Russia2019_14

Supporting the regular georeferencing of European Russia holdings in the Moscow Digital Herbarium (https://plant.depo.msu.ru/)

FINAL ACTIVITY REPORT

Contents

1. Executive summary

The Moscow Digital Herbarium is the largest biodiversity database in Russia. It covers imaged holdings of two biggest herbaria of Moscow–Moscow University Herbarium (MW) and Herbarium of Main Botanical Garden, RAS (MHA). Moscow University Herbarium is a core collection of the Moscow Digital Herbarium and yet the largest Russian dataset published in GBIF (https://www.gbif.org/dataset/902c8fe7-8f38-45b0-854e-c324fed36303). It consists of 979,496 specimen-based records, but only 31% of them were georeferenced by December, 30 2018. The main objective of the project was to increase the amount of

geodata associated with these records. The total number of geodata in MW dataset from European Russia and Northern Caucasus at the starting point of the project (March, 1 2019) was 121,690 georeferences. ***We newly georeferenced 68,166 specimens completing the proposed mission on 125%*** (out of 54,450). From March to May 2019, twelve larger spreadsheets from various regions of European Russia with 24,684 new georeferences were uploaded to the system. Additional 11,815 new georeferences were made by the herbarium staff members as their daily routine job and inserted through the web-interface. From June to August, other 17 larger spreadsheets from various regions of European Russia were uploaded giving us 25,010 new georeferences. These were supplemented by additional 6,657 new georeferences made by the herbarium staff members and inserted through the web-interface. As of September, 25 2019, Moscow University Herbarium (MW) holds 189,856 georeferenced specimens from the focused area. Many georeferences were figured out automatically using ISTRA system due to permanent adding of new label transcriptions. In March–May 2019, we added to the system 119,962 fully captured labels from European Russia and Northern Caucasus superseded by 23,170 fully captured labels in June–September 2019. The current rate of georeferencing in MW herbarium is 48.5% (48.6% for European Russia and 47.6% for Northern Caucasus) whereas the rate of label capturing is 50.2% (52.3% for European Russia and 38.0% for Northern Caucasus).

2. Contact information

Name	Alexey P. Seregin, Dr. Sci.
Institution	M.V. Lomonosov Moscow State University
Address	Leninskie Gory 1, Moscow, 119234, Russia
E-mail	
Phone	
Role in project	Project leader

3. The project and its objectives

The Moscow Digital Herbarium is the largest biodiversity database in Russia. Now, it covers imaged holdings of two biggest herbaria of Moscow–Moscow University Herbarium (MW) and Herbarium of Main Botanical Garden, RAS (MHA). Moscow University Herbarium is a core collection of the Digital Herbarium and at present the largest Russian dataset published in GBIF (https://www.gbif.org/dataset/902c8fe7-8f38-45b0-854e-c324fed36303). It consists of 978,868 specimen-based records, but only 31% of them were mapped by December, 30 2019. Therefore, the main objective of the project is to increase the amount of geodata

associated with holdings already imaged for the Moscow Digital Herbarium. We started the work in March, 01 2019 and successfully finished the mission by September, 25 2019.

The participants of the project are staff members of the Moscow University Herbarium, freelancers permanently involved in our activities, IT-staff of depo.msu.ru platform, and a commercial partner contracted for label capturing. The main stakeholder of the project is an international scientific community interested in filling geographical gaps in freely-accessible data relating to plant diversity of the European Russia. This is demonstrated by the growing number of peer-reviewed research using data from our dataset discovered through GBIF.

The complete grant budget was spent for complete text capturing of 26,000 labels by a commercial partner. Initially, we planed to contract manual text capturing of 33,800 labels, but this figure was ceased after budget cuts. Therefore, we pledged to georeference 54,450 specimens to reach the final number of 176,140 specimens with coordinates from the European Russia in Moscow University Herbarium (MW) published via https://plant.depo.msu.ru/ by 30.09.2019. After checking the text dataset performed by the commercial partner for quality and consistency, we will synchronize text data from the original labels with georeferencing activities implemented by the herbarium staff members and freelancers in line with the project. The synchronization will give us thousands of new georeferences identified automatically by the ISTRA system.

All results were being published on a regular basis as part of weekly synchronization of our database (available at https://plant.depo.msu.ru/) with Moscow University Herbarium GBIF-dataset (https://doi.org/10.15468/cpnhcc).

The report is highlighting some key quantitative indicators of the project progress with necessary explanations in two aspects. The first aspect focuses on a project timeline with a number of key dates when we assessed state of the deals. The second aspect is concerning a spatial coverage of the geodata by September, 25 2019 with a brief description of the spreadsheets integrated into the Moscow Digital Herbarium. The analytical data for the report timeline were mined directly from https://plant.depo.msu.ru/ at 01.03.2019 (starting point), 16.03.2019 (stage 1), 20.03.2019 (stage 2), 27.03.2019 (stage 3), 07.04.2019 (stage 4), 15.04.2019 (stage 5), 29.04.2019 (stage 6), 15.05.2019 (stage 7, mid-term report), 29.05.2019 (stage 8), 02.07.2019 (stage 9), 05.08.2019 (stage 10), 02.09.2019 (stage 11), 17.09.2019 (stage 12), and 25.09.2019 (final report). The spatial data were analysed at 15.05.2019 and 25.09.2019. Screenshots with maps were produced at 01.03.2019 (starting point) and 25.09.2019 (final report). The evaluation results are shown both as numbers of georeferenced records and in percentages to enable the partners and GBIF Secretariat to evaluate the project results in a simple manner.

