Biodiversity Databases in Russia: Towards a National Portal

Article · December 2016

DOI: 10.1139/AS-2016-0050

2 authors:

Natalya Ivanova
Institute of Mathematical Problems of Biology …

Maxim Shashkov
Institute of Physicochemical and Biological Pro…

Biodiversity Databases in Russia: Towards a National Portal

Natalya V. Ivanova

Institute of Mathematical Problems of Biology - the Branch of Keldysh Institute of Applied Mathematics, Russian Academy of Science, 1, Vitkevicha str., 142290, Pushchino, Moscow Region, Russia

Pushchino State Institute of Natural Sciences, 3, prospect Nauki, Pushchino, Moscow region, Russia

Natalya.dryomys@gmail.com

Maxim P. Shashkov

Institute of Mathematical Problems of Biology - the Branch of Keldysh Institute of Applied Mathematics, Russian Academy of Science, 1, Vitkevicha str., 142290, Pushchino, Moscow Region, Russia

Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Science, 2, Institutskaya str., 142290, Pushchino, Moscow Region, Russia

Pushchino State Institute of Natural Sciences, 3, prospect Nauki, Pushchino, Moscow region, Russia
Max.carabus@gmail.com

Corresponding Author's Name
Natalya V. Ivanova

Abstract

Russia holds massive biodiversity data accumulated in botanical and zoological collections, literature publications, annual reports of natural reserves, nature conservation and monitoring study project reports. While some data has been digitized and organized in databases or spreadsheets, most of the biodiversity data in Russia remains dormant and digitally inaccessible. Concepts of open access to research data is spreading, the lack of data publishing tradition and of use of data standards remain prominent. A national biodiversity information system is lacking and most of the biodiversity data is not available or the available data is not consolidated. As a result Russian biodiversity data remains fragmented and inaccessible for researchers. The majority of Russian biodiversity databases does not have web interfaces, and are accessible only to a limited numbers of researchers. The main reason for lack of access to these resources relates to the fact that the databases have previously been developed only as a local resource. In addition, many sources have previously been developed in the desktop database environments mainly using MS Access and, in some cases earlier DBMS for DOS, i.e. file-server system, which does not have the functionality to create access to records through a Web interface. Among the databases with a web interface, a few information systems have interactive maps with the species occurrence data and systems allowing registered users to upload data. It is important to note, that the conceptual structure of these databases were created without taking into account modern standards of the Darwin Core, furthermore, some data sources were developed prior to the first work version of the Darwin Core release in 2001. Despite the complexity and size of the biodiversity data landscape in Russia, the interest in publishing data through international biodiversity portals is increasing among Russian researchers. Since 2014, institutional data publishers in Russia have published about 140,000 species occurrences through GBIF.org. The increase in data publishing activity calls for the creation of a GBIF node in Russia, aiming to support Russian biodiversity experts in international data work.

Key words: Russia, gbif.org, biodiversity, databases.

Introduction

Russia plays a key role in world biodiversity conservation, including conservation of Arctic ecosystems: 80% of Arctic species diversity is represented in Russia (Climate Change Impacts in the Russian Arctic, Searching for Ways for Adaptation 2009). More than seventeen million square kilometers of the terrestrial area of the Russian territory is comprised of polar deserts, tundra, forest tundra, taiga, mixed and deciduous forests, broad-leaved forests, steppe, semi-deserts and subtropics. Mountain regions cover about a quarter of Russia, and significant territories are wetlands. The diversity of ecosystems translates into high species diversity, including more than 12,500 species of vascular plants, over 1,500 species of vertebrates, and 100,000 species of invertebrates (The National Strategy for biodiversity conservation in Russia 2002). A few hundred years of the exploration of Russian flora, fauna and mycota have generated a great body of biodiversity data. This data is found in different museums and herbaria, and is reflected in literature data from different countries and researchers. Some data are already digitized and organized in databases, but most of the data is disaggregated and presented in different formats (see below). A central national biodiversity system is missing. The need to create and develop such a resource has been repeatedly discussed, and even though data standards and technology is available, little progress has been observed. The largest international open biodiversity information systems: the Global Biodiversity

Information Facility, GBIF (2016), Encyclopedia of Life, EoL (2016), Integrated Digitized Biocollections, iDigBio (2016) and many others use the international data standards developed by the Taxonomic Database Working Group, TDWG (TDWG 2016). Open access technology and data standards allow all interested parties to upload and to publish their data through global portals, and therefore to improve the discoverability of their data, and significantly reduce the cost of the work using literature and collections. All this contributes to the development of international research cooperation.

Many researchers in Russia remain uninvolved in this activity. However, in recent years, the interest in publishing of data through GBIF.org and activity to popularize GBIF in the Russian-speaking environment has appeared. In this paper we summarize biodiversity data mobilization activities in Russia through the description of biodiversity databases, and report progress towards the creation of a national GBIF node in Russia.

1. Review of Russian biodiversity information systems

Here we summarized descriptions of biodiversity databases, using the available information from the literature and Russian Biodiversity information systems via the internet. Even though this summary covers the key biodiversity data resources in Russia, many personal, institutional and project databases remain unknown to us, or are inaccessible through the internet (e.g. Zeltyn and Insarov 1993; Knyazeva et al. 2007; Golub et al. 2009; Kryshen et al. 2009; Chernenkova et al. 2012 and many others).

