, during which the open access interactive multifunctional database management system (IMDS) was designed and developed for data storage and a wide range of statistical and search options, especially for language research purposes and comparative multilingual studies in linguistics and terminology.

The information system consists of multiple modules (input, statistics, export, etc.). The data input module was successfully designed and developed and has been used effectively by researchers for entering and collecting special lexis units.

The article mainly describes the statistical data that can be obtained from the IMDS, as well as problems encountered during the development of the system. The IMDS data collection has been updated with 50,846 entries of scientific (Latin) organism names, 60,585 entries of local names of different organisms (mainly in Latvian, but also in other languages), 1427 entries of names of organism-caused diseases, 2942 entries of dictionary words and 313 entries of terms with definitions collected based on excerpts from different publications containing common names linked with their scientific names.

**KEYWORDS:** special lexis, biology, terminology, data collection, design and development of an information management system, linguistic heritage.

---

## Introduction

Many jobs performed by people from ancient times like gardening, agriculture, forestry, medicine, crafts, and floristics are dealing with plants and, therefore, with names of organisms (e.g., plants and animals). A large part of the special vocabulary of biology are the names of plants and animals that appear in specialised, educational, and leisure literature.

These names are considered an exceptional thematic group because of their enormous diversity, which is different from other thematic groups of general vocabulary. For example, in German, this thematic group, plant and animal names, could be described by looking at the work of Dornseiff (2004) in which the entire vocabulary is divided into 22 large groups. Quasthoff comments on such division of words into thematic groups and emphasises that words in one thematic group in this dictionary usually occupy about half a page, but the thematic group for plants required 81 pages, and the animal species required 34 pages (Quasthoff, 2004, pp. 201–203). The above-mentioned aspect poses difficulties for translators, who have to find the terminological names of these organisms as well as variants of the plant or animal names or synonymically used plant names in the source and target language. It also poses difficulties for journalists preparing articles on this subject, as well as for anyone whose work or hobby is related to this subject.

Although the 30 terminology dictionaries in the fields of botany and zoology (including species name indexes) of all terminology dictionaries published in Latvia between 1900 and 2007 represent 6.90% of all terminology dictionaries, ranking these fields as the fourth most frequently published (Tulkošanas un terminoloģijas centrs, 2007, p. 13), this does not make working with this part of the specialised vocabulary significantly easier. Some of these resources are still only printed editions (see also Stalažs, 2015) and have not been digitised or updated in line with changes in terminology in this field at international level. Taking into account the authors' practical translation experience in translating environmental subject texts, data from lexicographic surveys (Balode, 2012, p. 41; Sviķe, 2018), as well as terminology work in translation studies at the Faculty of Translation Studies of Ventspils University of Applied Sciences (Sviķe, 2022), it can be concluded that there is a need for an electronic resource, where it would be easy to find at least most of the necessary information about this specific group of special lexis — names of organisms, such as scientific (Latin) names of taxa, correct terminological names in Latvian and other languages, synonymic names of organisms, frequency of use of these names, and analysis of examples of the use of organism names in a time scale.

The purpose of this study is to provide insights into design and development of the new open access interactive multifunctional database management system developed for special lexis research purposes and outline the problems faced during the implementation of the projects mentioned below and propose possible solutions. This study is about a project, funded by Latvian Science Council, Smart Complex of Information Systems of Specialized Biology Lexis for the Research and Preservation of Linguistic Diversity with the aim of developing this specific system from 2021 to 2023. This paper gives an overview of the different modules and data input and retrieval options of this system to show some research possibilities of this special lexis group. As the project is not yet finalised, the article describes the state of the art and presents interim results. The practical part of the study uses a mainly descriptive method to describe the experience of developing such a database from the very beginning. It could be useful for similar studies on other special lexical groups. The theoretical part includes a brief assessment of existing similar resources and justifies the need for the database. Statistical methods and qualitative analysis of language examples are used for data analysis and characterisation of examples.

## Theoretical Background

In 2011, researchers asked the question ‘How many species are there on earth and in the ocean?’ (Mora et al., 2011). The authors of the study made mathematical calculations and indicated that around 2010 there were more than 1.2 million documented species, but, according to predictions, this could be only 14% of the real number of species. However, more than 1.2 million documented species does not mean that there is the same number of scientific names of organisms in science. In fact, the scientific names of the organisms that have been introduced so far are much more. This is because science denominates not only the names of species, but also the names of taxa of other taxonomic levels, so-called ranks. There are also many synonyms and scientific names published without considering taxonomic principles that have been governed by particular International Codes.

Currently, there have not been any calculations that would allow us to determine how many scientific names for the classification of organisms have actually been introduced by now since the Swedish naturalist Carl Linnaeus published his first work ‘*Species plantarum..*’ in 1753 (Linné, 1753), in which he introduced modern traditions in the denomination of organisms. However, it is likely that this number could be closer to 2 million different scientific names for different organism taxa. Despite the large number of synonyms for the scientific names of organisms, these names are a fairly stable group of lexemes that are important in cross-linguistic communication. Scientists from different countries communicate with each other using mainly scientific names. However, in publications, the names of organisms sometimes appear without their scientific names, so the name of the organism in the local language can be the key term in a particular publication. Considering that scientific names of organisms are of international use, and they are being used together with local names in publications in many languages, it is possible to obtain equivalents in any language. This is crucial and important not only in studying linguistic aspects, but also in terminology and translation.

There are databases that aim to compile the scientific names of the world’s flora, for example <http://www.worldfloraonline.org/> as ‘an open access web-based compendium of the world’s 400,000 species of vascular plants and mosses’ (World Flora Online, 2022), but not fauna: such an electronic database would also be needed with the local names of organisms, which would be useful for terminology research, translators, and experts in the field.

Some local names of various plants and animals in Latvian have been published in the first religious publications in Latvian and other publications since the beginning of the 1600s (e.g., Šulcs, 2020). However, because of different spellings in several older works, it is not possible to know what these specific names correspond to. Since the classification of organisms using scientific names has been introduced, it can be assumed that these scientific names allow the more correct and precise matches to be obtained in a certain language. Therefore, in most cases, the matches are understandable through the attributed scientific names. Publications which mention both scientific and local names have been used to gather information in previous studies (Stalažs, 2015; Šulcs, 2020). The first publication, where both scientific names and local Latvian names are used, was published in 1777 (Hupel, 1777). Historically, since then and in paper format alone, starting with special scientific literature and ending with articles in encyclopaedias, as well as in popular publications and the press, several thousands of works containing the names of organisms in Latvian have been published. The amount of information and data is impressive, and it cannot be processed manually, therefore, modern information technology solutions are necessary.

Although many electronic publications, as well as individual species/taxa lists of specific organisms, have become available online over the last 25 years, they do not cover the diversity of lexemes introduced earlier or those required in terminological and linguistic work. There are several online databases where it is possible to search for scientific names and names of organisms in Latvian and major contact languages, e.g., German, English, Russian. For example, the Latvian National Terminology Portal (<https://termini.gov.lv/>) collects the names of organisms and their scientific name equivalents. There are organism names not only in Latvian but also in other languages, and data are linked to information source references. Unfortunately, this database covers only a selected number of information sources that does not include a remarkable number of organisms, as well as the names of the diseases caused by particular organism species. During the creation of this database, the classification of organisms and scientific names used in older publications have not been checked for compliance with

scientific requirements. Therefore, the database reflects only the equivalents of the terms that were published at a certain historical time. Considering that a lot has changed in biological science, part of the data in the information sources used in this database are outdated or incorrect.