4. Project implementation

4.1. Activities completed

Here we present a number of key dates when we assessed the current state of the deals. The analytical data for the report timeline were mined directly from https://plant.depo.msu.ru/ at 01.03.2019 (starting point), 16.03.2019 (stage 1), 20.03.2019 (stage 2), 27.03.2019 (stage 3), 07.04.2019 (stage 4), 15.04.2019 (stage 5), 29.04.2019 (stage 6), and 15.05.2019 (mid-term report).

Georeferences as of 01.03.2019 (starting point)

As of 01.03.2019, we imaged and published online 384,201 herbarium specimens from European Russia and Northern Caucasus, including 121,690 georeferenced specimens (31.6%).

Georeferences as of 16.03.2019 (stage 1)

European Russia, specimens	331,748
Northern Caucasus, specimens	55,037
TOTAL specimens	**386,785**
European Russia, georeferences	105,440
Northern Caucasus, georeferences	21,763
TOTAL georeferences	**127,203**
Georeferences added	**5,513**
Completion of the mission, %	10.1

Georeferences as of 20.03.2019 (stage 2)

European Russia, specimens	331,731
Northern Caucasus, specimens	55,039
TOTAL specimens	**386,770**
European Russia, georeferences	110,997
Northern Caucasus, georeferences	21,800
TOTAL georeferences	**132,797**
Georeferences added	**11,107**
Completion of the mission, %	20.4

Georeferences as of 27.03.2019 (stage 3)

European Russia, specimens	331,812

Northern Caucasus, specimens	55,035
TOTAL specimens	**386,847**
European Russia, georeferences	122,042
Northern Caucasus, georeferences	21,800
TOTAL georeferences	**143,842**
Georeferences added	**22,152**
Completion of the mission, %	40.7

Georeferences as of 07.04.2019 (stage 4)

European Russia, specimens	331,912
Northern Caucasus, specimens	55,035
TOTAL specimens	**386,947**
European Russia, georeferences	125,582
Northern Caucasus, georeferences	22,078
TOTAL georeferences	**147,660**
Georeferences added	**25,970**
Completion of the mission, %	47.7

Georeferences as of 15.04.2019 (stage 5)

European Russia, specimens	331,925
Northern Caucasus, specimens	55,035
TOTAL specimens	**386,960**
European Russia, georeferences	134,691
Northern Caucasus, georeferences	22,078
TOTAL georeferences	**156,769**
Georeferences added	**35,079**
Completion of the mission, %	64.4

Georeferences as of 29.04.2019 (stage 6)

European Russia, specimens	333,077
Northern Caucasus, specimens	55,035
TOTAL specimens	**388,112**
European Russia, georeferences	134,881
Northern Caucasus, georeferences	22,078

TOTAL georeferences	**156,959**
Georeferences added	**35,269**
Completion of the mission, %	64.8

Georeferences as of 15.05.2019 (stage 7, mid-term report)

European Russia, specimens	333,077
Northern Caucasus, specimens	56,611
TOTAL specimens	**389,688**
European Russia, georeferences	135,260
Northern Caucasus, georeferences	22,929
TOTAL georeferences	**158,189**
Georeferences added	**36,499**
Completion of the mission, %	67.0

Georeferences as of 29.05.2019 (stage 8)

European Russia, specimens	333,079
Northern Caucasus, specimens	56,611
TOTAL specimens	**389,690**
European Russia, georeferences	140,398
Northern Caucasus, georeferences	23,071
TOTAL georeferences	**163,469**
Georeferences added	**41,779**
Completion of the mission, %	76.7

Georeferences as of 02.07.2019 (stage 9)

European Russia, specimens	334,048
Northern Caucasus, specimens	56,611
TOTAL specimens	**390,659**
European Russia, georeferences	145,498
Northern Caucasus, georeferences	23,071
TOTAL georeferences	**168,569**
Georeferences added	**46,879**
Completion of the mission, %	86.1

Georeferences as of 05.08.2019 (stage 10)

European Russia, specimens	334,048
Northern Caucasus, specimens	56,611
TOTAL specimens	**390,659**
European Russia, georeferences	149,063
Northern Caucasus, georeferences	25,104
TOTAL georeferences	**174,167**
Georeferences added	**52,477**
Completion of the mission, %	96.4

Georeferences as of 02.09.2019 (stage 11)

European Russia, specimens	334,049
Northern Caucasus, specimens	56,611
TOTAL specimens	**390,660**
European Russia, georeferences	152,997
Northern Caucasus, georeferences	25,111
TOTAL georeferences	**178,108**
Georeferences added	**56,418**
Completion of the mission, %	103.6

Georeferences as of 17.09.2019 (stage 12)

European Russia, specimens	334,771
Northern Caucasus, specimens	56,611
TOTAL specimens	**391,382**
European Russia, georeferences	155,782
Northern Caucasus, georeferences	25,974
TOTAL georeferences	**181,756**
Georeferences added	**60,066**
Completion of the mission, %	110.3

Georeferences as of 25.09.2019 (final report)

European Russia, specimens	334,884
Northern Caucasus, specimens	56,596
TOTAL specimens	**391,480**

European Russia, georeferences	162,918
Northern Caucasus, georeferences	26,938
TOTAL georeferences	**189,856**
Georeferences added	**68,166**
Completion of the mission, %	125.2

As of 15.05.2019, we imaged and published online 391,480 herbarium specimens from European Russia and Northern Caucasus, including 189,856 georeferenced specimens (48.5% vs. 31% in the starting point).

