

III Russian National Conference  
“Information Technology in Biodiversity Research”

## Establishing the regional center on biodiversity data mobilization in the Northwestern Siberia (Russia)

Nina V. Filippova\* (a), Ilya V. Filippov (b), Elena D. Lapshina (c)

(a) ORCID: 0000-0002-9506-0991,

Yugra State University, 16 Chekhova Street, 628012 Khanty-Mansiysk, Russia

(b) ORCID: 0000-0002-3639-553X,

Yugra State University, 16 Chekhova Street, 628012 Khanty-Mansiysk, Russia

(c) ORCID: 0000-0001-5571-7787,

Yugra State University, 16 Chekhova Street, 628012 Khanty-Mansiysk, Russia

---

### Abstract

The initiatives on data mobilization in biodiversity science have been steadily on the rise throughout the past few decades. The Global Biodiversity Information Facility (GBIF) supports this development and enables open access to biodiversity data on an international scale. In Russia, with its immense territory, the initiative is following a unique course; while some regions are well-represented, others are hardly ever mentioned. The operational unit in the GBIF network is a GBIF Participant Node, which coordinates a country or organization to collectively develop an infrastructure for delivering biodiversity information. The Nodes, however, are formally established in countries which become Voting or Associate participants in GBIF, which is not yet the case for Russia. The large size of our country would justify the need to create sub-nodes of the next level. This publication is describing process of establishing a regional initiative in the northern part of West Siberia – within the administrative boundaries of the Khanty-Mansi Autonomous Okrug-Yugra, Yamal-Nenets Autonomous Okrug and Tyumen region. The regional initiative started in 2018, although some biodiversity information initiatives had been developing earlier (since 2012). The work currently covers four main areas of GBIF-nodes responsibilities and services: coordination the landscape of initiatives, support of biodiversity data mobilization, biodiversity data analysis and use, and support of data management and curation. These services are described in more detail in the present paper, and problems in the development of individual services are discussed. By the end of 2020, there are 11 registered publishing organizations and 37 published datasets from the Siberian Northwest. The total number of published records is about 120 000, with 30% published by the Yugra State University; 20% are observations from iNaturalist.org; all organizations in Northwestern Siberia as a whole represent about 50% of the published records.

*Keywords:* digitization, GBIF, global portals, open data

© 2020 Nina V. Filippova, Ilya V. Filippov, Elena D. Lapshina

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published by Ural Federal University and peer-reviewed under responsibility of BDI-2020 (III Russian National Conference “Information Technology in Biodiversity Research”)

---

\* Corresponding author. *E-mail address:* filippova.courlee.nina@gmail.com

## **Introduction**

The history of biodiversity research and biodiversity informatics initiatives in Northwestern Siberia was described by the authors in previous publications (Filippova et al. 2017, 2019). In the past two years, the initiative has become more organized and now covers a large portion of the responsibilities and services published in the document Establishing an Effective GBIF Participant Node (GBIF Secretariat 2019). The aim of this paper is to describe the regional program in more detail in relation to the abovementioned document, to show examples of existing services, and to discuss problems and plans for the future. Inspired by this document, we will focus on its structure as well as on the initiatives of the national GBIF community (<http://gbif.ru>).

A GBIF Participant node is a team designated by Participant Node – a country, economy or organization that joins GBIF by signing a Memorandum of Understanding ([www.gbif.org/document/80661/gbif-memorandum-of-understanding](http://www.gbif.org/document/80661/gbif-memorandum-of-understanding)). Russia is not currently a Participant country, but we have a working group that promotes the ideas and already, to a large extent, plays the functional role of a Participant Node. Since the “GBIF Participant Node” is not formally applicable to this community in Russia, we will use the term “GBIF center” to refer to the respective initiative or team at the national and regional level.

## **Coordinating the landscape of biodiversity-related initiatives including participating in the GBIF network**

The collaboration between the various organizations and initiatives related to biodiversity data, as well as with GBIF itself, is a key moment of the development of the biodiversity data mobilization center. In the Northwestern Siberia (NWS), there has been some progress in coordinating data mobilization, but many potential contributors remain unaware and thus uninvolved, especially in the Yamal-Nenets Autonomous Okrug.