General online dictionaries also include the names of organisms, for example, those available on *Letonika.lv* (<https://www.letonika.lv/>), a resource of bilingual translation dictionaries, and on the consolidated dictionary portal *Tēzaurus.lv* (<https://tezaurus.lv/>). The information available on *Tēzaurus.lv* is not intended for research of historical language changes, as it does not provide information on exact data sources of specific entry units and time periods used to compile an entry. Like the Latvian National Terminology Portal database, there is also no verification of organism classification information here. Although *Tēzaurus.lv* offers the possibility to use the Latvian Language Historical Dictionary, which covers historical lexeme sources from the 16<sup>th</sup>–17<sup>th</sup> centuries, including the source references, there is no connection with the scientific names of organisms, since modern scientific classification was only introduced in the 18<sup>th</sup> century.

The largest electronic encyclopaedia of species that provides information on scientific names of plants and animals with names in local languages is *Latvijasdaba.lv* (<https://www.latvijasdaba.lv/>). This is the main electronic encyclopaedia, which covers a large number of species found in Latvia but not in other regions of the world. This encyclopaedia also does not examine changes in the classification systems of organisms. In some organism groups, a number of species do not have their names in Latvian. Although the range of species is relatively wide, it does not cover all groups of organisms, especially names of diseases caused by particular organisms. However, it should be noted that *Latvijasdaba.lv* is the most comprehensive electronic publication covering the organisms found in Latvia.

A special database dedicated to vascular plants was created at the Institute of Biology of the University of Latvia. Unfortunately, this is not a public database for the needs of specific researchers. The database covers the period of the 17<sup>th</sup> century and includes data on information sources (Šulcs, 2011, 2020). Since this database is not available to the public and does not cover all groups of organisms, but only a part of the plants (vascular plants only), it is not available for translators, gardeners, and the general public.

From a terminological point of view, the problem is that most information resources cover only a limited data or publication-specific group of organisms, rather than all taxonomic groups. Traditionally, researchers specialise in studying narrow groups of organisms, and authors, through their publications, convey to the public the names of organisms related to their interests. The practical experience of the authors of this article shows that currently there are no electronic databases dedicated for research purposes, e.g., for analysing the use of organism names over time, frequency of name usage, etc.

Other different resources such as *Periodika.lv* (<http://periodika.lv/>) or language corpora can also be used for translation and research, but in practice it has been found that the performance of the available corpus of modern Latvian texts is not inadequate for translating and studying texts in the field of biology (see Jasmonts et al., 2022). Translators render texts on diverse subjects, so they often need different names of organisms, not only the organism groups that have been interesting in recent decades, with their names covered by internet resources. However, most publications were printed books: botanical dictionaries [e.g., *Botaniskā vārdnīca* by Galenieks (1950)], which are not digitised), textbooks (e.g., first botany textbooks), plant indexes, and a monograph *Latviešu valodas augu nosaukumi* (Ēdelmane & Ozola, 2003) which collects a wide range of plant names, including regional plant names, and is not completely digitised. These printed books contain the largest range of taxon names of organisms. Since internet resources cover only a short period of time, generally 30 years only, they alone are no longer sufficient for linguistic research and terminological needs. There are two aspects to consider. If all the historical material is not digitised, it is impossible to trace the changes in language traditions over time. To enable such studies in the future, the IMDS is developed considering the gaps described above. Focussing on terminological issues, it has previously been concluded that it is important to follow the historical usage of organism names when making decisions about changing particular names (Stalažs, 2015). To ensure this, extensive information resources are needed, (1) including as many taxa of organisms as possible; (2) covering the longest possible time period from the past to the present; and (3) linking the use of local names of organisms in the past with the modern taxonomic changes of organisms.

Modern technical capabilities allow to obtain new information that changes the perception of the classification of organisms. These changes affect specific groups of organisms that have been studied in the past or groups whose earlier names even not correlate at all with actual herbarium materials. Changes in organism classification have a significant impact on terminological and linguistic aspects as well, since the named realities change (Stalažs, 2022).

There are some specific problems. For example, it is important for both linguists and terminologists to specify whether the scientific names *Arvicola amphibius* and *Arvicola terrestris* refer to one species or two species. An answer would enable us to determine how many local names are needed to denote realia. To develop effective new terminology dictionaries that recommend appropriate terms to translators, it is essential to keep up with advancements in the biological sciences and methodically account for terminology in the relevant field. Any taxonomic changes must consider whether new local names need to be created to prevent multiple names from being given to a single species or other taxon, when in reality there is only one species or other taxon with several scientific names assigned.

These issues are essential not only for terminology, but also for technical solutions during the development of the database when comparing publications issued at different time periods. There are taxa of organisms that were named with different scientific names or alternate spellings of names (e.g., *Pinus sylvestris* and *Pinus silvestris*) in different historical periods. If terminologists are not aware of the variants, a faulty decision could be made. Therefore, it is important to provide digital tools to keep track of historical changes in the classification of organisms, representing these changes according to international codes of taxonomy and the latest research results in biology, while ensuring that it is technically possible to automatically link future changes to historical usage of local names. The same applies to technical solutions that would allow distinguishing homonymous organism names between different groups of organisms (e.g., *Pieris* as plants and *Pieris* as animals), as well as homonymous local names (e.g., Latvian *kaulenes* as plants and *kaulenes* as animals), etc. This is a significant aspect, which would provide further assistance to translators, for example, if they have to translate a text in the subject domain where such information cannot be determined without context.

To collect data related to the biology field, multiple approaches can be used, for example, machine learning (ML) algorithms for optical character recognition (OCR), integrations with existing information systems, and manual data entry in a file, database or system.

In recent years, machine learning algorithms have been developed in various fields, including linguistics (Kaur & Garg, 2021). It should be noted that ML algorithms cannot provide 100% precision and manual testing is necessary. Also, a large data set is necessary for ML algorithms to learn and it is not always possible to collect enough data for learning purposes, especially in languages with a relatively small number of users, such as Latvian (Jasmonts et al., 2022). Integration with other systems is possible, but this specific set of biology data may not be sufficient for neural network training (Darģis et al., 2020). In addition, not all information systems permit access via application programming interface to collect data dynamically and more effectively. And the third option, manual entering, is time consuming. However, it can be reduced by developing an information system that is intuitive and offers multiple autocompletion possibilities.

In this research, authors aim to achieve linguistic and terminological needs with new information technology solutions, especially for data collection and visualisation. First technical solutions that would allow creating a dynamic open access database system dedicated to the names of organisms will be examined. It would also ensure the following options: (1) the collection of data on as many taxa of organisms as possible; (2) tracking the use of local names of organisms during different times; and (3) linking the organism names used in specific time periods with taxonomic changes by considering the latest discoveries in biology.

### Data Coverage

#### Data and Methods

Lexemes (names of organisms) from different publications entered into the IMDS created during this study were used to build and test the database system and for further analysis. Also, this data set was used to create technical solutions when: (1) homonymous names of organisms are used in the publications, as well as in the classification of organisms, and (2) different spelling variants of the organism names are used. In addition, technical solutions are implemented for automated linking of recurring

data, so that data linking does not have to be done manually. This is a preliminary report that examines the results as of 16.12.2022.