Many georeferences were figured out automatically using ISTRA system due to ***permanent adding of new label transcriptions***. In March–May 2019, we added to the system 143,132 fully captured labels from European Russia (137,641) and Northern Caucasus (5,491). This work was made by the freelancers involved in our activities and regarded here as a co-funding of the project. Text dataset submitted by our commercial partner (26,000) is not yet uploaded. It is in the process of thorough checking on data consistency.

Label transcriptions as of 01.03.2019 (starting point)

European Russia, specimens	331,748
Northern Caucasus, specimens	55,037
TOTAL specimens	**386,785**
European Russia, transcriptions	37,349
Northern Caucasus, transcriptions	16,006
TOTAL transcriptions	**53,355**
Level of full-text capturing, %	13.8

Label transcriptions as of 15.05.2019 (mid-term report)

European Russia, specimens	333,077
Northern Caucasus, specimens	56,611
TOTAL specimens	**389,688**
European Russia, transcriptions	151,813
Northern Caucasus, transcriptions	21,504
TOTAL transcriptions	**173,217**
Transcriptions added (from March)	**119,962**
Level of full-text capturing, %	30.8

Label transcriptions as of 25.09.2019 (final report)

European Russia, specimens	334,884
Northern Caucasus, specimens	56,596
TOTAL specimens	**391,480**
European Russia, transcriptions	174,990
Northern Caucasus, transcriptions	21,497
TOTAL transcriptions	**196,487**
Transcriptions added (from June)	23,170
Transcriptions added (from March)	**143,132**
Level of full-text capturing, %	36.6

Spatial coverage. The total number of geodata at the starting point of the project (March, 1 2019) was 121,690 georeferences (Fig. 1). As of September, 25 2019, we georeferenced 189,856 specimens (Fig. 2).

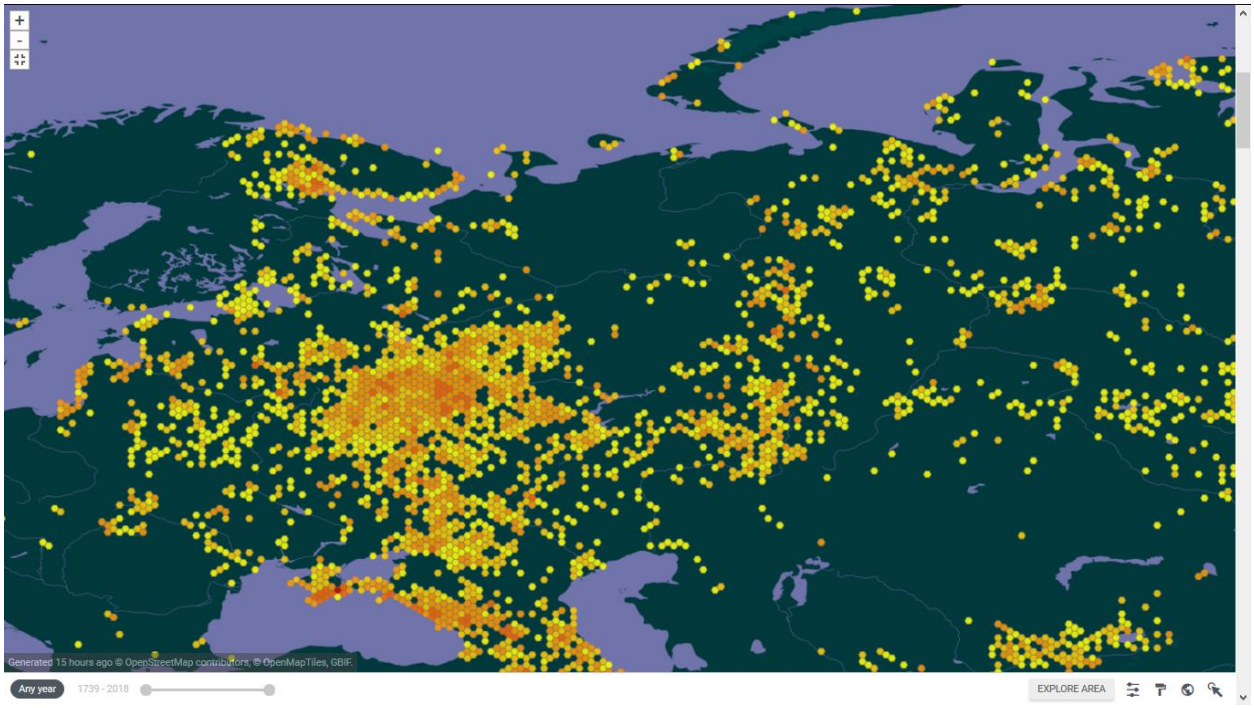

Fig. 1. Eastern Europe and adjacent areas covered by the "Moscow University Herbarium (MW)" dataset (https://doi.org/10.15468/cpnhcc) showing ***121,960 georeferences*** within European Russia (as of 01.03.2019).

