

Russia2019_12

Mobilization of biodiversity data on soil mesofauna across European Russia biomes

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

Executive summary	1
Contact information	2
Introduction	2
The project and its objectives	2
Project activities completed by mid-term.....	4
Project communications	7
Mid-term evaluation findings and recommendations for the remaining project implementation period	7
Annex – Sources of verification	9

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 first part of the project was predominantly devoted to the strategical planning and developing of the most efficient practices of data collection and standardization. The developed practices will serve as the practical guide for the second phase of the project and future work on data standardization for soil zoologists in Russia and beyond. The following steps were completed: (1) designing the road map of the project; (2) establishing the data

collection and standardization protocols; (3) collection of data; (4) standardizing and publishing first dataset online; (5) planning of the cross-institutional workshop; (6) designing the information web-page for the project. Detailed explanations are given in the report below.

Contact information

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

Anton Potapov

A.N. Severtsov Institute of ecology and evolution Russian academy of sciences

Leninsky prospect 33, 119071, Moscow, Russia

Project coordinator

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.*

Current report compiles achievements and issues emerged during the first phase of the biodiversity data mobilization project Russia2019_12. Report is based on the communication between the involved researchers during their joint work on the project activities and is seen as a practical guide for the second phase of the project and future work on data standardization for soil zoologists in Russia and beyond.

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 aims at aggregating and standardizing biodiversity data on the two most abundant groups of soil-dwelling arthropods in European Russia – springtails (Entognatha: Collembola, ca. 30,000 occurrences) and oribatid mites (Arachnida: Sarcoptiformes, ca.

15,000 occurrences). The data was accumulated by two soil ecologists and taxonomic experts (Dr. Natalia Kuznetsova and Dr. Andrey Zaitsev) from a range of biomes from the North to the South of European Russia. In addition, a cross-institutional workshop is organised and training activities are performed to increase awareness of the open science concept among Russian soil zoologists. The project is implemented during February – September 2019 and involves taxonomic experts (Dr. Natalia Kuznetsova and Dr. Andrey Zaitsev) and data ecologists (Dr. Anton Potapov, Dr. Natalya Ivanova and Maxim Shashkov).

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 by filling in the column 'State by mid-term report'. 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
Collembola communities of European Russia (collection of Natalia Kuznetsova)	Collembola/ across European Russia/1975- 2018	30 000	Soil core heat extraction	10-1000m	ca. 95% Digitalized, not standardized	9114 occurrences were standardized and published. About 25 000 more occurrences is being digitalized and standardized

						following the developed protocols
Oribatida communities of European Russia (collection of Andrey Zaitsev)	Oribatida/ across European Russia/1968-2010 (including unpublished and published literature data of A. Zaitsev's supervisor D.A. Krivolutsky.	15 000	Soil core heat extraction	10-1000m	ca. 95% Digitalized, not standardized	About 15 000 occurrences is being standardized following the developed protocols for Collembola

b. Other deliverables

Description	State by mid-term report
At least 1 open cross-institutional seminar/workshop will be conducted at the Institute of Ecology and Evolution RAS or at the Moscow State Pedagogical University, involving soil zoologists from several institutes in Moscow and other cities. The following topics will be addressed: (1) the open-science concept, (2) data standardization, (3) the GBIF	Information on the workshop "Diversity of soil animals in Russia: publication and effective use of initial data" was distributed among the soil zoologists in Russia. Preliminary program of workshop was

platform, (4) publishing datasets through GBIF.	defined. Registration form is open (https://forms.gle/Fjrbf7nTHtDWDkSr5). The workshop is planned for 29-30 of August, 2019.
Individual on-line consultations for soil biodiversity experts across Russia. The addressed topics will be as listed above.	Information about the possibility was distributed recently. Registration of Moscow state pedagogical university as a new GBIF publisher was initiated.
Advertising the initiative and promoting GBIF Russia to a broader scientific audience by posting general information, web links and methodological materials on institutional web sites.	Information web page about the project was designed and currently is under development. Information for online resources about the project is being prepared and refined.

Expected milestones by mid-term report

Milestone	Status by mid-term progress report
At least one dataset has been published to GBIF.org	Dataset including 9114 occurrences were standardized and published (https://www.gbif.org/dataset/f2e0dd25-5b04-4a90-bc86-3144659d5574)

Project communications

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

Information web page about the project will be developed and distributed among colleagues within and outside Russia through e-mail and social media (e.g. Twitter). Potential prospects of further data standardization during joint scientific projects and publications will be discussed during upcoming conferences (e.g., “International seminar on Apterygota” in Paris <https://isa10.sciencesconf.org>, and “Biodiversity next” in Leiden <https://biodiversitynext.org>).

