БИОЛОГИЯЛЫҚ ҒЫЛЫМДАР / БИОЛОГИЧЕСКИЕ НАУКИ / BIOLOGICAL SCIENCES

DOI 10.54596/2958-0048-2025-4-51-57
UDK 633.31(574.22)
IRSTI 68.35.31

DIGITIZATION OF THE COLLECTION OF THE FAMILY *FABACEAE* IN THE HERBARIUM FUND OF THE "BIOLOGY" DEPARTMENT OF KOZYBAYEV NORTH KAZAKHSTAN UNIVERSITY AND ITS INTEGRATION INTO THE GBIF PLATFORM

Kuandykova M.K.[1*], Tleubergenova G.S.[1], Galaktionova E.V.[1]
[1]Kozybayev North Kazakhstan University, Petropavl, Kazakhstan*
[*]Corresponding author: mkuandykova511@gmail.com

Abstract

The article presents the results of digitizing herbarium specimens of the family *Fabaceae* (Leguminosae) collected in the Kyzylzhar district of the North Kazakhstan region and stored in the herbarium fund of M. Kozybayev North Kazakhstan University (SQU). As a result of this work, 200 specimens belonging to 46 species were systematized, georeferenced, and published on the international platform GBIF (Global Biodiversity Information Facility) for open scientific access.

A structured database in the Darwin Core Archive format was created, providing open access to regional floristic materials. Analysis of the species composition revealed the dominance of meadow-steppe genera (*Lathyrus*, *Astragalus*, *Trifolium*, *Melilotus*), and showed that the peak of herbarium collection occurred in the 1970s–1980s, during the period of active field training.

The modern stage (2020–2025) is characterized by the transition to digital methods of data processing and publication, reflecting the integration of regional botanical research into the global biodiversity monitoring system. The obtained results demonstrate the importance of herbarium digitization for preserving botanical heritage and promoting scientific data exchange within global initiatives for studying the flora of Kazakhstan.

Keywords: *Fabaceae*, herbarium, digitization, GBIF, herbarium fund, biodiversity.

М. ҚОЗЫБАЕВ АТЫНДАҒЫ СҚУ "БИОЛОГИЯ" КАФЕДРАСЫНЫҢ ГЕРБАРИЙ ҚОРЫНДАҒЫ *FABACEAE* КОЛЛЕКЦИЯСЫН ЦИФРЛАНДЫРУ ЖӘНЕ GBIF ПЛАТФОРМАСЫНА ИНТЕГРАЦИЯСЫ

Қуандықова М.К.[1*], Тілеубергенова Г.С.[1], Галактионова Е.В.[1]
[1]«Манаш Қозыбаев атындағы Солтүстік Қазақстан университеті» КеАҚ Петропавл, Қазақстан*
[*]Хат-хабар үшін автор: mkuandykova511@gmail.com

Аңдатпа

Мақалада Солтүстік Қазақстан облысы Қызылжар ауданының аумағында жиналған және Манаш Қозыбаев атындағы Солтүстік Қазақстан университетінің (SQU) гербарий қорында сақталған *Fabaceae* тұқымдас өсімдіктердің гербарий үлгілерін цифрландыру нәтижелері келтірілген. Жұмыс нәтижесінде 46 түрге жататын 200 үлгі жүйеленді, геореференцияланды және еркін ғылыми қолжетімділік үшін GBIF (Global biodiversity Information Facility) халықаралық платформасына орналастырылды.

Жұмыс нәтижесінде аймақтық флористикалық материалдарға ашық қол жетімділікті қамтамасыз ететін Darwin Core Archive форматындағы құрылымдық мәліметтер базасы құрылды. Түрлер құрамын талдау шалғынды-дала туыстарының (*Lathyrus, Astragalus, Trifolium, Melilotus*) үстемдігі анықталды, сондай-ақ гербарий материалдарын жинау шыңы 1970-1980 жылдары, далалық тәжірибелер белсенді жүргізілген кезде болғанын көрсетті.

Қазіргі кезең (2020-2025) Биоәртүрлілікті мониторингтің әлемдік жүйесіне аймақтық ботаникалық зерттеулердің интеграциясының көрсететін деректерді өңдеу мен жариялаудың цифрлық әдістеріне көшумен сипатталады. Алынған нәтижелер Қазақстан флорасын зерттеу жөніндегі жаһандық бастамалар шеңберінде ботаникалық мұраны сақтау және деректермен ғылыми алмасу үшін гербарий қорларын цифрландырудың маңыздылығын көрсетеді.