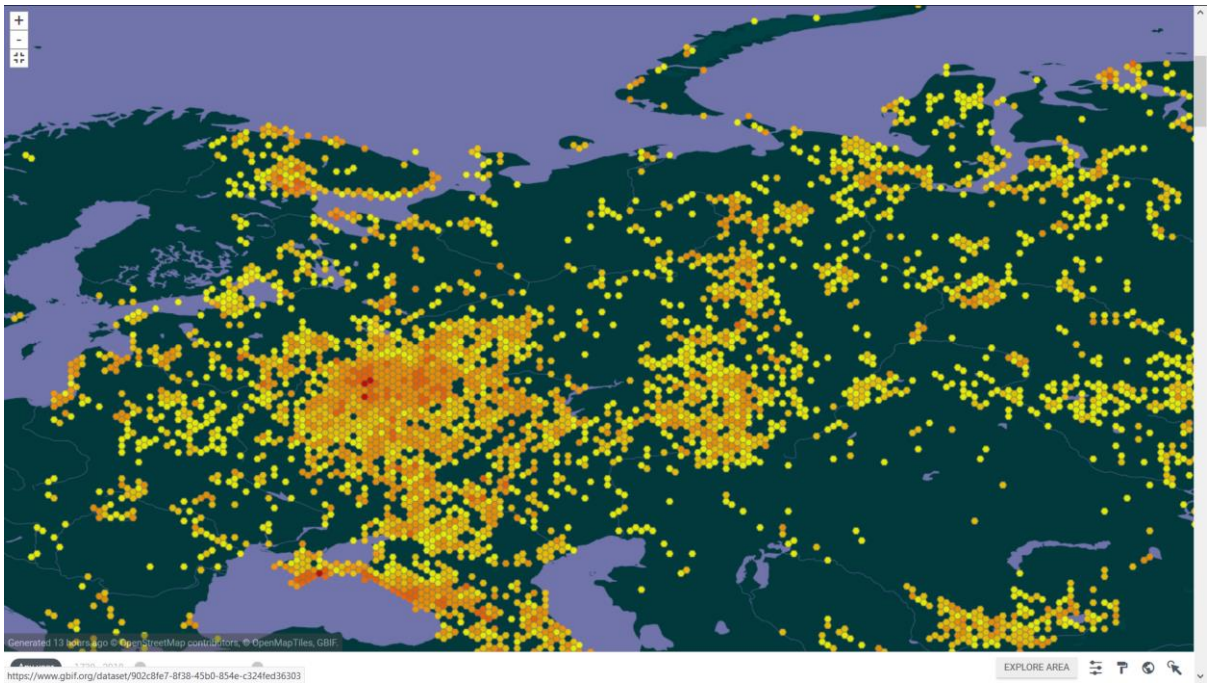

Fig. 2. Eastern Europe and adjacent areas covered by the "Moscow University Herbarium (MW)" dataset (https://doi.org/10.15468/cpnhcc) showing ***189,856 georeferences*** within European Russia (as of 25.09.2019).

The difference between two maps is not much impressive due to good sampling of the collection prior to the starting point of the project. Nonetheless, one should mind new plant diversity data available for (1) Komi Republic, denser sampling of (2) western slope of the Middle Urals, (3) Bashkortostan and (4) Central Chernozem Region, better data availability in (5) the Moscow area and (6) several provinces around Moscow.

The spatial coverage of the geodata with a brief description of the spreadsheets integrated both into the Moscow Digital Herbarium (https://plant.depo.msu.ru/) and the Moscow University Herbarium GBIF-dataset (https://doi.org/10.15468/cpnhcc) is given below. The spatial data were analyzed at 25.09.2019.

Herbarium Curatorial Areas	Code	Georefs Present	Georefs Absent	Number of Specimens	With Georefs, %
EUROPEAN RUSSIA					
N Region	E1	18,285	18,917	37,202	49
NW Region	E2	6,269	7,424	13,693	46
W Region	E3	2,724	3,993	6,717	41
Central Region	E4	34,576	17,772	52,348	66
Moscow Region	E4a	37,638	37,158	74,796	50
Central Forest Region	E5	12,026	2,706	14,732	82
Central Chernozem Region	E6	16,533	31,221	47,754	35
Volga-Kama Region	E7	3,919	9,547	13,466	29
Middle Volga Region	E8	6,010	15,448	21,458	28
Lower Volga Region	E9	8,783	13,786	22,569	39
E Region	E10	10,191	12,648	22,839	45
Rostov Oblast	E12a	5,982	1,348	7,330	82
NORTHERN CAUCASUS					
Northern Caucasus – W	K1a	8,561	5,999	14,560	59
Northern Caucasus – Central	K1b	7,873	11,210	19,083	41
Northern Caucasus – E	K1c	2,867	5,417	8,284	35
Dagestan	K2	2,010	3,701	5,711	35
Russian Black Sea Coast	K3	5,627	3,330	8,957	63

Phase 1: from March, 1 to May, 15 2019

We added ***36,499 georeferences*** in this period. The following twelve major spreadsheets with geodata incorporated during phase 1 were the most essential:

1) Georeferences from the Urals by V.P. Travkin, set #1 (462 georefs);

2) Georeferences from Eastern Europe and Northern Caucasus (general) by gbif.ru team (3,387 georefs);

3) Georeferences from Eastern Europe for specimens collected after 2000 by N. Kopylova (399 georefs);

4) Georeferences from Moscow Region for specimens collected in 1950-1999 by I. Krivokorin (1,032 georefs);

