

Russia2019_04

European North-East Russia mosses occurrence data mobilization on the base of SYKO herbarium mosses collection

MID-TERM ACTIVITY REPORT

Guidelines on how to complete the activity report are included in italics. You are welcome to remove the guideline text from the document before you submit the report.

Please note that once the activity report has been approved, it will be added to your project page. Therefore, we kindly ask you not to add any contact details of persons in the report unless you have permission from the person to do so.

Contents

MID-TERM ACTIVITY REPORT	1
Executive summary	1
Contact information	2
Introduction	2
The project and its objectives	3
Project activities completed by mid-term.....	5
Project communications	8
Mid-term evaluation findings and recommendations for the remaining project implementation period	8
Annex – Sources of verification	10

Executive summary

Provide a brief explanation of the context and the approach taken for the mid-term evaluation, and a summary of the main conclusions, lessons learned and recommendations for the remaining project period.

The aim of the project is to digitize and publish label information from at least 8,000 moss specimens in the herbarium (SYKO) moss collection of the Institute of Biology of the Komi Scientific Center of the Ural Branch of the Russian Academy of Sciences. As the non-

dataset project deliverable the project will document the process of mobilizing data from herbarium labels in multiple forms: an article in the Russian open access peer-reviewed journal; a webinar on the institution's YouTube channel, and a report at one Russian scientific conference. Most of these tasks have been completed. We were able to develop a data entry system that significantly improves the efficiency of monotonous manual labor. Up to 27 May 2019 there are 11,771 labels images uploaded to the database and 11,257 of them were digitized with minimum set of fields: catalog number, species name, collection date, geographic coordinates, names of persons who collected and identified the species in the sample. This number of digitized labels will allow us to publish about 13,000 mosses occurrences. The first version of dataset with 4,991 moss occurrences were published and registered in GBIF. The project results were presented in regional scientific conference "Symbiosis 2019" (Perm city, 13.05.2019) and in the form of YouTube webinar "Digitization of herbarium label data and their preparation for publication in GBIF. Practical case of bryophytes collections of the SYKO herbarium". The feedback given showed that it is necessary to describe our experience not only in form of oral presentation but also in the form of scientific publication. The software tools developed for the project should be documented and published with source code in one of open repository (e.g. GitHub).

Contact information

Provide the name, institutional affiliation, role in the project and contact details of the author(s) of the report.

Ivan Chadin, Institute of Biology of Komi Science Centre of the Ural Branch of the Russian Academy of Sciences (IB Komi SC UB RAS, team leader).

Introduction

This section should explain to readers what they will find in this report. It should include:

- *A description of how the evaluation has been carried out (e.g. consultation or surveys with project partners and participants).*
- *A description of how the project partners will use the evaluation results.*

The report is designated to the mid-term evaluation of the project "European North-East Russia mosses occurrence data mobilization on the base of SYKO herbarium mosses collection" (Russia2019_04). The evaluation was performed with the help of project participants and colleagues feedback received after project results presentation on the

scientific conference and webinar. The evaluation results may be useful for project partners as a part of benchmarking for the similar data mobilization projects.

The project and its objectives

A brief summary of the project to help readers understand its objectives, including, for example:

- *The project's start date and expected duration*
- *A list of project participants and description of the main stakeholders*
- *The targeted capacity needs as outlined in the project proposal*
- *The project objectives and expected deliverables as included in the project proposal*

The project's start date: 04.02.2019

Expected duration: 6 months and 26 days (calendar days)

A list of project participants and description of the main stakeholders

Participants:

Chadin Ivan – team leader, program coding, dataset publishing with IPT

Zheleznova Galina – briologist, collection curator, labels data entering, verification of entered data

Shubina Tatyana – briologist, labels data entering, verification of entered data

Rubtsov Mikhail – engeneer, geo-referencing

Litvinenko Galina – technician, labels image digitization, labels data entering

Stakeholder

Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences (IB Komi SC UB RAS). It was was organized in 1962. The Institute is the largest the academic center for ecological and biological research on the European North-East Russia. The Institute consists of six departments and four laboratories, Zoological museum, Botanical garden and herbarium (SYKO). Main areas of research are: Study of biodiversity, structural and functional organization, stability and productivity of taiga and tundra ecosystems; Biological action of ionizing radiation and other physico-chemical factors on cells, living organisms and natural ecosystems; problems of radiation and ecological genetics; Study of physiological-biochemical basis of adaptation and reproduction of plants in cold climates; Study of biologically active compounds in plants of natural flora and introduced plants; Development of methods for monitoring, bioindication; creation of

inventories and databases of biological resources of the European North-East with the use of remote sensing and GIS technologies.