Кілт сөздер: Fabaceae, гербарий, цифрландыру, GBIF, гербарий қоры, биоәртүрлілік.

ОЦИФРОВКА КОЛЛЕКЦИИ *FABACEAE* В ГЕРБАРНОМ ФОНДЕ КАФЕДРЫ «БИОЛОГИЯ» СКУ ИМЕНИ М. КОЗЫБАЕВА И ЕЁ ИНТЕГРАЦИЯ НА ПЛАТФОРМУ GBIF

Куандыкова М.К.[1*], Тлеубергенова Г.С.[1], Галактионова Е.В[1].
[1]НАО «Северо-Казахстанский университет имени М. Козыбаева», Петропавловск, Казахстан*
[*]Автор для корреспонденции: mkuandykova511@gmail.com

Аннотация

В статье представлены результаты оцифровки гербарных образцов растений семейства *Fabaceae* (Бобовые), собранных на территории Кызылжарского района Северо-Казахстанской области и хранящихся в гербарном фонде Северо-Казахстанского университета имени Манаша Козыбаева (SQU). В результате работы 200 образцов, относящихся к 46 видам, были систематизированы, георефернцированы и размещены на международной платформе GBIF (Global Biodiversity Information Facility) для свободного научного доступа.

В результате работы создана структурированная база данных в формате Darwin Core Archive, обеспечивающая открытый доступ к региональным флористическим материалам. Анализ видового состава выявил доминирование лугово-степных родов (*Lathyrus*, *Astragalus*, *Trifolium*, *Melilotus*), а также показал, что пик сбора гербарных материалов пришёлся на 1970–1980-е годы, когда активно проводились учебно-полевые практики.

Современный этап (2020–2025 гг.) характеризуется переходом к цифровым методам обработки и публикации данных, что отражает интеграцию региональных ботанических исследований в мировую систему мониторинга биоразнообразия. Полученные результаты демонстрируют значимость цифровизации гербарных фондов для сохранения ботанического наследия и научного обмена данными в рамках глобальных инициатив по изучению флоры Казахстана.

Ключевые слова: *Fabaceae,* гербарий, цифровизация, GBIF, гербарный фонд, биоразнообразие.

Introduction

Modern plant biology and taxonomy actively utilize digital technologies to preserve and disseminate herbarium data. Digitization of collections enables the integration of local botanical research with global databases, providing access to biodiversity information on a universal scale. This paper focuses on the digitalization of herbarium materials from the *Fabaceae* family (legumes) collected in the North Kazakhstan region between 1960 and 2025 and their posting on the GBIF platform.

The use of digital technologies in research and education, as well as in modern botanical and ecological research, is finding application. One key area of this transformation is the digitization of herbarium collections - a process that includes the creation of electronic images of specimens, the addition of metadata, and their publication in international databases.

Herbarium collections are a vital source of information on the floristic composition, distribution, and variability of plants. They serve as a scientific archive reflecting the long history of flora studies in the region [1]. However, physical herbarium specimens are subject to

aging, deterioration, and the risk of loss, making the task of their digital preservation particularly pressing.

In recent years, particular attention has been paid to the inclusion of local botanical collections in international information resources, such as the Global Biodiversity Information Facility (GBIF), an international platform that brings together biodiversity data from around the world. This integrates regional data into an international scientific context, increasing their value for comparative, ecological, and taxonomic research.

Herbarium digitization also contributes to increased transparency and accessibility of scientific data: any researcher can use published information to analyze species ranges, assess rarities and flora dynamics, and predict changes in vegetation cover in the context of global climate change.

The *Fabaceae* family was chosen for this study as one of the largest and most ecologically significant families of the flora of Northern Kazakhstan [2, 3]. Its representatives are widespread in meadow, steppe, and forest edge ecosystems, play an important role in maintaining soil fertility, forming vegetation cover, and serve as valuable forage, medicinal, and honey plants [4].

This study aims to digitize herbarium materials of the *Fabaceae* family collected in the Kyzylzhar district of the North Kazakhstan region during the period 1960-2025 and integrate them into the GBIF database.