5) Georeferences from the City of Moscow (retrieved by OCR) by I. Melnik (453 georefs);

6) Georeferences from the Urals by V.P. Travkin, set #2 (1,095 georefs);

7) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #1 (1,958 georefs);

8) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #2 (752 georefs);

9) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by A.P. Seregin (4,814 georefs);

10) Georeferences from Arkhangelsk Oblast and Komi by K.V. Dudova (5,773 georefs);

11) Georeferences from Moscow Region, Central Region, and Central Chernozem Region (general) by I. Krivokorin, set #3 (1,685 georefs);

These twelve spreadsheets include 24,684 georeferences. Other 11,815 newly incorporated georeferences were made by the herbarium staff members as their daily routine job and inserted through the web-interface directly into the Moscow Digital Herbarium (https://plant.depo.msu.ru/).

We also uploaded one large text dataset with label transcriptions om phase 1:

12) Text transcriptions of ca. 144,000 labels (made by the commercial partner in 2018) with consequent automatic georeferencing (2,874 georefs).

Phase 2: from May, 16 to September, 25 2019

We added ***31,667 georeferences*** in this period. The following 17 major spreadsheets with geodata incorporated in phase 2 were the most essential:

13) Georeferences from the Urals by V.P. Travkin, set #3 (2,424 georefs);

14) Georeferences from European Russia (general) for specimens collected in 1991–2018, by I. Melnik (1,795 georefs);

15) Georeferences from Lower Volga Region by A. Salmin (3,231 georefs);

16) Georeferences from North-Western Region by A. Sefilyan (348 georefs);

17) Georeferences from Krasnodar Krai and Adygea by A. Salmin, set #1 (2,089 georefs);

18) Georeferences from Eastern Europe (general) by I. Melnik (3,962 georefs);

19) Coordinates mined from labels from miscellaneous regions (general) by A. Bortsova (1,368 georefs);

20) Georeferences from Moscow Oblast and the City of Moscow by N. Ivanova, set #1 (837 georefs);

21) Georeferences from Krasnodar Krai and Adygea by A. Salmin, set #2 (283 georefs);

22) Georeferences from Krasnodar Krai and Adygea by E. Kashirina (582 georefs);

23) Georeferences from Tula Oblast by A. Lakomov (255 georefs);

24) Georeferences from Moscow Oblast and the City of Moscow by N. Ivanova, set #2 (743 georefs);

25) Georeferences from European Russia (general) with focus on Moscow Region, Central Region, and Central Chernozem Region by A.P. Seregin, set #1 (3,360 georefs);

26) Georeferences from Krasnodar Krai and Adygea by A. Salmin, set #3 (873 georefs);

27) Georeferences from European Russia (general) with focus on Moscow Region, Central Region, and Central Chernozem Region by A.P. Seregin, set #2 (1,134 georefs);

28) Georeferences from N edge of the City of Moscow (N.I. Kuznetsov's 1920s collections) by A.P. Seregin (841 georefs);

29) Georeferences from European Russia (general) with focus on Moscow Region, Central Region, and Central Chernozem Region by A.P. Seregin, set #3 (885 georefs).

These twelve spreadsheets include 25,010 georeferences. Other 6,657 newly incorporated georeferences were made by the herbarium staff members as their daily routine job and inserted through the web-interface directly into the Moscow Digital Herbarium (https://plant.depo.msu.ru/).

Georeferencing was supplemented on phase 2 by massive uploads of label transcription datasets made by freelancers permanently involved in our activities:

30) Label transcriptions from old Zinger's Herbarium by D. Borodina, set #3 (1,178 labels);

31) Label transcriptions from old Zinger's Herbarium by D. Borodina, set #4 (2,555 labels);

32) Label transcriptions from Northern Region by T. Sukhova (263 labels);

33) Label transcriptions from old Zinger's Herbarium by D. Borodina, extra set (378 labels);

34) Label transcriptions from Moscow Oblast and the City of Moscow by D. Borodina (10,231 labels);

35) Label transcriptions from Moscow Oblast and the City of Moscow by I. Melnik and T. Sukhova (11,385 labels).

4.2. Ongoing and post-project activities

The plan of the project submitted to GBIF Secretariat as part of the application form (and later included to the contract) did not include any unrealistic intensions. Therefore, we did not made any substantial changes in the project plans and successfully achieved proposed results. We completely finished georeferencing by September, 25 2019. The final report was submitted to GBIF Secretariat in September, 30 2019.

26,000 label transcriptions made by the commercial partner have been received by us completely by September, 30 2019. Currently, we are checking this text dataset for quality and consistency. We are planning to synchronize geodata and these text data on the portal by October, 15 2019.

5. Project deliverables

a. Data

Title of dataset	Taxonomic/ geographic/ temporal scope	Approximate number of records	Sampling methodology/protocol used (if relevant)	Geographic accuracy for most records (in m or km, or province, country etc.)	Current state (e.g. undigitized, digitized)	State by final report
Moscow University Herbarium (MW)	Vascular plants; Global; 1739–2018	975,000 records, including 54,450 newly georeferenced ones	Standard procedures of manual georeferencing using open-source web-platforms and map libraries; ISTRA algorithm for aggregation of identical labels	1–10 km	Herbarium specimens are imaged and published in GBIF; 14% got captured labels	68,166 new georeferences added out of expected 54,450 (125%); text capturing of 26,000 labels was performed by a commercial partner.