The Institute maintain rich biological collections. The SYKO herbarium keep 315 045 samples (including 206 445 samples of vascular plants and 57 000 samples of mosses). IB Komi SC UB RAS is interested in it's biological collections digital mobilization.

Project activities completed by mid-term

This section should summarize the project activities completed by the mid-term, with a description of the associated outputs and deliverables. Please highlight any changes from the original plans provided in the full project proposal.

If no result has been achieved on a specific point, please indicate it as a “no result achieved yet”.

Also, please comment on the expected milestones for the mid-term reporting as defined in the contract.

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 mid-term report
Mosses Collection of SYKO Herbarium (Syktyvkar, Russia)	Plantae, Bryophyta 1947-2017	8,000	Data will be acquired from herbarium labels. Georeferencing will be performed according Zheleznova G, Shubina T, Degteva S, Rubtsov M, Chadin I (2018): Moss occurrences in Yugyd Va National Park ... http://ib.komisc.ru:8088/ipt/resource?r=mosses_occurrence_yugyd_va&v=1.5	1.6–3.0 km	February 2019 – undigitized May 2019 – digitized	11,257 labels with minimal set of information were digitized

b. Other deliverables

Description	State by mid-term report
In the first half of the project period, the experience gained during its implementation will be presented in the form of a webinar on the YouTube channel for biodiversity specialists.	The webinar was organized on 22.05.2009. It was attended by about 60 experts in the field of biodiversity (40 online and c.a 20 live in conference hall). Webinar recording is available here: https://www.youtube.com/watch?v=TMenzugsCaY and here (trimmed video): https://www.youtube.com/watch?v=IMs6k8PUrN8
Prior to the end of the project, the results will be presented in the preprint in Russian, which will later be submitted for publication in an open access peer-reviewed journal, as well as in a report at the scientific conference.	No result achieved yet

Expected milestones by mid-term report

Milestone	Status by mid-term progress report
At least one dataset has been published to GBIF.org	The first version of dataset with 4,991 moss occurrences were published and registered in GBIF: Zheleznova G, Shubina T, Rubtsov

	M, Litvinenko G, Chadin I (2019). SYKO Herbarium Mosses Collection. Version 1.2. Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences. Occurrence dataset https://doi.org/10.15468/yjdjs4 accessed via GBIF.org on 2019-05-27.
--	---

Project communications

Describe the plans to communicate and share the results of your project with the project stakeholders and broader GBIF community.

In the first half of the project duration, the obtained experience was presented in the form of webinar on the YouTube channel (<https://www.youtube.com/watch?v=IMs6k8PUrN8>). The webinar was combined with a seminar held in the conference hall of the Institute of Biology. C.a 20 botanists and zoologists from the Institute attended the seminar in conference hall.

The masterclass “The concept of “Open Science” on the example of biodiversity primary data publication through the global portal GBIF.org: theory and practice” for students and teachers of Perm State University was held by Ivan Chadin on 13.05.2019.

Prior to the end of the project, the results will be presented in the preprint in Russian, which will later be sent for publication in an open access peer-reviewed journal, as well as in form of report at the scientific conference.

Mid-term evaluation findings and recommendations for the remaining project implementation period

This should be the main section of the report, covering for example:

- *An evaluation of the project activities by the mid-term and their outputs/deliverables*
- *Any feedback on the project's relevance from the partners and stakeholders*
- *Comments on the project implementation, its efficiency and effectiveness*
- *The management arrangements for the project, including support from the GBIF Secretariat*
- *Any reflection on the mid-term evaluation itself that could help inform the project's final evaluation and final report*
- *Areas of success to build on during the remainder of the project implementation*

This section is also an opportunity to draw out the main lessons from the project experience that could be applied in other contexts, including any best practice that others in the GBIF community could apply.

Try to clearly document any changes to the project plans that will be made based on the findings of the mid-term evaluation. Please discuss any substantial changes with the GBIF Secretariat (dschigel@gbif.org). In addition, please outline any recommendations for the GBIF Secretariat or the community to reinforce the initial successes of the project.