A broad coverage of organizations and initiatives is essential for successful data mobilization. Regionally, due to the recent history of the region's industrial development, the number of relevant organizations is comparatively low. All organizations that possess and manage biodiversity data answer to three main ministries: the Ministry of Science and Education of Yugra (including 3 universities and research institutes), the Ministry of Natural Resources of Yuga (about three dozen protected areas), and the Ministry of Culture of Yugra (museums, which also maintain biological collections). The data users are scientific and environmental institutions, the Department of Subsoil Use and Natural Resources of Yugra (<https://depprirod.admhmao.ru>), the Service for control in the field of environmental protection, wildlife and forest relations of Yugra (<https://prirodnadzor.admhmao.ru>), the Yugra Research Institute of Information Technologies, as well as private and state oil and gas companies. The data is also used by educational institutions and the general public.

We were able to reach certain collaboration agreements with biodiversity data providers, related to publishing biodiversity data openly in GBIF and using aggregated GBIF data for biodiversity monitoring programs, as well as developing educational infrastructure in the field of biodiversity informatics in the region. Most of the biodiversity research organizations and nature protection organizations in Yugra had been registered in GBIF and have begun data publishing. We are holding regular training workshops on data mobilization (<https://nwsbios.org/events>) and submitted a series of applications for funding the initiative, including a recently supported

application for the Presidential Grant aimed at an educational program on data mobilization in the region (No. 20-2-014584).

We are continuously negotiating with stakeholders and decision-makers on open data publishing and use of aggregated open data. The Department of Natural Resources of Yugra showed interest to begin the initiative on standardization of data collection from protected areas, as well as in the use of the collected data.

Collaboration with public stakeholders such as educational organizations started with the region's protected areas and the Forestry schools. Public outreach events are organized in collaboration with educational and nature protection organizations using iNaturalist platform (<https://www.inaturalist.org/>) in the form of bioblitzes and challenges.

Another project on working with data mobilization in Russia is the "Eurasian chronicles of nature" (<http://chronicleofnature.com>), which collaborates with the regional data mobilization center.

Unfortunately, an important sector that is actively developing in Northwestern Siberia – the oil and gas companies currently remains largely unaware of the initiative. These companies are providing environmental monitoring in the fields of gas and oil extraction and can act as stakeholders, as well as participate in decision-making and funding.

The northernmost part of the area (Yamal-Nenets Autonomous Okrug) is currently the least involved in data mobilization initiatives. Stakeholders in this area are government organizations, nature conservation areas and the Scientific Center for the Study of the Arctic, which coordinates research activities in the region.

Communication with the GBIF national community is supported through its Facebook group (<http://www.facebook.com/groups/477172382940229>), Google groups ([gbif-in-russian@googlegroups.com](mailto:gbif-in-russian@googlegroups.com)) and workshops on data mobilization and publishing (<http://gbif.ru/gbifruTEAM>). Communication with the GBIF Secretariat is supported through publications and during the regular national conference "Information Technologies in Biodiversity Research" (<http://gbif.ru/conf>).

The working potential of the mentors community is enhanced at national workshops and regional educational events (<https://nwsbios.org/events>). For instance, a data publishing workshop was held for the museums, universities and protected areas in Khanty-Mansiysk in the spring of 2019 ([https://nwsbios.org/collections\\_seminar\\_2018](https://nwsbios.org/collections_seminar_2018)). The first student workshop was organized as part of a field practice course in July 2019 ([https://nwsbios.org/yugra\\_workshop](https://nwsbios.org/yugra_workshop)). Each of the registered publishing organizations has appointed a contact person (future mentor) for coordination of data mobilization initiative within the organization.

We set up a website for the regional initiative, containing information about the project and its news, as well as training materials and analyses of the mobilized data (<https://nwsbios.org>). The social network group serves to outreach a wide audience (<https://vk.com/nwsbios>). There is a mailing list to keep the community informed about current activities and results and the messages are saved in the website blog (<https://nwsbios.org/blog>).

### Supporting biodiversity data mobilization

To ensure successful data mobilization in the Northwestern Siberia, a technical infrastructure has been installed on the server of Yugra State University. We are using SPECIFY software for managing biological collections databases (<https://www.specifysoftware.org>, <http://bioportal.ugrasu.ru>) and Integrated Publishing Toolkit software (IPT, <https://www.gbif.org/ipt>, <http://ipt.ugrasu.ru:8080/ipt>) to publish data in GBIF. At present, five regional collections have been imported in SPECIFY, the rest are on hold (herbaria and other collections of protected areas, universities and private collections of individual researchers). Eleven organizations are registered in the IPT installation of the YSU, all of them from Northwestern Siberia. In total, 37 datasets were published via this IPT installation (Table 1).