Expected milestones by final report:

Milestone	Status by final report
1. All mobilized data has been published to GBIF.org	1. Confirmative. See our dataset https://www.gbif.org/dataset/902c8fe7-8f38-45b0-854e-c324fed36303 and free text below.
2. (All published data meet the minimum requirements outlined in the Data Quality Requirements available at http://bid.gbif.org/en/community/data-quality/)	2. Confirmative. See “Issues and flags”: https://www.gbif.org/dataset/902c8fe7-8f38-45b0-854e-c324fed36303/metrics
3. Best practices and lessons learned have been documented	3. Confirmative. See free text below.

The dataset "Moscow University Herbarium (MW)" (https://doi.org/10.15468/cpnhcc) which includes 68,166 newly georeferenced specimens from European Russia and the following formal disclaimer in the metadata (Description of the dataset section): "68,166 occurrences were georeferenced from March, 1 2019 to September, 24 2019 by the support of FinBIF project ‘Supporting the regular georeferencing of European Russia holdings in the Moscow Digital Herbarium’ (Russia2019_14, see details at https://www.gbif.org/project/2dfnq4VJQxHVSoPZWykiCb/supporting-the-regular-georeferencing-of-european-russia-holdings-in-the-moscow-digital-herbarium). They include newly georeferenced records from Eastern European plain, western slope of the Urals and Northern Caucasus."

6. Project communications

To deliver our experience in the collection digitization and data mobilization to a wider audience the project leader Dr. A.P. Seregin made three invited talks on the seminar “Natural sciences collections of Yugra: collecting, fixation, storage, delivering to a scientific

community” (Khanty-Mansiysk, March 25-26, 2019). The links are already published on project page.

Another invited lecture entitled “Why publish data through GBIF and other open repositories? Moscow Digital Herbarium experience: benefits for the scientific community and for individual researchers” was given by Dr. A.P. Seregin on the Seminar “Diversity of soil animals in Russia: publication and effective use of source data” (August 29-30, 2019, Moscow)).

Later on, Dr. A.P. Seregin gave an invited lecture on 114th Società Botanica Italiana Congress (4–7 September 2019, Padova) entitled "Moscow Digital Herbarium: a global approach through the regional actions" within the congress’ Symposium "Herbaria: still relevant in the 21th century?"

Some key points of the project are being published regularly in the Herbarium blog in the Russian social network VK (https://vk.com/mw_herbarium) in the form of short daily posts. We are communicating with a wider audience this way and getting some feedback from the users.

Links to be added:

https://www.researchgate.net/publication/335662295_Moscow_Digital_Herbarium_a_global_approach_through_the_regional_actions (a presentation of the talk given on 114th Società Botanica Italiana Congress (4–7 September 2019, Padova)), in English

https://www.researchgate.net/publication/335475069_Zacem_publikovat_dannye_v_GBIF_i_na_drugih_otkrytyh_resursah_Opyt_Cifrovogo_gerbaria_MGU_vygody_dla_naucnogo_soobsestva_i_dla_otdelnyh_issledovatelej (a presentation of the talk given on Seminar “Diversity of soil animals in Russia: publication and effective use of source data” (August 29-30, 2019, Moscow)), in Russian

7. Evaluation: findings and conclusions

An assessement of the overall outcomes and impacts of your project, including strengths and weaknesses in its implementation and results. Try to identify clear conclusions from your experience during the implementation of the project. If any changes have been made to the project plans please clearly indicate this in the report and the reasons for this. Also report on any feedback on the project’s relevance from the partners and stakeholders

Data usage. Using the "Activity" bookmark, we detected an intensification of uploading data events from our GBIF-dataset. The average number of daily uploading events in which MW

dataset was involved was ca. 46 in late March 2019, ca. 52 in mid-May 2019, and finally 52.3 in late September 2019. About 25% of download events employs "Has coordinates" filter, therefore the growing number of georeferences in the Moscow University Herbarium dataset leads to the better visibility of our holdings.

Data citations. The number of citations of our GBIF-mediated data (as seen in gbif.org counter) is also steadily growing. GBIF has established itself as an essential infrastructure underpinning biodiversity-related science. This is clearly demonstrated by the increasing volume of peer-reviewed research using GBIF-mediated data. As of 15.05.2019, our GBIF-dataset accumulated 55 citations—this number doubled since mid-November 2018. As of 25.09.2019, we gained 80 citations. We could assume that the growing number of citations is one of the most prominent project's deliverables.

World positions. In the series of ranked lists (see below) one can identify easily, that Moscow University Herbarium is one of the world's leaders in digitization of herbarium specimens. Being only 63rd herbarium in the world as for physical collections (Thiers 2019), MW is the 11th largest GBIF-donor across herbaria for specimen-based records, the 12th largest GBIF-donor across herbaria for georeferenced specimen-based records, the 6th largest GBIF-donor for plant images, and the 5th largest GBIF-donor for imaged specimens of plants.