To create the database system, various types of publications, both paper and electronic, were used for data collection. Excerpts are gathered from these publications by entering them into the database system of local names of organisms that are linked to the particular scientific names. All publications are divided into six groups according to their types (**Table 1**). During the development of IMDS, publications in Latvian is primarily used, but for testing purposes, selected publications in other languages are used as well. Regardless of the base language of the publication, the local names of all languages that are clearly linked to a specific scientific name are entered into the IMDS. If synonyms are used in the publications, each name is entered separately to allow statistical data processing. If there are spelling variants in the publications, such lexemes are also considered. If the names of organisms are used in inflections in the original publications, in the IMDS they are entered in the nominative form. All lexemes are linked to publication data and scientific names in their original spelling.

**Table 1.** Coverage of publications by type of publications\*

Type	Number of monographs, books	Number of papers	Total	Distribution
Scientific publications	80	1070	1150	26.10%
Educational and other special literature (text books, field guides, etc.)	175	5	180	4.09%
Reference literature (dictionaries, encyclopaedias, etc.)	96	1968	2064	46.85%
Popular and popular scientific literature	126	699	825	18.72%
Media (e.g., newspapers, press releases, etc.)	68	44	112	2.54%
Legislative acts (laws, regulations, etc.)	1	74	75	1.70%
Total	546	3860	4406	100.00%

\*Data as of 16.12.2022

Both our experience and the problems identified in the beginning of this article and the problems faced during the implementation of the project are considered in order to identify and implement the most effective technical IT solutions in the creation of the IMDS. The data presented in **Table 1** is only mid-project data, which is subject to change by the end of the project (December 2023). The authors intend to produce a separate scientific publication dedicated to the publications used for collecting the data in the IMDS.

### **New Solutions and Innovations to be Offered**

Within the framework of this project, it is planned to cover as many publications as possible, published starting from the beginning of the 20<sup>th</sup> century, in order to create the first and the largest database system of the Latvian special biology lexicon. The planned solutions: (1) possibilities of automated name usage statistics of entry units; (2) possibilities of tracking historical changes of entry units with a time scale; (3) linking the names of organisms with a modern classification of organisms in accordance with the International Codes of Nomenclature; (4) automated identification of homonymous terms; (5) automated linking of organism names in local languages with the scientific names of organisms.

### **Description of the Technical Solutions of the IMDS**

Before the development stage of the information system, requirements were defined. Thus, the system should: (1) be open access with an impressive amount of linguistic and terminological heritage of many languages; (2) facilitate comparative studies in linguistics and terminology; (3) be available online without additional software



installation; (4) be scalable and with the possibility to effectively add new IMDS modules. A team of linguists, terminologists, and programmers are working to design and developing the IMDS with data storage and a wide range of statistical or search options designed especially for language research purposes and comparative multilingual studies in linguistics and terminology. At the end of the project, more than 4000 referenced information sources are planned to be included in the IMDS, and at least 600,000 linkages of local names of different organisms in different languages are planned to be collected.

The developed IMDS consists of multiple modules which are mutually connected (see Fig. 1):

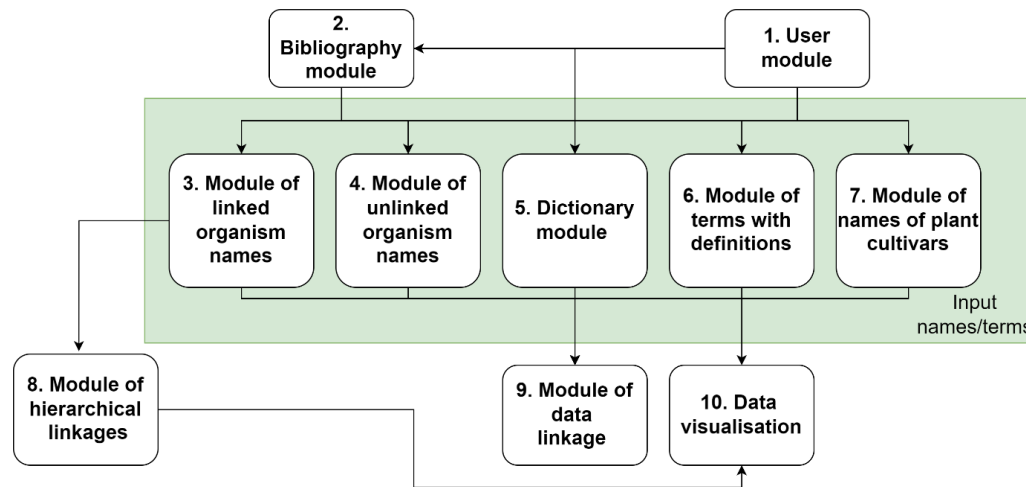


Fig. 1. Modules of the IMDS

- 1 User module. All data entry is stored with information about the user who entered the data. This module consists of user registration, authentication, and authorisation.
- 2 Bibliography module. All lexeme information is linked to the bibliographic information. This module provides the ability to store all publication data: monographs, journals/proceedings, and papers with linked additional information, for example, publisher, ISBN, alternative title, place, Digital Object Identifier (DOI), author/authors, etc. In this case, it was important: (1) to link the bibliographic source correctly with the specific excerpted unit; (2) to organise the collection of such publication data, which will allow further provision of an appropriate reference to the source of information in the database. This reference is formatted as a Harvard-style bibliographic citation.
- 3 Module of linked organism names. In this module, three types of organism names are entered: (1) scientific name, (2) local name, and (3) names of diseases caused by organisms (since the names of organisms can also be the names of diseases). In this module, the linkage between the organism name and bibliography unit is stored with additional information, for example, language mark, page number on the source bibliography, system user data, date and time, user comments. In addition, the linkage between the root element and the names of other organisms in the same linkage group is carried out in this module.
- 4 Module of unlinked terms and special lexis units. In this module, the linkage between a specific organism name (not linked to other languages) and bibliography unit is stored with additional information, for example, language mark, page number on the source bibliography, system user data, date and time, user comments. There are no linkages between other names of organisms.
- 5 Dictionary module. It is for lexemes of different languages that are not related to the scientific name of organisms and where entries are linked to specific bibliography units with additional information, for example, part of speech, gender, number, language mark, page number on the bibliographic source, system user data, date and time of the entry, user comments. Dictionary entries are linked if there is a terminological or linguistic connection between them. This module also provides links to synonyms of the same language, if the scientific name of the organisms is not used in the original publication used for data collection in the database.