Research Methods

The herbarium collection of the Department of Biology at Manash Kozybayev North Kazakhstan University (SQU) served as the material for the study. A total of 200 herbarium sheets were collected. The region of distribution is the Kyzylzhar district of the North Kazakhstan region. The process of digitizing herbarium specimens and integrating them into the GBIF international platform consisted of several stages, in accordance with international recommendations:

Preparatory stage and digitization: Each herbarium specimen was labeled (assigned a unique occurrence ID, e.g., SQU00601).

Database creation: Information from the original labels was entered into an electronic catalog (database) in the Darwin Core Archive (DwC) format, the primary standard for biodiversity data exchange. The information was entered in full, including entries in the original language, after which a preliminary data verification was performed.

Georeferencing: For specimens without precise GPS coordinates, manual georeferencing was performed using the point-radius method (determining the geographic coordinates of latitude and longitude) based on text descriptions of the collection site (verbatimLocality).

Publication via IPT (Integrated Publishing Toolkit) and Organization Registration: North Kazakhstan University (SQU) is registered with GBIF as a publishing organization. IPT Installation and Configuration: Integrated Publishing Toolkit (IPT) software, which serves as a tool for creating and publishing datasets, was used on the organization's server.

Data Upload: The prepared dataset in DwC format (GBIF.xlsx - Sheet1.csv) and metadata (dataset description) were uploaded to IPT and published. The data then became available for indexing and searching on the main GBIF portal [5].

Study Results

The digitized collection includes 200 herbarium specimens representing 46 plant species of the *Fabaceae* family. Data analysis revealed key quantitative characteristics and ecological features of the collection. A diagram of the relationship between *Fabaceae* genera is shown in Figure 1.

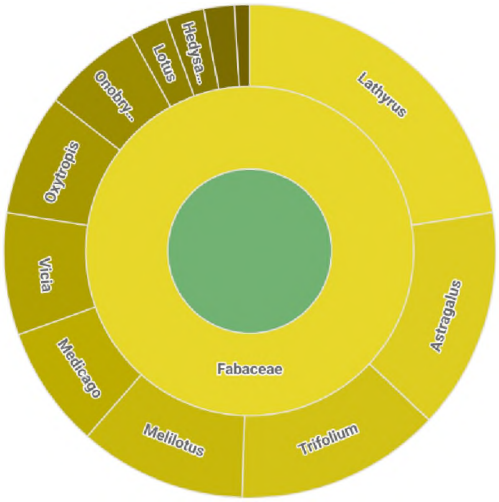

Figure 1. Genera distribution within the *Fabaceae* family, based on 200 digitized accessions on the GBIF platform

The center of the diagram features a common taxon, the family *Fabaceae*, with sectors representing the various genera of plants in this family represented in the collection.

The most numerous genera (occupying the largest sectors of the diagram) are: *Lathyrus* (pea vetch) – 45 accessions (22.5%); *Astragalus* (lowleaf milk vetch) – 29 accessions (14.5%); *Trifolium* (clover) – 27 accessions (13.5%); *Melilotus* (sweet clover) – 22 accessions (11%). Less represented in the collection are the following genera: *Hedysarum* (hedysarum) – 5 accessions (2.5%); *Lotus* (Lotus fulvus) – 5 samples (2.5%), *Glycyrrhiza* (Licorice) – 4 samples (2%), *Caragana* (Caragana) – 2 samples (1%) [6].

The collection of samples spans over sixty years – from 1960 to 2025. The dynamics of the collection of herbarium specimens of the *Fabaceae* family are presented in Figure 2.