All project activities fulfilled up to date have been carried out according to the plan stated in project proposal. Digital images of herbarium labels began to be obtained with the help of a homemade planetary scanner since the 01 February 2019. The first version of database and the web interface were ready to 24 February 2019 with the help of Django framework. About 2000 labels images were uploaded to the database and all project participants began to decrypt the label data. The workflow of data entering was modified during project implementation and after several trials and errors the high productivity workflow was

implemented. Entering data from labels is basically manual labor and therefore, we have applied to it the well-known Frederick Taylor principles of increasing the productivity of manual labor. The “Scientific management” principles were implemented with modification of web interface. All labels information were splitted in portions and for every portion the series of special web-forms were created: form to input the elements of the collection date; form for entering species names; form for entering the names of collectors and determinators. Since labels are stored in the catalog in a certain order, the preservation of this order when entering into label images database allowed to apply another method of acceleration of manual processing: the “Copy from previous” button. This button speeds up data entry by several times, keeping the ability to quickly modify incorrectly entered data. The most difficult to decipher labels data is the georeferencing of occurrence on the base of text description.

This work was simplified by combining labels into groups by the name of the collector, and the year of collection of samples. It is known that botanists make most of their routes by walking and in the same date they usually work in radius 5-10 km. So it is usual to have tens of labels with the same or close geographic coordinates that were collected in one day by same collector.

Up to 27 May 2019 there are 11,771 labels images uploaded to the database and 11,257 of them were digitized with minimum set of fields: catalog number, species name, collection date, geographic coordinates, names of persons who collected and identified the species in the sample. This number of digitized labels will allow us to publish about 13,000 mosses occurrences.

The first version of dataset with 4,991 moss occurrences were published and registered in GBIF(Zheleznova G, Shubina T, Rubtsov M, Litvinenko G, Chadin I (2019). SYKO Herbarium Mosses Collection. Version 1.2. Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences. Occurrence dataset <https://doi.org/10.15468/yjdjs4> accessed via GBIF.org on 2019-05-27). All digitized data were checked by bryologists. The geographic coordinates were checked for gross error with the help of occurrence location visualization in QGIS software.

The intermediate results of the project were presented to the professional community in two forms: masterclass “The concept of “Open Science” on the example of biodiversity primary data publication through the global portal GBIF.org: theory and practice” for students and teachers of Perm State University on the regional scientific conference “Symbiosis 2019” (Perm city, 13.05.2019) and in the form of webinar “Digitization of herbarium label data and their preparation for publication in GBIF. Practical case of bryophytes collections of the SYKO herbarium”. The feedback of this events showed that researchers in the field of

biodiversity have a great interest in the possibilities of publishing their data through GBIF. At the same time, the level of awareness about the procedure of publication and use of data with help of GBIF remains relatively low. The feedback of main stakeholder – Institute of Biology of Komi Science Centre of the Ural Branch of the Russian Academy of Sciences the was very positive. Our colleges were very interested in possibilities to publish their primary data through IPT instance of our Institute. The director of Institute Svetlana Degteva (she is professional botanist) was impressed by the productivity of project team. The SYKO herbarium has several collections (fungi, lichens, algae) with number of specimens comparable with number of labels digitized by our team in 4 months. The feedback given showed that it is necessary to describe our experience in the form of scientific publication. The software tools developed for the project should be documented and published with source code in one of open repository (e.g. GitHub).

Annex – Sources of verification

Sources of verification are for example links to relevant digital documents, news/newsletters, brochures, copies of agreements with data holding institutions, workshop related documents, pictures, etc.

1. Zheleznova G, Shubina T, Rubtsov M, Litvinenko G, Chadin I (2019). SYKO Herbarium Mosses Collection. Version 1.2. Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences. Occurrence dataset <https://doi.org/10.15468/yjds4> accessed via GBIF.org on 2019-05-27.
2. The Program of XI All-Russian Congress of young scientists-biologists with international participation: "Symbiosis–Russia 2019» <http://imbiocom.ru/wp-content/uploads/2019/05/PROGRAMMA-Simbioz-Rossiya-2019-1.pdf>
3. The webinar “Digitization of herbarium label data and their preparation for publication in GBIF. Practical case of bryophytes collections of the SYKO herbarium” recording: <https://www.youtube.com/watch?v=TMenzugsCaY>. Trimmed video:

h

t

t

p

s

:

/

/

w

w

w