- 6 Module of terms and definitions. In this module, terms with their corresponding definitions are linked to specific bibliography units, for example, identified subdomains for term and for definition, related scientific term (e.g., scientific designation in Greek or Latin) and umbrella term, language mark, page number on the source bibliography, system user data, date and time of the entry, user comments, etc. The entered terms are stored with linkage to term equivalents in other languages and their corresponding definitions.
- 7 Module of names of plant cultivars. In this module, cultivar names are linked to bibliography units with additional information, such as: country of origin, cultivar breeder's rights owner, breeder's rights protection date, system user data, language, related organism taxon name, time and date of entry, user comments, cultivar's description, species epithet, group affiliation, etc.
- 8 Module of hierarchical linkages. This module consists of organism names that are linked in a hierarchical tree with specified root element (higher taxa), for example, *Plantae*, *Animalia*, etc. In this hierarchical tree, information about scientific names of organisms, their taxon level, and linkages to parent name and children names are described. This module also includes information about the linked local organism names accepted in the IMDS. This module is integrated with Module 3 and data entering is managed by using autocomplete options because organism names are mostly entered by using Module 3 whereas this module provides the creation of hierarchical linkage between the names of organisms. This module will provide hierarchical linkage with the correct scientific names of organisms according to modern classification of organisms.
- 9 Module of data linkage. In this module, the link of scientific names entered in Module 8 with the excerpt from different publications will be ensured and controlled. This will ensure that, regardless of spelling differences, scientific names will be linked to the appropriate and correctly spelt names.
- 10 Data visualisation. This is a module for information filtering and retrieving from the databases of the developed IMDS. Results are reproduced in multiple ways, for example, in graphs, plots of time series, etc.

## Results and Discussion

### Data Coverage

On 16.12.2022, the IMDS databases contained 4406 bibliographical references and 283,618 linkages between records and bibliographical data. The IMDS database contains 50,846 scientific names (including spelling variants as used in original publications), as well as 60,585 local names in Latvian and in several other languages, and 1427 names of organism-caused diseases (Fig. 2). At this stage of the project, the main focus is on Module 3 (linked organism names), while the other modules are under development or already as finished prototypes. For testing purposes, the following data have been entered in these modules: Module 4—1712 scientific names of organisms and 204 other terms not linked to any language; Module 5—2942 terms linked among different languages including Latvian, German, English, Russian, Latin, French, Spanish; Castilian, Danish, Swedish, Polish, Greek, Norwegian, and Portuguese; Module 6—313 terms with a total of 329 definitions.

### Problems and Solutions of Automated Data Linkage

The many synonyms of scientific names of organisms that have been used in different time periods present one of the main problems. However, it was found that even greater problems were found to be caused by different spelling variants, which sometimes exceed the number of synonyms. These problems are essential when trying to develop IT solutions for linking the scientific names used in the publications as easily as possible with the modern classification of organisms in accordance with the International Codes of Nomenclature. The more spelling variants there are, the more difficult it is to ensure that an automated system works and manual control is required. It is especially necessary to prevent possible errors.

Firstly, all names are linked to their scientific name. Considering that these names are entered in the IMDS as they are in the original publications and with the authors of the names (if any), technically it is possible to filter the parts of the names (ignoring the authors of the names). However, if the original names are misspelt, each variant creates technical problems. Further problems arise from homonymous names that can be distinguished either by the author of the scientific name or by the local name used. This requires manual monitoring of name linkages. Sometimes the data is misspelt at the time of data entry and the data entry specialist cannot add a comment that would help to understand the applicability of particular names.



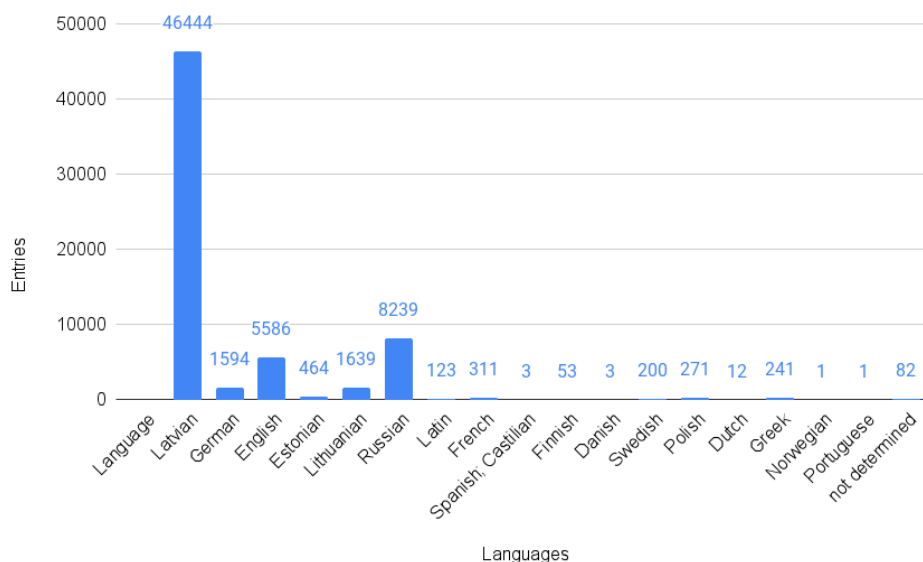


Fig. 2. Distribution of entries by languages in Modules No. 3–6

Table 2. Examples of data entries with an identical or similar scientific name

Scientific name	Latvian name (binding names in bold)
<i>Pinus silvestris</i>	<i>meža priede</i>
<i>Pinus silvestris</i>	<b><i>parastā priede</i></b>
<i>Pinus silvestris</i>	<i>priede</i>
<i>Pinus silvestris</i> L.	<i>parastā egle</i>
<i>Pinus silvestris</i> L.	<b><i>parastā priede</i></b>
<i>Pinus silvestris</i> L.	<i>priede</i>
<i>Pinus sylvestris</i>	<b><i>parastā priede</i></b>
<i>Pinus sylvestris</i>	<i>priede</i>
<i>Pinus sylvestris</i> L.	<i>priede</i>
<i>Pinus sylvestris</i> L.	<b><i>parastā priede</i></b>
<i>Pinus sylvestris</i> L. (1753)	<b><i>parastā priede</i></b>
<i>Alnus glutinosa</i>	<b><i>melnalksnis</i></b>
<i>Alnus glutinosa</i>	<i>alksnis</i>
<i>Alnus glutinosa</i> (L.) Gaertn.	<b><i>melnalksnis</i></b>
<i>Alnus glutinosa</i> Gaertn.	<b><i>melnalksnis</i></b>
<i>Alnus glutinosa</i> Gaertn.	<i>alksnis</i>
<i>Helix pomatia</i>	<i>gliemezis</i>
<i>Helix pomatia</i>	<i>īstais vīngliemezis</i>
<i>Helix pomatia</i>	<b><i>parka vīngliemezis</i></b>
<i>Helix pomatia</i>	<b><i>parku vīngliemezis</i></b>
<i>Helix pomatia</i>	<i>vīngliemezis</i>
<i>Helix pomatia</i> L.	<b><i>parka vīngliemezis</i></b>
<i>Helix pomatia</i> L.	<b><i>parku vīngliemezis</i></b>
<i>Helix pomatia</i> L.	<i>vīngliemezis</i>
<i>Helix pomatia</i> Linnaeus, 1758	<b><i>parka vīngliemezis</i></b>
<i>Helix pomatia</i> Linnaeus, 1758	<b><i>parku vīngliemezis</i></b>

Several categories can be used to classify typical problematic cases: (1) scientific names that are originally written identically (excluding authors of the names); (2) the variety of spelling variants of the same scientific name; (3) scientific names, the first word of which is abbreviated; (4) significant differences in the spelling of scientific names, as well as synonyms; and (5) homonyms of scientific names, distinguished by designation of the author of the scientific name or local name. Homonymy is also observed in local names, but if there are no problems with the recognition of the scientific name, then homonymy of local names does not cause problems in the automated data linkage.