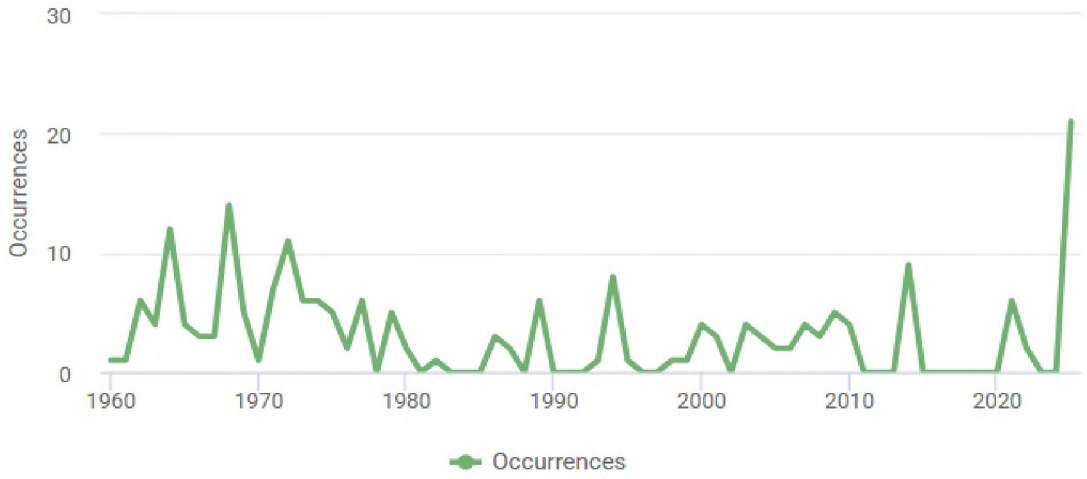

Figure 2. Dynamics of herbarium specimen collection in the *Fabaceae* family on the GBIF platform (1960–2025).

The graph shows the change in the number of herbarium specimens in the Fabaceae family (Y-axis: number of specimens; X-axis: collection years) from 1960 to 2025. From 1960

to 1980, the number of specimens totaled 101 herbarium sheets. From 1980 to 2010, the number of specimens reached 58 sheets. Currently, the period spanning 2010 to 2025 has 38 herbarium sheets.

An analysis of the distribution by year showed that the peak of collecting activity occurred in the 1970s and 1980s (due to the university students' field training). In the 2000s and 2020s, there was a decline in the number of collections, due to a decline in student enrollment and shorter fieldwork periods during those years. Since 2020, a new phase has begun: processing herbarium material for digitalization and updating the metadata base on the GBIF platform. This demonstrates the continuity of research generations and the transition of herbarium work to the digital stage.

All collection sites are located within the Kyzylzhar district (Figure 3). Geographic coordinates (latitude ≈ 54°, longitude ≈ 69°) indicate that the samples cover both northern and southern parts of the district, including river floodplains, meadows, forest edges, and agricultural lands.

The distribution of samples by habitat shows the dominance of meadow ecosystems, which is characteristic of the flora of the North Kazakhstan Region.

Meadow communities: the vast majority of samples (approximately 80%) were collected in meadows (meadow, floodplain meadow, floodplain meadow, steppe meadow, grass-sedge meadow).

Anthropogenic and disturbed habitats: a small proportion of samples (approximately 15%) were collected in disturbed communities or agrocenoses (e.g., fields, roadsides, crop fields).

Forest and forest edge communities: a small number of samples (less than 5%) were collected in wooded areas (forest, birch forest).