**Table 1.** The organizations of Northwestern Siberia registered in IPT of Yugra State University for 10.11.2020.

Organization	Number of published datasets	Number of published records	Number of citations
Biological collection of Yugra State University	9	45 695	14
Malaya Sosva State Nature Reserve	10	6 969	13
Natural park Kondinskies lakes named after L.F. Stashkevich	15	5 396	14
Surgut State University	6	3 735	1
Natural park Numto	10	1 226	0
Yuganskiy State Nature Reserve	2	1 096	0
Institute of the problems of Northern development	1	262	0
Tobolsk complex scientific station of Ural Branch of the Russian Academy of Sciences	1	253	0
Regional historical, cultural and ecological Center	1	165	7
Museum of Nature and Man	0	0	0
United Directorate of Reserves of Taimyr	0	0	0

The development of a DwC export from KAMIS system (a national information system for the museum collections <https://kamis.ru>) has been started to mobilize the biological collections stored in museums.

We are using several channels to source the funding for the regional initiative, e.g. through grants and awards. In the past, the initiative was supported by the Department of Nature Resources of Yugra. The initiative is currently funded by the Yugra State University with the university grant (No. 13-01-20/39), as well as through the EU INTERACT project for the development of the Mukhrino field station (No. 730938). In 2020, the project “Biodiversity data digitization and mobilization workshop in Yugra” received support in the form of the Presidential Grant (No. 20-2-014584). In 2018–2020, we submitted nine applications to the Russian Foundation for Basic Research (RFBR) and social initiatives foundations (Governor's Grant, Presidential Grant, etc.). Technical support of data mobilization is provided in

the form of training seminars and through the helpdesk (individual consultations). The SPECIFY installation is developed in cooperation within the SPECIFY Collections Consortium (Yugra State University has been a member of the consortium since 2020, <http://www.sustain.specifysoftware.org>). The installation and management of the regional IPT is supported by the national GBIF community. Currently, two people are involved in providing technical support. Hopefully, the number of these mentors will grow after organization of training seminars in the future (with a basic aim to reach at least one person in each organization). We use educational materials on the following topics in the field of biodiversity informatics:

- Darwin Core Standard ([http://gbif.ru/DwC\\_spec](http://gbif.ru/DwC_spec))
- How to create a dataset in GBIF (<http://gbif.ru/datapublish>)
- How to publish a dataset in GBIF via IPT (<http://www.gbif.org/ipt>)
- SPECIFY collections management system (<http://gbif.ru/specify>)
- iNaturalist portal use ([http://gbif.ru/inat\\_publish](http://gbif.ru/inat_publish))

Instructions and manuals on data mobilization (text- or video presentations) are published online in Russian (some are prepared by the national GBIF community; others have a regional context and are prepared locally).

The development of the open data culture in the region had been ignited by several examples. Recently, we published a data paper in collaboration with several colleagues working in the region (Filippova et al. 2020a). The growing number of citations of the published datasets had been advertised in the community correspondence. The raising efficiency of usage of scientific collections was confirmed by the Fungarium of Yugra State University, which is exchanging its collections regularly since it has been represented in GBIF. The culture of open data in citizen science (iNaturalist) is motivated by the examples of rare species observation and total impact of citizen science in biodiversity research (<https://nwsbios.org/inat-yugrabilio>). Some attempts are made to include the publication of datasets in GBIF in the financial credits for the universities and other research organizations.

The call for open data, being the key part of GBIF, is stressed in every presentation and publication of the regional initiative. Specific actions that we took in this direction were: 1) the Rules for authors of the journal *Environmental Dynamics and Global Climate Change* (published by Yugra State University) include a paragraph on the raw data publication as an electronic application in GBIF (<https://edgccjournal.org/EDGCC/about/submissions>) (for example, Lapshina et al. 2018); 2) as a mandatory rule when visiting the Mukhrino field station, a metadata should be provided by the end of working period (for biodiversity research – in DwC standard). An important outcome of open data culture is data verification and countering falsification of the research results. This aspect should be accounted for when shaping the policies of journals, grant proposals and research evaluation in general.

The initiative is promoted to a wide audience of citizen scientists through the development of the iNaturalist network in the region (<https://nwsbios.org/inat-yugrabilio>). In recent years, we organized a number of regional events, including an online workshop for the community members on the development of a regional umbrella project on iNaturalist. Among other things, a number of ecological education departments of the region's nature conservation areas are involved in the management of the regional iNaturalist projects and organization of educational iNat-events.

### **Supporting biodiversity data analysis and use**

Due to the relatively recent origin of the initiative, biodiversity data analysis remains insufficiently supported in the region, with only a few examples to be provided.