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.

The first part of the project was predominantly devoted to the strategical planning and developing of the most efficient practices of data collection and standardization. The developed practices will serve as the practical guide for the second phase of the project and future work on data standardization for soil zoologists in Russia and beyond. The following steps were completed:

(1) Designing the road map of the project. Current stage of the data was revised and priorities were defined accordingly. At the initial stage, well-managed part of the data was used to develop a data standardization protocol. In parallel, continuous workflow on gathering and checking of various data from papers and excel spreadsheets were established. Further, workshop program, web page content and information distribution

strategies were discussed. All collective discussions were done online via e-mail and *google documents*.

(2) Establishing the data collection and standardization protocols. For data collection we first used volunteers and technical assistants to compile data from different sources in standard type *excel* spreadsheets (community matrix form). Next, spreadsheets were checked by corresponding taxonomic experts for inconsistencies and corrected, if necessary. Third, corrected tables were checked, cleaned, and standardized in Darwin core (according to the GBIF template) in *R* with packages *reshape*, *reshape2*, *dplyr* and *taxize*. Finally, standardized data comprised of three files, (a) occurrences, (b) sampling events and (c) dataset metadata, were delivered to an IPT administrator for publishing in GBIF.

(3) Collection of data. Data on Collembola were collected from the personal archives of Dr. Natalia Kuznetsova for the period of 1975-2018. The data comprised both paper and digitalized sources. For development of the standardization protocol, the well-managed dataset on seasonal dynamics (1991-1997) of collembolan communities from Moscow region has been chosen. Other data sources were listed and prioritized for standardization in account to have a broad range of geographical regions and habitat types. Data on Sarcopiformes were collected from the personal archives of Dr. Andrey Zaitsev for the period of 1968-2010. The data comprised digitalized sources that were prioritized for standardization in account to have a broad range of geographical regions and habitat types.

(4) Standardizing and publishing first dataset online. The dataset on seasonal dynamics of collembola was processed in *R* software to fit the data into the GBIF template (<https://github.com/gbif/ipt/wiki/samplingEventData#templates>). Specifically, we changed the data format from a matrix to a long table, added missing information from the metadata, renamed the variables, changed the format of data whenever it was necessary (e.g. for dates), created sampling event table and metadata table. The resulting table was checked and cleaned using diagnostic summaries and further published through the IPT at the Institute of Mathematical Problems of Biology RAS.

(5) Planning of the cross-institutional workshop. Workshop was planned as a basis for efficient communication between the project stakeholders and community of soil zoologists

in Russia. During joint work on the workshop description and program, project stakeholders agreed on the following main topics that will be focused on the workshop:

- Principles and advantages of the 'open science'. Where and how publish data? What is gbif.org?
- Data on soil animal diversity of Russia: state of the art and perspectives of establishing and developing of a national web-portal.
- Biodiversity data standards and data quality.
- How to search and download published data from online resources?
- How to prepare your initial data for publication?
- How to publish your data (for example, in gbif.org)?

Preliminary program of the workshop is available under

https://drive.google.com/file/d/1i7PnYO85e8JEeWSB_YzDDcsUlnHBVYEg/view?usp=sharing. Workshop is titled "Diversity of soil animals in Russia: publication and effective use of initial data" and is planned for 29-30 of August, 2019. Registration was opened recently (<https://forms.gle/Fjrbf7nTHtDWDkSr5>).

(6) Designing the information web-page for the project. After joint online discussions, the following working packages were included in the terms of reference, according to the web page contents: news tray, descriptive page, description of related projects, contacts of stakeholders and web resources (databases and learning materials), content modification interface. The development will be initiated in early June. Textual descriptions for the information web page and for web pages of involved organizations are under preparation.

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.

Published dataset with 9114 occurrences of Collembola:

<https://www.gbif.org/dataset/f2e0dd25-5b04-4a90-bc86-3144659d5574>

Preliminary program of the workshop:

https://drive.google.com/file/d/1i7PnYO85e8JEeWSB_YzDDcsUlnHBVYEg/view?usp=sharing
[g](#)