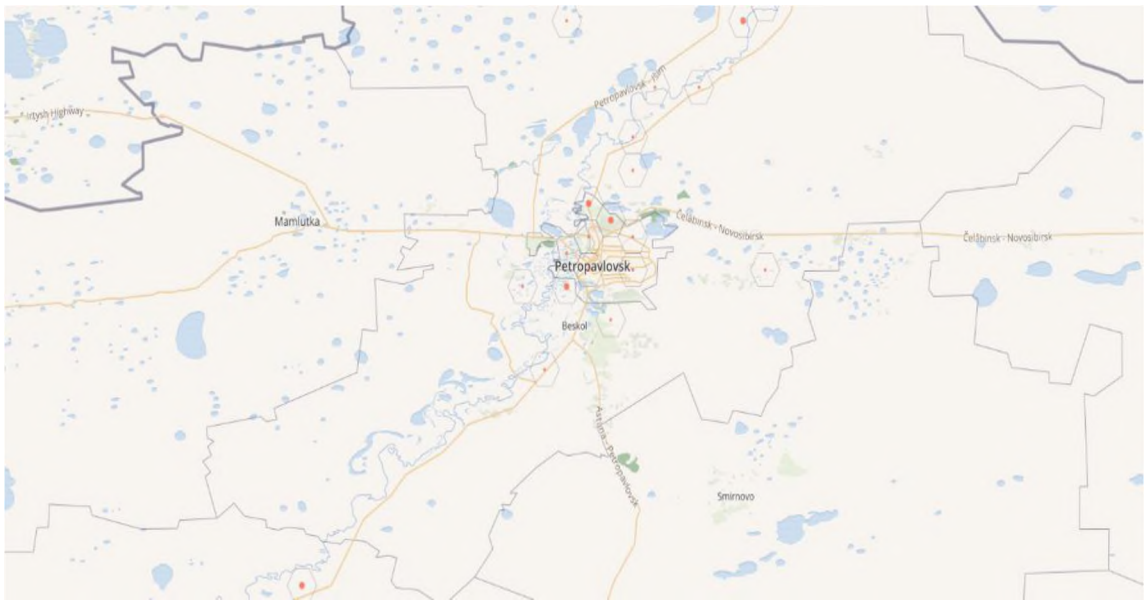

Figure 3. Locations of *Fabaceae* representatives in the Kyzylzhar District of the North Kazakhstan Region

The dominance of species associated with meadows and pastures, such as *Melilotus officinalis* (yellow sweet clover) (4% of the collection) and representatives of the genera

Medicago and *Trifolium* (a combined 12%), highlights their important role in the region's herbaceous cover. The resulting dataset is a valuable resource for further research on the taxonomic structure of *Fabaceae* in the North Kazakhstan region.

The historical part of the herbarium collection of the Department of Biology at North Kazakhstan University includes 54 herbarium specimens collected between the 1960s and 1970s (Table 1). They represent 19 species from 9 genera of the *Fabaceae* family.

Table 1. Number of historical herbarium specimens of the *Fabaceae* family digitized on the GBIF platform

№	Species	Number of samples	Year
1	*Glycyrrhiza uralensis* Fisch. ex DC. (Ural licorice)	1	1960
2	*Melilotus albus* Medik. (White sweet clover)	1	1961
3	*Trifolium pratense* L. (Red clover)	1	1962
4	*Astragalus danicus* Retz. (Danish milk vetch)	1	
5	*Oxytropis pilosa* (L.) DC. (Hairy oxytropis)	2	
6	*Lathyrus pratensis* L. (Meadow pea)	1	
7	*Astragalus stenoceras* C.A. Mey. (Narrow-horned milk vetch)	2	1963
8	*Trifolium lupinaster* L. (Lupine clover)	1	
9	*Lathyrus pratensis* L. (Meadow pea)	1	
10	*Astragalus onobrychis* L. (Sainfoin milk vetch)	4	1964
11	*Astragalus corniculatus* M.Bieb. (Carob milk vetch)	2	
12	*Astragalus tanaiticus* K.Koch. (Don Milkweed)	3	
13	*Oxytropis pilosa* (L.) DC. (Hairy Oxytropis)	2	
14	*Trifolium lupinaster* L. (Lupine Clover)	1	
15	*Astragalus austriacus* Jacq. (Austrian Milkweed)	1	
16	*Oxytropis pilosa* (L.) DC. (Hairy Oxytropis)	2	1965
17	*Melilotus albus* Medik. (White Sweet Clover)	2	
18	*Lotus corniculatus* L. (Corned Lotus)	1	1966
19	*Astragalus austriacus* Jacq. (Austrian Milkweed)	1	
20	*Melilotus albus* Medik. (White Sweet Clover)	1	
21	*Lathyrus palustris* L. (Marsh Pea)	1	1967
22	*Lathyrus pratensis* L. (Meadow Pea)	1	
23	*Vicia cracca* L. (Mouse Pea)	1	
24	*Melilotus albus* Medik. (White Sweet Clover)	1	1968
25	*Lathyrus palustris* L. (Marsh Lath)	3	
26	*Oxytropis pilosa* (L.) DC. (Hairy Oxytropis)	2	
27	*Lathyrus pisiformis* L. (Pea Lath)	2	
28	*Vicia sepium* L. (Vegetable Pea)	1	
29	*Astragalus austriacus* Jacq. (Austrian Milkweed)	2	
30	*Hedysarum gmelinii* Ledeb. (Gmeelian Sweet Clover)	1	
31	*Astragalus danicus* Retz. (Danish Milkweed)	2	
32	*Lathyrus tuberosus* L. (Tuberous Lath)	1	1969
33	*Oxytropis pilosa* (L.) DC. (Hairy Oxytropis)	2	
34	*Hedysarum gmelinii* Ledeb. (Gmeelian Sweet Clover)	1	
35	*Vicia cracca* L. (Mouse Pea)	1	
36	*Hedysarum gmelinii* Ledeb. (Gmeelin's kopeechnik)	1	1970