The largest specimen-based datasets in GBIF for plants:

1. The vascular plants collection (P) at the Herbariu… 5,427,934
2. MEL AVH data 5,071,885
3. Naturalis Biodiversity Center (NL) - Botany 4,824,265
4. Tropicos Specimen Data 4,672,826
5. The New York Botanical Garden Herbarium (NY) 3,290,076
6. NMNH Extant Specimen Records 3,010,783
7. Meise Botanic Garden Herbarium (BR) 1,424,968
8. PRECIS 1,117,942
9. Harvard University Herbaria: All Records 1,060,056
10. Phanerogamic Botanical Collections (S) 1,018,284
11. ***Moscow University Herbarium (MW) 979,480***

The largest specimen-based datasets in GBIF for georeferenced records of plants:

1. MEL AVH data 4,667,907
2. Tropicos Specimen Data 3,128,898
3. Naturalis Biodiversity Center (NL) - Botany 1,488,245
4. The New York Botanical Garden Herbarium (NY) 1,042,284
5. PRECIS 903,698
6. Lund Botanical Museum (LD) 871,712
7. Vascular Plant Herbarium, Oslo (O) 598,158
8. Field Museum of Natural History (Botany) Seed Pla… 554,080
9. NMNH Extant Specimen Records 518,234
10. BRI AVH data 502,070
11. 500,000 plant Specimens from PE Herbarium in China… 500,000
12. ***Moscow University Herbarium (MW) 417,376***

The largest datasets in GBIF for plants with images (all record types):

1. The vascular plants collection (P) at the Herbariu… 5,378,032
2. Naturalis Biodiversity Center (NL) - Botany 4,539,421
3. iNaturalist Research-grade Observations 3,542,636
4. The New York Botanical Garden Herbarium (NY) 2,488,309
5. NMNH Extant Specimen Records 2,389,000
6. ***Moscow University Herbarium (MW) 978,224***
7. Royal Botanic Gardens, Kew - Herbarium Specimens 922,351
8. Herbarium specimens of Université de Montpellier 2… 830,324
9. Vascular Plant Herbarium, Oslo (O) 711,375
10. RB - Rio de Janeiro Botanical Garden Herbarium Col… 662,799

The largest specimen-based datasets in GBIF for plants with images:

1. The vascular plants collection (P) at the Herbariu… 5,378,032

2. Naturalis Biodiversity Center (NL) - Botany 4,539,169
3. The New York Botanical Garden Herbarium (NY) 2,488,309
4. NMNH Extant Specimen Records 2,352,939
5. ***Moscow University Herbarium (MW) 978,224***

National positions. In Russia, we are still the leaders in data mining. Moscow University herbarium dataset is the largest source of primary biodiversity information published by Russian institutions and the largest dataset for the territory of the Russian Federation (within borders accepted by ISO standards).

The second largest dataset in GBIF from Russian institutions is managed in the Netherlands (EOD), whereas the third largest is also published by the Moscow University Herbarium ("A grid-based database…").

The second largest dataset in GBIF for the territory of Russia is iNaturalist. Its rapid growing in Russia is a result of our tremendous activity in herbarium-associated project "Flora of Russia" (https://www.inaturalist.org/projects/flora-of-russia) managed by the Moscow University staff members.

The largest datasets in GBIF from Russian institutions:

1. ***Moscow University Herbarium (MW) 979,496***
2. EOD - eBird Observation Dataset 171,034
3. A grid-based database on vascular plant distributi… 123,054 (managed by MSU)
4. Birds of Northern Eurasia 65,917
5. Birds and Mammals Collections of the Zoological… 54,120 (managed by MSU)

The largest datasets in GBIF for the territory of Russia:

1. ***Moscow University Herbarium (MW) 622,788***
2. iNaturalist Research-grade Observations 208,623 (largely managed by MSU)
3. Geographically tagged INSDC sequences 195,599
4. EOD - eBird Observation Dataset 171,034
5. A grid-based database on vascular plant distributi… 123,054 (managed by MSU)

8. Recommendations and lessons learned

What ***best practices and lessons*** we learned after the finishing of the Noah's Ark megagrant (2015–2018) which aimed to support the Moscow University biological collections?

(1) Never give up! We applied for 16 grants, competition and prizes to keep our work moving. The most precious trophies were not conquered, but nonetheless we have four modest grants to promote regional missions and a contract with GBIF Secretariat funded by FinBIF.

(2) Do not rely on volunteers in your work. You will spend dozens of hours to teach a volunteer with no guarantee that he will come tomorrow. Think twice, do you really need him? Only staff members of the Moscow University Herbarium and freelancers permanently involved in our daily activities are involved in the project.

(3) Virtual consolidation of lesser collections within a larger web-hub is a budget-efficient solution. We mobilized ca. 16,000 herbarium specimens from the Main Botanical Garden of RAS (MHA) within "Moscow project". Herbarium of the Main Botanical Garden is the second largest in Moscow, and this will be their first appearance in GBIF as soon as they will be registered as a publisher. Their dataset will be published via the Moscow University IPT.

9. Future plans

A description of how the partners involved will build on the results of this project in their future work. This could include future collaborative activities, such as plans to complete any unfinished project activities and how the future impact of the project could be monitored or measured.

The GBIF project entitled “Supporting the regular georeferencing of European Russia holdings in the Moscow Digital Herbarium (https://plant.depo.msu.ru/)” is one of the small-scale projects aimed to support digitisation of the Moscow University Herbarium after the completion of the Noah's Ark megagrant (2015–2018). Other three projects are still active. Here is the brief review of our activities in near future.