1 To avoid introducing errors, primary name linkage is done manually, but IMDS automatically suggests identical scientific names. This can be done easily if the scientific names (excluding the designation of the author of the names) are identical or differ by one letter. Table 2 shows examples in which the initial part of the scientific name is identical or differs only by one character. Since the name of the authors is part of the scientific names entered in the IMDS, the different spellings of the following name of the authors are also indicated in the examples (Tables 2–5). This is important because it is also necessary to provide technical possibilities for distinguishing homonyms separated by the names of the authors and to ensure that IMDS automatically distinguishes the

**Table 3.** Examples of data entry with spelling variations on the side of the scientific and Latvian name of one species. The correct spelling of the scientific name of the species is *Juglans mandshurica*

Scientific name	Latvian name
<i>Juglans mandschurica</i>	<i>Mandžūrijas riekstkoks</i>
<i>Juglans mandshurica</i>	<i>Mandžūrijas valrieksts</i>
<i>Juglans mandshurica</i>	<i>Mandžūrijas riekstkoks</i>
<i>Juglans mandshurica</i> Max.	<i>Mandžūrijas riekstkoks</i>
<i>Juglans mandshurica</i> Max.	<i>Mandžūrijas valrieksts</i>
<i>Juglans mandshurica</i> Maxim.	<i>Mandžūrijas riekstkoks</i>
<i>Juglans mandshurica</i> Maxim.	<i>Mandžūrijas valrieksts</i>
<i>Juglans mandshurica</i> Maxim. (1857)	<i>Mandžūrijas riekstkoks</i>
<i>Juglans manshurica</i>	<i>Mančūrijas riekstkoks</i>
<i>Juglans manshurica</i> Max.	<i>Mančūrijas valrieksts</i>
<i>Juglans manshurica</i> Max.	<i>Mandžūrijas riekstkoks</i>
<i>Juglans manshurica</i> Max.	<i>Mandžūrijas valrieksts</i>
<i>Juglans manshurica</i> Maxim.	<i>Mandžūrijas riekstkoks</i>

**Table 4.** Examples of data entries when scientific names (also abbreviated) match a local name, allowing automatic linking of records by local name

Scientific name	Binding local name (in bold, if several names)
<i>A. calamus</i>	<b>smaržīgā kalme</b>
<i>Acorus calamus</i>	<b>smaržīgā kalme</b>
<i>Acorus calamus</i> L.	kalme
<i>Acorus calamus</i> L.	<b>smaržīgā kalme</b>
<i>Acorus calamus</i> L. (1753)	<b>smaržīgā kalme</b>
<i>A. crecca</i> L.	<b>krīklis</b>
<i>Anas crecca</i>	<b>krīklis</b>
<i>Anas crecca</i>	parastais krīklis
<i>Anas crecca</i> L.	<b>krīklis</b>
<i>Anas crecca</i> Linnaeus, 1758	<b>krīklis</b>
<i>Rh. caucasicum</i> Pall.	<i>Kaukāza rododendrs</i>
<i>Rhododendron caucasicum</i> Pall.	<i>Kaukāza rododendrs</i>
<i>Rh. maximum</i> L.	<i>lielais rododendrs</i>
<i>Rhododendron maximum</i>	<i>lielais rododendrs</i>
<i>Rhododendron maximum</i> L.	<i>lielais rododendrs</i>
<i>Rh. viscosum</i> (L.) Torr.	<i>lipīgais rododendrs</i>
<i>Rhododendron viscosum</i>	<i>lipīgais rododendrs</i>
<i>Rhododendron viscosum</i> (L.) Torr.	<i>lipīgais rododendrs</i>
<i>Rhododendron viscosum</i> Torr.	<i>lipīgais rododendrs</i>

scientific parts of the names (without the designation of the author's name). For example: *Pieris D. Don* as plants and *Pieris Schrank* as animals.

- In the original publications, the number of variants of the spelling of certain scientific names can be observed, which can significantly complicate the automatic finding of the names in the system. In this case, the local name equivalent is required. However, the spelling of equivalents of local organism names may also vary (Table 3). If the spelling of local names does not match, it can be problematic to automatically find connections between the names corresponding to the same taxon (usually to the same species). In such cases, the greatest possible number of bibliographic sources is necessary.
  - In some publications, the generic epithet of scientific names is neither written in full nor explained elsewhere in the text, which can make automatic recognition of names difficult. There is an option to automatically select potential equivalents in cases where the given local name equivalent matches the equivalents of fully written scientific names (Table 4). However, even these situations may require a one-time manual check to avoid potential mistakes. In cases where local name equivalents do not match, there may be problems in correctly deciphering the attribution of names. To be able to find binding names, it is necessary to collect a large amount of terminological data.
- Homonymy on the local name side can cause true problems, and in these cases linking should be done manually (Table 5). It is enough to do it once, though, as the next identical records will link automatically.
- The examples shown in Table 3 involve a mixed variation in spelling, both on the side of scientific and local names. If the spelling of scientific names varies significantly, which also applies to synonyms, then only the local language name

**Table 5.** Examples of data entry when the local name given as equivalent to the abbreviated scientific name is homonymous and used as a local name for different taxa

Scientific name	Homonymous local name used in some publications (in bold, if several names)
<i>B. verrucosa</i>	<i>bērzis</i>
<i>Betula</i>	<i>bērzis</i>
<i>Betula alba</i>	<i>bērzis</i>
<i>Betula L.</i>	<i>bērzis</i>
<i>Betula pendula</i>	<i>bērzis</i>
<i>Betula pendula</i> Roth	<i>bērzis</i>
<i>Betula platyphylla</i>	<i>bērzis</i>
<i>Betula pubescens</i>	<i>bērzis</i>
<i>Betula sp.</i>	<i>bērzis</i>
<i>Betula spp.</i>	<i>bērzis</i>
<i>Betula verrucosa</i>	<i>bērzis</i>
<i>Betula youngii</i>	<i>bērzis</i>
<i>Rh. luteum L.</i>	<b><i>dzeltenais rododendrs</i></b>
<i>Rh. luteum</i> Sweet	<b><i>dzeltenais rododendrs</i></b>
<i>Rhododendron aureum</i> Georgi	<b><i>dzeltenais rododendrs</i></b>
<i>Rhododendron luteum</i>	<i>dzeltenā acālija</i>
<i>Rhododendron luteum</i>	<i>dzeltenais rododendrons</i>
<i>Rhododendron luteum</i>	<b><i>dzeltenais rododendrs</i></b>
<i>Rhododendron luteum</i> Sw.	<i>dzeltenais rododendrons</i>
<i>Rhododendron luteum</i> Sw.	<b><i>dzeltenais rododendrs</i></b>
<i>Rhododendron luteum</i> Sweet	<b><i>dzeltenais rododendrs</i></b>
<i>Rhododendron luteum</i> Sweet var. <i>luteum</i>	<b><i>dzeltenais rododendrs</i></b>

**Table 6.** Examples of data entries where the local name helps to link different variants of scientific names that are equivalent to one taxon

Scientific names (synonyms, spelling variants; correct name in bold)	Binding local name
<i>Cerasus vulgaris</i> <i>Cerasus vulgaris</i> var. <i>frutescens</i> <i>Cerasus</i> × <i>vulgaris</i> <i>Prunus</i> × <i>gondouinii</i> <b><i>Prunus cerasus</i></b> <i>Prunus vulgaris</i>	<i>skābais ķīrsis</i>
<b><i>Euonymus</i></b> <i>Evonymus</i>	<i>segliņi</i>
<i>Hyponomeuta malinella</i> <i>Hyponomeuta malinellus</i> <i>Yponomeuta malinella</i> <b><i>Yponomeuta malinellus</i></b>	<i>ābeļu tīklkode</i>

should be used for automatic name linking (Table 6). In this situation, manual control and verification is mandatory, as local names can also be homonyms. Local names can also help to find correct equivalents for records where significant typographical errors have been made in the scientific name. These errors can be so significant that they prevent automatic finding of equivalents by scientific name alone.