A prerequisite for efficient data use is ease of access. Therefore, for local tasks, organization of a regional portal is important. Currently, the regional collections portal in SPECIFY contains five collection databases and other collections digitization only begins. Data can be obtained directly from local SPECIFY as well as from global portals, via GBIF and iNaturalist, etc. Instructions have been prepared by the national community on the use of data from these systems; training courses are held at seminars and in individual consultations. Promotion of correct citation of data is taught during workshops and through the helpdesk (<https://gbif.ru/datause>).

The coverage of the area with mobilized data has not yet been assessed. The need for such analysis exists, since the territory is currently being actively developed by oil and gas industries. An example of such analysis for a particular group (fungi) is available in the publication (Filippova et al. 2020a). As the result of literature data digitization, a dataset was published in GBIF (Filippova et al. 2020b) and subsequent data paper analyzes the study extent.

An example of the use of data for compiling national and regional checklists is represented by the same paper (Filippova et al. 2020a). A complete electronic checklist of fungi for the Northwestern Siberia was compiled based on digitized literature records.

Providing data for political decision-making and conservation priorities is in future plans of the regional center, including the program of reprint of the Red Data Book. From the very beginning of the development of the initiative, the program for the Red Data Book of the Department of Natural Resources was in line with the data mobilization program. Thus, archives of protected species records were digitized and organized in a local information system (publication in GBIF in future plans). This system is currently closed and the Department is relying on support of the data mobilization center to provide data on rare and protected species via GBIF.org.

### **Supporting biodiversity data management and curation**

There is a wide range of different software solutions for data management, including SPECIFY for managing database collections, GBIF species matching tool (<http://www.gbif.org/tools/species-lookup>) for validating taxonomic lists, GBIF data validator (<http://www.gbif.org/tools/data-validator>) to verify that a dataset complies with GBIF requirements, etc. The regional community is promoting the use of these tools in the region and the whole country.

Maintaining data quality to meet GBIF standards is also part of the regional initiative (<http://www.gbif.ru/dataquality>). In addition to courses and technical support, the standardization of monitoring data is in nearest future plans for the region. Creation of templates and databases to simplify and standardize data input (based on SPECIFY software) will help improve data quality.

### **What do regional center need to be effective?**

The document Establishing an effective GBIF participant Node (GBIF Secretariat 2019) provides key points for its successful development, which includes functionality and technical capabilities.

*Functionality* stays for ability to formulate strategies, plans and policies for community development. The regional community has informal goals and objectives, as well as an informal group of stakeholders (summarized at <https://nwsbios.org>). For the near future, we aim at creating a collaboration agreement between stakeholders and setting up a council of representatives; we also plan to develop short and long-term plans. Currently, about 10 organizations are participating in the initiative that have confirmed their cooperation by letters of support (<https://nwsbios.org/documents>).

*The technical capabilities* include software products and the necessary server hardware for publishing and storing data, as well as technical documentation (manuals) for publishing data and improving data quality. A software product for publishing data in GBIF (IPT) and software for data management of biological collections (SPECIFY) had been installed on the YSU server. In addition, technical capabilities include software tools for managing data for particular regional tasks (for example, monitoring reports, special checklists, rare species reports, etc.), which will be developed in the future.

The four main qualities that a GBIF node should meet in order to fulfill its mission: neutrality, leadership and initiative, service orientation and adaptability. Establishing the regional data mobilization center, we are trying to meet the above qualities, but some problems may be discussed.

*Neutrality* is essential in relation for the benefits that a node provides for hosting institution in term of data use. This implies neutrality with all stakeholders, regardless of preferences and priorities. Until now, we have not experienced any problems with this issue, apart from the widespread reluctance of authors to open data before publication. We hope that building a culture of open data and correct citation practices will help address this problem.

*Leadership and initiative* are recognized as important qualities in the GBIF Node mission. These qualities should help to change the habits and practices associated with the open publication and storage of biodiversity data. Information and Internet technologies are rapidly developing and biologists often find it difficult to make new practices. It also important to understand that the GBIF Node performs its functions at the complex landscape, including scientific, governmental, educational, public, etc. Leadership and initiative coming from young employees of the “basic level” must therefore be supported by managers who are familiar with the landscape and ready to provide lobbying and introducing the initiative into existing institutions.

GBIF Nodes are created to support communities of people and institutions and therefore *focus on service*, representing GBIF and the broader community of biodiversity informatics to the wider community of people.

The last property – *adaptability*, is associated with the need for a node to respond and adapt to different stakeholders, sponsors and partners. In order to meet this requirement it is important to have an evolving, relevant development strategy. The intent to formulate such a strategy for the region was described above.