The analysis shows that the dominant genera by number of specimens are: *Astragalus* - 18 specimens (33.3%), *Lathyrus* - 10 specimens (18.5%); *Oxytropis* - 10 specimens (18.5%); *Melilotus* - 5 specimens (9.2%). A smaller proportion are genera such as *Trifolium* - 3 specimens (5.5%); *Hedysarum* - 3 specimens (5.5%); *Vicia* - 3 specimens (5.5%); *Lotus* and *Glycurrhiza*, each with 1 specimen (1.85%), typical of meadow and forest edge biotopes.

Thus, three genera - *Astragalus, Lathyrus*, and *Oxytropis* - account for approximately 70% of the entire historical collection of the *Fabaceae* family, which emphasizes their leading role in the flora of the region and their active study in the second half of the 20th century. The historical collection lays the foundation for further expansion of the herbarium, and its digitalization on the GBIF platform allows these early materials to be preserved and made accessible for modern floristic and ecological research.

Conclusion

The digitalization of the *Fabaceae* herbarium collection at M. Kozybayev North Kazakhstan University has become an important step in integrating regional floristic data into the global scientific community. The resulting database (200 specimens, 46 species, 1960–2025) enables tracking changes in floristic composition, conducting a detailed analysis of species distribution across biotopes, and serves as an example of the successful digitalization of local herbarium collections. Posting data on the GBIF platform ensures open access to information, which is critical for comparative floristic, systematic, and ecological research.

References:

1. Tleubergenova, G.S., & Kuznetsova, M.A. Flora and Vegetation of the North Kazakhstan Region. Petropavl: M. Kozybayev North Kazakhstan State University, 2017. – 150 p.
2. Tleubergenova, G.S., & Pashkov, S.V. Taxonomic structure of the Family Fabaceae (Leguminosae) in the Flora of the North Kazakhstan Region. // Bulletin of Moscow City Pedagogical University. Series “Natural Sciences”, 2020, No. 2 (38), pp. 60–67.
3. Kalkabaeva, G.M., & Esimzhanova, A.M. Stages of flowering of Fabaceae plants in the North Kazakhstan Region. // Bulletin of Karaganda University. Series “Biology. Medicine. Geography”, 2024, No. 1 (113) pp 34-41.
4. Galiev, Zh.M., & Kalkamanova, A.B. Herbarium Fund of the Department of General Biology of M. Kozybayev North Kazakhstan State University: A Teaching Aid for Students of Botanical Disciplines. Petropavl: M. Kozybayev NKSU, 2006. - 56 p.
5. Buivolov Yu.A., Ivanova N.V., Shashkov M.P. Digitization of data from Chronicles of Nature and scientific biological collections of specially protected natural areas. Study guide. Federal State Budgetary Institution Prioksko-Terrasny State Natural Biosphere Reserve, 2019.- 38 p.
6. Kuandykova M.K. (2025). Herbarium Collection of Fabaceae family North Kazakhstan University named after Manash Kozybayev (SQU). Version 1.2. NLS "M. Kozybayev North Kazakhstan University". Occurrence dataset https:// doi.org/10.15468/j95r2t accessed via GBIF.org on 2025-12-23

Information about the authors:

Kuandykova M.K. – corresponding author, Master’s student, Manash Kozybayev North Kazakhstan University NPLC, Petropavl, Kazakhstan; e-mail: mkuandykova511@gmail.com;
Tleubergenova G.S. – Associate Professor of the Department of Biology, Candidate of Biological Sciences, Manash Kozybayev North Kazakhstan University NPLC; Petropavl, Kazakhstan; e-mail: gbattalova@ku.edu.kz;
Galaktionova E.V.– Senior Lecturer, Master of Biology, Manash Kozybayev North Kazakhstan University NPLC, Petropavl, Kazakhstan; e-mail: evgalaktionova@ku.edu.kz.