- 5 Homonymous scientific names can be found within a group of organisms or between different groups of organisms (Table 7). These names are indistinguishable when standing alone without additional information. The author of the scientific name or the local name can be used to distinguish the names. If such names are within the same group of organisms, one of these names eventually becomes a synonym. However, homonyms may have been used in different periods of time in the literature, so a manual check of the names is required to ensure that there are no erroneous linkages of the IMDS database records.

Since manual linking has been done once, the homonymy of local names (Table 8) is no longer a problem if new records repeat the same scientific and local name combinations as previously registered. However, if a scientific name with a different spelling is added, the equivalence ratio of the new name combination must be manually checked considering the new combination only.

**Table 7.** Some examples of homonymy of scientific names between different organism groups

Scientific name	Organism group, Latvian equivalents and synonyms (if any)
<i>Arenaria</i>	plants — <i>smiltenītes, arenārijas</i> ; animals/birds — <i>akmeņtārtiņi</i>
<i>Morus</i>	plants — <i>zīdkoki</i> ; animals/birds
<i>Pieris</i>	plants — <i>kalnērikas, pieras, piēras, pīeras, pīeres, pjēres</i> ; animals, butterflies — <i>balteņi</i>
<i>Prunella</i>	plants — <i>brūngalvītes, zilgalvītes</i> ; animals/birds — <i>peļkājītes</i>
<i>Trichia</i>	fungi — <i>trihijas</i> ; animals, as synonyms/snails — <i>matiņgliemeži</i> , and butterflies

**Table 8.** Examples of homonymy of local names of organisms, which may make it difficult to perform terminological work when developing IT solutions

Scientific name	Homonymous Latvian name		Organism group
	in some publications	correct name (if yes)	
<i>Bellis</i> <i>Coccinella</i> <i>Coccinellidae</i>	<i>mārītes</i> <i>mārītes</i> <i>mārītes</i>	<i>mārpuķītes</i> also: <i>mārīšu ģints</i> <i>mārīšu dzimta</i>	plants animals (insects) animals (insects)
<i>Empetrum</i> <i>Tricholomopsis</i>	<i>vīstenes</i> <i>vīstenes</i>		plants fungi
<i>Hemipholiota</i> <i>Pholiota</i> <i>Thysanura</i> (now deprecated name)	<i>zvēņenes</i> <i>zvēņenes</i> <i>zvēņenes</i>		fungi fungi animals (insects)
<i>Festuca</i> <i>Calocybe</i>	<i>auzenes</i> <i>auzenes</i>		plants fungi

### Practical Significance and Novelties of the IMDS

Before the launch of the IMDS database project, the need for such a resource was expressed and supported by experts in the field and by students and translators specialising in translation of texts related to the field of biology. On the other hand, while preparing to implement the project, we received feedback from reviewers on the usefulness of the project and on the expected difference of the IMDS from existing solutions. In the Theoretical Background chapter, we already mentioned some online resources, listing the shortcomings of these resources and the fact that some of them are not open access resources or do not include special lexis units in Latvian. Within the framework of this project, we focus on online resources that include both the scientific names of organisms and their Latvian equivalents. After briefly describing the resources that are already available, we found that they have shortcomings, which are listed here, and we offer new solutions:

- 1 Limited terminological and data source coverage—when developing the IMDS, we anticipated that the terminological coverage should be broad and should cover: (a) all groups of organisms, not just a few specific ones; (b) as many sources of information as possible during the project implementation period; (c) the widest possible time period for collection with methodology developed within this project. One of the main focusses is also the referencing of data sources, which also provides statistical possibilities in studies of terminological culture across various time periods.
- 2 Wide range of possibilities, including open access and personalised use of the database system—the creation of an open access database system that would allow users to participate in adding data themselves,

as well as provide opportunities for wider use of data, was defined as a mandatory requirement. The IMDS is designed to provide terminological information available to a diverse range of international interest networks, not only to those interested in Latvia. All of the given options are crafted to make the IMDS useful information to specialists in various fields who use the names of organisms on a daily basis, e.g., biologists, farmers, foresters, doctors, translators, linguists, terminologists, etc.

- 3 Broad cross-linguistic coverage and linkages — one of the important disadvantages of many sources of information is the limited variety of the languages covered. Many sources offer single-language equivalents or a limited number of languages covered. Consequently, many organism taxa remain incorporated if only the literature published in Latvian is considered. The IMDS is built to be able to link the terms from different languages in one place, providing an unlimited number of languages and reference source accessions.

Here is an example from the data collected in the IMDS in the first half of the project, when scientific names related to the plant genus *Pinus* were selected as examples. On 16 December 2022, the IMDS already covered 322 scientific names related to this plant genus, which included different spellings of the names. These names were associated with more than 300 local names in seven languages (Fig. 3), covered by more than 444 citation links to more than 275 publications published between 1926 and 2022. The three main local names of the *Pinus* species with the most linkages are *parastā priede* (in Latvian, 112 linkages), *melnā priede* (in Latvian, 47 linkages), and *kalnu priede* (in Latvian, 43 linkages). These seven languages covered are just the beginning, as the IMDS has the ability to add an unlimited number of languages based on the standard ISO 639-1, specifically covering European languages. IMDS provides technical support for languages using Latin, Greek, and Cyrillic alphabets.