## Conclusion

As the result of organized systematic work on the regional GBIF initiative, we already achieved some results and gained valuable experience. This includes four main areas of the GBIF-nodes responsibilities and services: coordinating the landscape of initiatives, supporting biodiversity data mobilization, biodiversity data analysis and use, and supporting data management and curation. These services were described in more detail in the paper, and the problems were discussed.

The total number of mobilized records for the Northwestern Siberia (within the administrative boundaries of the Khanty-Mansi Autonomous Okrug-Yugra, Yamal-Nenets Autonomous Okrug and Tyumen region) has exceeded 120 000 by the end of 2020, with approximately 40% published by the Yugra State University; 13% are observations from iNaturalist.org (Table 2). Comparing to other Siberian regions, data mobilization in this region is progressing rapidly, on a par with the Novosibirsk region. However, compared to the Moscow region and some other regions of Central Russia, these values are lower by an order of magnitude. Regional contribution to the common pool of mobilized data currently reaches 60%, the rest being the result of data mobilization of large collections in Russia and abroad.

**Table 2.** The first ten publishing organizations of Northwestern Siberia (within the administrative boundaries of the Khanty-Mansi Autonomous Okrug-Yugra, Yamal-Nenets Autonomous Okrug and Tyumen region) with the number of published records (totaling about 120 000 records for 10.11.2020).

Organization	Number of published records	Percentage from the total number
Yugra State University Biological Collection	40 091	33
iNaturalist.org	24 422	20
Lomonosov Moscow State University	10 839	9
Malaya Sosva State Nature Reserve	6 969	6
Central Siberian Botanical Garden SB RAS	9 358	8
Natural park Kondinskies lakes named after L.F. Stashkevich	5 394	4
Surgut State University	3 735	3
Tyumen State University	3 472	3
PlutoF	2 436	2
Cornell Lab of Ornithology	1 535	1

## Competing interests

The authors have declared that no competing interests exist.

## Acknowledgements

We are very grateful for the fruitful discussions during preparation of the manuscript to Natalia Ivanova and for the help with English translation to Tatiana Bulyonkova. The project was funded by the grant 13-01-20/39 of the Yugra State University, by the Presidential Grant 20-2-014584 and by the INTERACT Virtual Access under EU-H2020 Grant Agreement No. 730938.

## References

- Filippova NV, Filippov IV, Schigel DS, Ivanova NV, Shashkov MP (2017) Biodiversity informatics: global trends, national perspective and regional progress in Khanty-Mansi Autonomous Okrug. *Environmental dynamics and global climate change* 8 (2): 46–56. <https://doi.org/10.17816/edgcc8246-56>.
- Filippova N, Filippov I, Ivanova N (2019) Biodiversity Portal of the Northern Part of West Siberia, Russia. *Biodiversity Information Science and Standards* 3: e37067. <https://doi.org/10.3897/biss.3.37067>.
- Filippova N, Arefyev S, Bulyonkova T, Zvyagina E, Kapitonov V, Makarova T, Mukhin V, Stavishenko I, Tavshanzhi E, Shiryaev A, Tolpysheva T, Sedelnikova N, Ryabitseva N, Paukov A, Zhurbenko M (2020a) Fungal literature records database of the Northern West Siberia (Russia). Occurrence dataset. Version 1.13. Yugra State University Biological Collection (YSU BC). <https://doi.org/10.15468/hfje31> [Accessed on 15.09.2020]
- Filippova N, Arefyev S, Zvyagina E, Kapitonov V, Makarova T, Mukhin V, Sedelnikova N, Stavishenko I, Shiryaev A, Tolpysheva T, Ryabitseva N, Paukov A (2020b) Fungal literature records database of the Northern West Siberia (Russia). *Biodiversity Data Journal* 8: e52963. <https://doi.org/10.3897/BDJ.8.e52963>.
- GBIF Secretariat (2019) Establishing an Effective GBIF Participant Node: Concepts and general considerations. Copenhagen. <https://doi.org/10.15468/doc-z79c-sa53>.
- Lapshina ED, Ganasevich GN, Vasina AL (2018) Redkie rasteniya i rastitel'nye soobshchestva bolot bogatogo gruntovogo pitaniya zapovednika «Malaya Sos'va» (Zapadnaya Sibir') [Rare plants and plant communities of rich fens of Malaya Sosva Nature Reserve (Western Siberia)]. *Environmental dynamics and climate change* 9 (1): 72–92 [in Russian]. <https://doi.org/10.17816/edgcc9020>.