"Flora of Moscow" Information System on Moscow Digital Herbarium web-platform (4,000,000 RUR from RFBR, deadline October, 15 2020). The City of Moscow is a unique area that requires special attention to the study of various components of the environment. At the present stage of technological development, it is right time to transfer into the digital form extensive diverse materials on the flora of Moscow, to standardize the data and to provide open access to the data on plant diversity of our metropolis. For the first time in Russia, a spatial database of the urban flora of a large city will be created: (1) data from labels will be fully captured, (2) all specimens will be georeferenced and (3) all data will be placed in the public domain in the form of a convenient and interesting web-resource for Moscow plants. As a result of the project, a database with records on 20,000 specimens collected in Moscow and a 10-kilometer-long strip along Moscow borders and their georeferences will be published on Moscow Digital Herbarium web-platform. The database will integrate electronic information on the flora of Moscow from various sources, as well as 3,000 new photo records of individual species and 1,000 new herbarium specimens collected within the project framework. Based on mobilized data and other available sources of information, an actual checklist of the flora of Moscow will be compiled and maps of distribution in Moscow for each species drawn up. Flora of Moscow on Moscow Digital Herbarium web-platform will be a unique resource covering the flora of the city in three aspects – botanical, spatial and dynamic. The results of the research will form a firm foundation for further study and comprehensive protection of the local flora.

The most diverse regional flora in Russia - biodiversity informatics of vascular plants in Krasnodar Krai (1,200,000 from RFBR, deadline April, 30 2021). The Caucasus is the only Russian hotspot that is present in all published schemes of distribution of world's

biodiversity, which makes it the top priority area for floristic studies in Russia. No one tried to characterize the internal structure of the Caucasian hotspot and its most diverse sector, i.e. the Western Caucasus. The goal of the project is to create the most complete dataset on the floristic diversity of the Russia's richest regional flora. We are going to give a description of the spatial structure of the flora within this territory and to reveal the patterns of distribution of individual species, their groups and floristic diversity in general. For the first time in Russia, a complete regional dataset comprising at least 25,000 images, labels and georeferences based upon herbarium specimens will be compiled and published in the public domain. For the first time, interactive distributional maps for all species of the regional flora will be compiled, with each point on the map associated with a specific record in the database and an image of the herbarium specimen. For the first time, a cartographic description of the spatial structure of the Russia's richest regional flora will be given. For the first time, it will be possible to generate automatically floristic checklists of any spatial dimension. According to the results of the project, we will create a regional open access web-resource "Flora of the Krasnodar Krai based upon the materials of the Moscow Digital Herbarium". The portal will include at least 25,000 specimens from the Moscow University Herbarium. The resource will have two versions, i.e. the operational version for data entry and data editing and the user version. The portal will deploy a library of images, which will consist of images of all specimens at 300 dpi resolution and photographs of plants in nature. All samples will have fully captured labels and geolocation. The analytical result of the study will include the creation of 2,000 electronic interactive maps of the distribution of all species in Krasnodar Krai. It will be the basis for identifying phytogeographical patterns using the biodiversity informatics methodological apparatus.

Diversity assessment and visualization of the Tula Oblast flora using modern data technologies (2,400,000 from RFBR, deadline April, 30 2022). In the Tula Oblast characterized by a highly developed industry, a high percentage of land development and a lack of protected natural areas, an environmental challenge is particularly acute. The project is aimed at the inventory of biodiversity (i.e. diversity of vascular plants of Tula Oblast) at a fundamentally new level. For the first time on the regional level text database and GIS-module with geolocations will compliment the library of high quality images. The goal of the project is to assess the actual diversity and spatial structure of the flora of the Tula region by combining various data sources on a single platform, visualizing all the available information by digitizing data and creating a multifunctional electronic resource with open access. The main scientific idea is to make the Tula flora visible to the entire world community, and first of all, to specialists working in the field of studying and preserving biodiversity, solving problems of nature conservation at all levels up to global. To do this, it is supposed to do the

following: 1) combine the data of 5 herbarium collections made in the region; 2) scan the entire regional herbarium; 3) supplement the materials with photographs of plants in nature; 3) create a database of labels; 4) establish the geo-referencing of the finds; 5) on the basis of all received information to create a powerful information online resource "Digital Herbarium of Tula Oblast". According to the results of the project, a database with 23,000 records will appear on this platform. This will enable us to standardize data and to provide an open access to data on flora. The results of the study will form the foundation for studying and conserving the biodiversity of the region, form a powerful basis for the modern regional checklist and atlas of the spatial distribution of species, and will make a significant contribution to the 'Atlas of the Russian Flora'.

10. Annex – Sources of verification

Links:

https://www.researchgate.net/publication/335662295_Moscow_Digital_Herbarium_a_global_approach_through_the_regional_actions (a presentation of the talk given on 114th Società Botanica Italiana Congress (4–7 September 2019, Padova)), in English

https://www.researchgate.net/publication/335475069_Zacem_publikovat_dannye_v_GBIF_i_na_drugih_otkrytyh_resursah_Opyt_Cifrovogo_gerbaria_MGU_vygody_dla_naucnogo_soobsestva_i_dla_otdelnyh_issledovatelej (a presentation of the talk given on Seminar “Diversity of soil animals in Russia: publication and effective use of source data” (August 29-30, 2019, Moscow)), in Russian

https://www.gbif.org/project/2dfnq4VJQxHVSoPZWykiCb/supporting-the-regular-georeferencing-of-european-russia-holdings-in-the-moscow-digital-herbarium (project description and mid-term report)

Signed on behalf of the project partners Date

---------- ----------