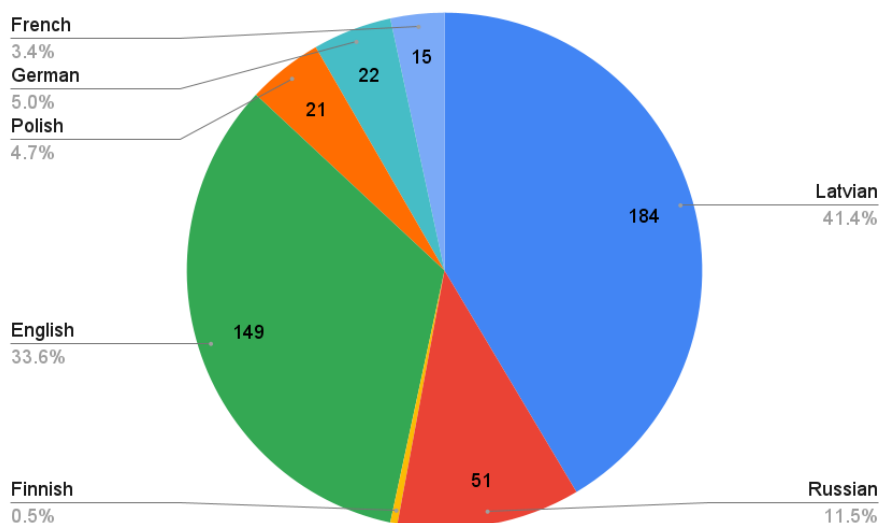
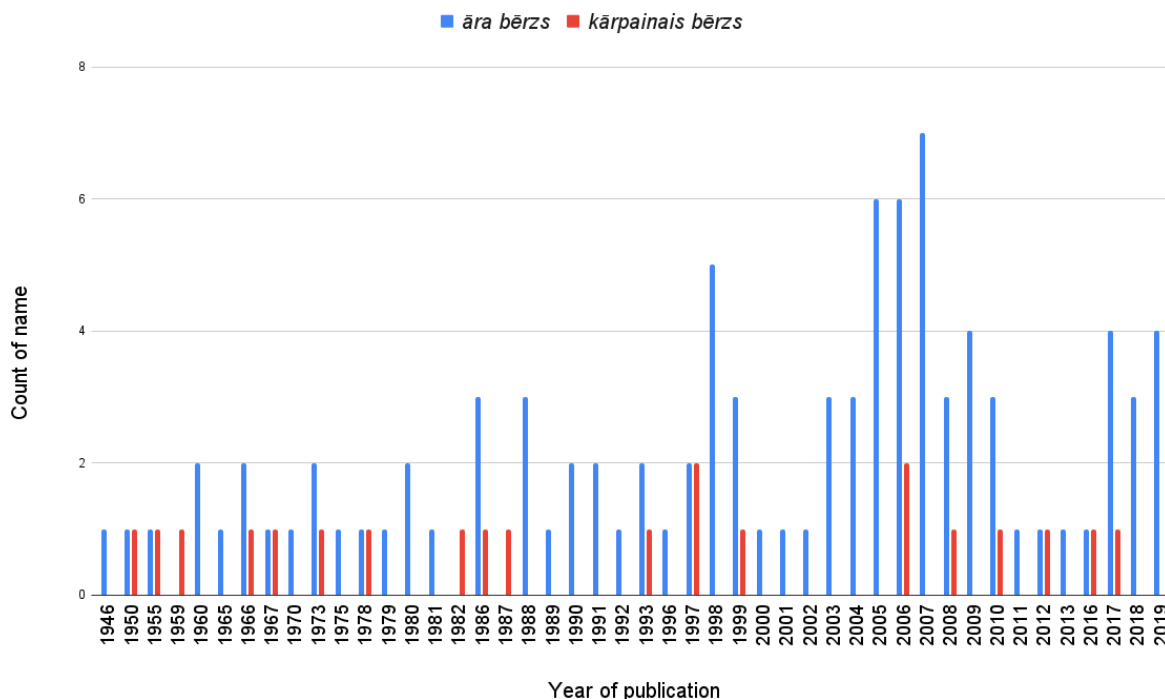


Fig. 3. Distribution of language linkages with local plant names related to taxa of the plant genus *Pinus*

- 4 Dynamic features of the IMDS, as well as association with the most modern classifications of organisms according to the International Codes of Nomenclature — since discoveries in science bring new knowledge that makes it possible to clarify the classification systems of organisms, it is necessary that the terminological work is dynamic and keeps up with such discoveries in science. This is a key condition set during the creation of the IMDS. Therefore, the IMDS provides a technical option to link outdated historical naming systems with modern classifications in accordance with the International Codes of Nomenclature. Technically, solutions for the problems listed above and will be described in future scientific publications by the authors.
- 5 Preserving linguistic heritage to track changes in the development of the terminological culture of different languages, since the oldest terms and organism name variants can serve as a basis for the formation of new terms and names of organisms. Therefore, the historical material is also practical significance.

- 6 Timeline options for statistics and historical usage of the language unit—to keep terminology and language science up-to-date, it is necessary to use all the technical possibilities of information technology to provide statistics and related automated research options. These features are currently under development and will be described in separate publications by the authors. Below is an example from the records already included in the IMDS database, which refer to Latvian names of the plant species *Betula pendula*, *āra bērzs* and *kārpainais bērzs* (Fig. 4).



**Fig. 4.** Historical usage of the linked parallelly used Latvian names, *āra bērzs* and *kārpainais bērzs*, of the plant species *Betula pendula* (IMDS records on 16.12.2022)

## Brief Conclusions and Prospects for the Future

Designing of the IMDS is a challenging task to ensure effective performance in all IMDS data collection, processing, and retrieval operations especially because of the data linkages in multiple directions: for example, the linkages of organism names to bibliography units and linkage to other organism names, hierarchical linkages according to the taxonomy structure, etc. During the development of the IMDS, various technical problems have been identified, which need to be solved in order to ensure automated linking of organism names recorded in the IMDS database. However, it is evident that partial manual control and monitoring will always be necessary to eliminate the probability of errors. The first solutions have already been developed, and they will be fully completed by the end of the project in December 2023. The current solutions provide wide language coverage according to ISO 639-1, and an unlimited number of source references. The IMDS is an attempt to create a large multilingual linguistic heritage database dedicated to organism names and intended especially for research. The IMDS is intended as an international database system; therefore, we invite colleagues from other countries to join us in pursuit of solutions and new opportunities in the future.

**Acknowledgements.** This research has been funded by the Latvian Council of Science. The project ‘Smart complex of information systems of specialized biology lexis for the research and preservation of linguistic diversity’ No. Izp-2020/1-0179. The authors thank Aiga Bādere and Aiga Veckalne for improving the English language of the manuscript.



## References

- 1 Balode, I. (2012). Vācu-latviešu un latviešu-vācu leksikogrāfija (1991-2010). In A. Lauzis (Ed.), *Vārdnīcu izstrāde Latvijā 1991-2010* (pp. 16-61). Rīga: Latviešu valodas aģentūra.
- 2 Dargis, R., Levāne-Petrova K. & Poikāns I. (2020). Lessons learned from creating a balanced corpus from online data. In A. Utka et al. (Eds.), *Human Language Technologies - The Baltic Perspective*, *Frontiers in Artificial Intelligence and Applications*, 328, 127-134. <https://doi.org/10.3233/FAIA200614>
- 3 Dornseiff, F. (2004). *Der deutsche Wortschatz nach Sachgruppen*. 8. völlig neu bearb. Aufl. von Uwe Quasthoff. Berlin; New York: de Gruyter. <https://doi.org/10.1515/9783110901009>
- 4 Ēdelmane, I., & Ozola, Ā. (2003). *Latviešu valodas augu nosaukumi*. Rīga: Augsburgas institūts.
- 5 Galenieks, P. (1950). *Botaniskā vārdnīca*. Rīga: Latvijas Valsts izdevniecība.
- 6 Hupel, A. W. (1777). *Topographische Nachrichten von Lief- und Ehtland* (Bd. 2), Riga.
- 7 Jasmonts, G., Sviķe, S., & Šķirmante, K. (2022). New information extracting and analysis methodology for the terminology research purposes: The field of biology. In *CEUR Workshop Proceedings: Vol. 3160. IRCDL 2022: 18th Italian Research Conference on Digital Libraries*. Retrieved December 2022 from <https://ceur-ws.org/Vol-3160/paper2.pdf>
- 8 Kaur, G., & Garg, T. (2021). Machine Learning for Optical Character Recognition System. In M. Malarvel et al. (Eds.), *Machine Vision Inspection Systems*, Vol. 2: *Machine Learning-Based Approaches* (pp. 91-107). Beverly: Scrivener. <https://doi.org/10.1002/9781119786122.ch5>
- 9 Linné, C. (1753). *Species plantarum: exhibentes plantas rite cognitatas ad genera relatas, cum differentiis specificis, nominibus trivialis, synonymis selectis, locis natalibus, secundum systema sexuale digestas* (Tom. 1), Holmiae: Laurentii Salvii. <https://doi.org/10.5962/bhl.title.59734>
- 10 Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biology*, 9(8), e1001127. <https://doi.org/10.1371/journal.pbio.1001127>
- 11 Quasthoff, U. (2004). *Methodologische Einführung*. In Franz Dornseiff. *Der deutsche Wortschatz nach Sachgruppen*. 8. völlig neu bearb. Aufl. von Uwe Quasthoff (pp. 193-206). Berlin & New York: de Gruyter.
- 12 Stalažs, A. (2015). Par oficiālajai saziņai atzītu organismu latvisko nosaukumu neveiksmīgu ieviešanu, aizstājot jau iegājušos nosaukumus: piemērs ar *Deroceeras* un *Limax* ģinšu sugām. *Scripta Letonica*, 2(1), 15-23. Retrieved December 2022 from <http://www.rpd-science.org/ScrLet/Vol2015-1.html>
- 13 Stalažs A. (2022). Consequences of taxonomic changes of organisms in local names: examples of the organism names in Latvian. In A. Stalažs et al. (Eds.), *Book of Abstracts of the International Interdisciplinary Scientific Conference Linguistic Diversity, Terminology and Statistics* (pp. 44-48). Ventspils: Ventspils Augstskola.
- 14 Sviķe, S. (2018). A New Dictionary of Botanical Terms: Data Analysis of a Lexicographic Survey. *Economics World*, 6(3), 228-241. <https://doi.org/10.17265/2328-7144/2018.03.007>
- 15 Sviķe, S. (2022). Survey analysis of dictionary-using skills and habits among translation students. In A. Klosa-Kückelhaus et al. (Eds.), *Proceedings of XX EURALEX Congress: Dictionaries and Society* (pp. 494-508). Mannheim: IDS-Verlag. <https://doi.org/10.14618/phpy-6r66>
- 16 Šulcs, V. (2011). *Vaskulāro augu nosaukumterminu datu banka (DB) - pamats zinātniskās nosaukumterminoloģijas pilnveidei Latvijā*. In *Ģeogrāfija, ģeoloģija, vides zinātne. Referātu tēzes* (pp. 240-242). Rīga: Latvijas Universitāte. <https://doi.org/10.22364/ggvz.2011>
- 17 Šulcs, V. (2020). *Nacionālā botāniskā nomenklatūra: mantojums, problēmas un perspektīva*. Latvijas Nacionālā Bibliotēka, *Zinātniskie Raksti*, 6, 117-130.
- 18 Tulkošanas un terminoloģijas centrs (2007). *Latviešu valodas terminoloģijas resursu kvalitātes un pieejamības apzināšana dažādās zinātnes un praktiskās darbības nozarēs*. Rīga: Tulkošanas un terminoloģijas centrs.
- 19 World Flora Online (2022). Retrieved from <http://www.worldfloraonline.org>

## Santrauka

### Arturs Stalažs, Karina Šķirmante, Silga Sviķe, Gints Jasmonts, Roberts Ervīns Ziedīņš Naujos atviros prieigos interaktyvios daugiafunkcės duomenų bazių valdymo sistemos, skirtos specialiajai biologijos leksikai, projektavimo ir kūrimo patirtis

Atliekant terminologijos tyrimus, reikia apdoroti daug duomenų, pavyzdžiui, atlikti gretinamuosius kalbotyros tyrimus ir stebėti kalbos bei terminologijos pokyčius. Elektroninių priemonių ir duomenų bazių trūkumas lėtina procesą, todėl terminologijos ir gretinamosios kalbotyros tyrimų efektyvumui gerinti reikia veiksmingo informacijos valdymo sprendimo.

Nuo 2021 m. terminologų, vertėjų, tyrėjų, programuotojų darbo grupė dalyvauja bendrame projekte ir kuria naują interaktyvią daugiafunkcę informacijos valdymo sistemą (toliau — IMDS), skirtą duomenų paieškai, statistikai ir saugojimui. Ji sukurta kalbos tyrimų ir daugiakalbių lyginamųjų kalbotyros ir terminologijos tyrimų tikslams. Šis tyrimas apibūdina vidurinį projekto etapą, dabartinę būklę, rezultatus ir problemas.

Interaktyvią sistemą sudaro šie moduliai: registracija, duomenų įvedimas, statistika, duomenų eksportas. Duomenų įvedimo modulis jau sukurtas ir naudojamas specialioms leksikos elementams įvesti bei duomenims apibendrinti. Šiame tyrime daugiausia aprašomi iš IMDS gaunami statistiniai duomenys bei sistemos problemos. IMDS (2022 m. gruodžio 16 d.) yra apibendrinta 50 846 vnt. mokslinių (lotyniškų) organizmų pavadinimų, 60 585 skirtingų organizmų pavadinimai latvių, anglų, vokiečių, rusų, estų, lietuvių ir kt. kalbomis, 1 427 organizmų sukeltų ligų pavadinimai, 2 942 žodyno skirsnii įrašai ir 313 terminų su apibrėžimais įrašų. Duomenys surinkti remiantis daugiau nei 4 400 informacijos šaltiniais, kuriuose pateikti organizmų pavadinimai kartu su moksliniais (lotyniškais) pavadinimais.

## About the Authors

### ARTURS STALAŽS

Dr. agr., leading researcher, Institute of Horticulture, Latvia

#### Research interests

Biological terminology, generally organism naming; also plant pests and invasive organisms

**Address** Graudu iela 1, Ceriņi, Krimūnu pagasts, Dobeles novads, LV-3701, Latvia

**E-mail** arturs.stalazs@llu.lv

**Orcid iD**  
0000-0002-3971-6255

### KARINA ŠķIRMANTE

Mg.sc.comp., researcher and lecturer, Ventspils University of Applied Sciences, Latvia

#### Research interests

Development of IT solutions and tools for terminology research

**Address** Inženieru iela 101a, Ventspils, LV-3601, Latvia

**E-mail** karina.krinkele@venta.lv

**Orcid iD**  
0000-0002-2906-4660

### SILGA SVIķE

Dr. philol., Assistant Professor, Researcher, Ventspils University of Applied Sciences, Latvia

#### Research interests

Lexicography, terminology and translation

**Address** Inženieru iela 101a, Ventspils, LV-3601, Latvia

**E-mail** silgas@venta.lv

**Orcid iD**  
0000-0001-8374-3237

### GINTS JASMONTS

Mg.sc.comp., programmer, Ventspils University of Applied Sciences, Latvia

#### Research interests

Use of computer science for problem solving in novel ways, research of new methods

**Address** Inženieru iela 101a, Ventspils, LV-3601, Latvia

**E-mail** gints.jasmonts@venta.lv

**Orcid iD** 0000-0002-5256-1585

### ROBERTS ERVĪNS ZIEDIŅŠ

Programmer, Ventspils University of Applied Sciences, Latvia

#### Research interests

Computer science

**Address** Inženieru iela 101a, Ventspils, LV-3601, Latvia

**E-mail** s20ziedrobe@venta.lv

**Orcid iD** 0000-0002-4000-591X