The main reason for a lack of access to these resources relates to the fact that the databases have previously been developed only as a local resource. The authors did not wish to share their data and only announced in publications the fact of the existence of the database. Also many resources have previously been developed in the desktop database environment (more often Microsoft Access), which did not have the functionality to create access to the data through Web interface. Furthermore, descriptions of such hidden data resources have not been published in literature. While biodiversity papers do mention databases used for certain analyses, descriptions of the database structure, software, programming languages and other details are much less visible or missing.

We have reviewed Russian biodiversity databases for the following characteristics: type of data, data standard, number of records, availability of the primary data, and web interface. Based on the content and primary foci of the reviewed systems, they were divided into three groups: occurrence databases (Table 1a) taxonomic databases (Table 1b), and digital collections (Table 1c).

1.1 Occurrence databases

A huge amount of different resources in terms of volume, quality and functionality have been developed over the last 20–25 years. In this section the databases on species distribution are described.

While many databases of various scales exist and operate in isolation, technical specifications of the databases (such as the structure, data formats and software used) are typically described very poorly.

The analysis of available metadata showed that information about the occurrence of different taxonomic groups of plants and animals is available via the internet (Table 1a). Unfortunately, most of these resources have a local data standard even in the case where different resources contain similar data (e.g. Morozova and Borisov 2010 and Dahlke et al. 2014 or Koropachinsky et al. 1999; Abdrahimov et al. 2011; Biodiversity of Altai-Sayan Ecoregion 2016), and the conceptual structure of the databases were created without taking into account modern international standards (Wieczorek et al. 2012).

Database topics are often repeated. For example, invasive alien species are a very important group and a common target for the creation of the databases. One of the most significant initiatives, an information system *Alien species of Russia* (2016) is maintained by the Institute of Ecology and Evolution Russian Academy of Science (RAS) and covers plant species, insects, fishes, and mammals. Most of the data on the distribution of alien plant species are presented on the Web-Oriented Geoinformation System of *Alien Plant Species of European Russia* (Morozova and Borisov 2010). Information on alien plant species can also be found in the information system *The Black Data Book of Russian Flora* (2016). Furthermore, databases on individual alien species do exist, for example on *Heracleum sosnowskyi* (Dahlke et al. 2014).

Most of the existing database resources are published as finished closed system with little or no updates after their initial release (Table 1a). This becomes apparent from the absence of any updates after several years. Far too often, the databases developed for research projects and hosted by the commercial web services become forever unavailable soon after the completion of the project (Shashkov and Ivanova 2012).

Spatial information on occurrences is often presented as raster images, not through mapping services (Table 1a). Undoubtedly, such systems contain important information on biodiversity. Especially such data are important for the assessment of biodiversity of insufficiently studied regions. Below we outline a few key examples of such systems.

The information system on vertebrates in Russia (Vertebrate Animals of Russia 2016) includes information about taxonomic status, distribution, recordings of voice and acoustic signals. Unfortunately, the Web interface of this resource provides only the metadata and does not provides access to the database.

Information Retrieval System for Fauna and Flora in Protected Natural Areas of the Russian Federation (2016) integrates distribution data on fish, amphibians, reptiles, birds, mammals, vascular plants, lichens, mosses, hepaticae and anthocerotae from Russian protected areas. Information about the biodiversity of Russian protected areas is also available via the portal *Protected areas of Russia* (2016). This is an ongoing project aiming at a mobilization and generalization of knowledge about the protected areas and providing information support for monitoring these areas.

Information about the biology and distribution of some taxonomic groups is summarized by researchers at the Institute of Biology of Komi. Available data about biodiversity of dipteran insects of the “gnus” (midges) complex (parasitic Diptera) of northeastern European Russia (Panyukova et al. 2014). Data about Siberia is available in *Flora Baikal Siberia* (Abdrahimov et al. 2011), *Biodiversity of animals and plants of Siberia* (Koropachinsky et al. 1999) and *Biodiversity of Altai-Sayan Ecoregion* (2016).

Dynamic updatable maps (based on OSM web-service) includes *Cryptogamic Russian Information System*, CRIS (Melechin et al. 2013), one of the most successful developments of its kind. The system has been developed as tool for convenient storage, organization, integration, visualization and analysis of data on the biodiversity of cryptograms. Currently data from the Polar-Alpine Botanical Garden-Institute of N.A. Avrorin RAS herbarium collection (KPABG) and literature data (mainly for the Murmansk Region) are included in the CRIS. The system is fully developed using open-source software and a multi-user platform. Registered users can upload primary data. Special controlled vocabulary is used for description of species occurrences. These terms, except the general taxonomic and georeference terms, are to describe features specific to cryptogams, such as substrate type. Custom queries with different search criteria can be created. Maps of the occurrences are also available to users. Part of this information is published through gbif.org (Table 2, doi:10.15468/nctfm2, 10.15468/80tu83, 10.15468/yxt7co).

Since the 1990s, biodiversity inventories and surveys, especially on rare species, are carried out by non-governmental organizations in Russia. The data from such initiatives are often more easily accessible than data from the RAS institutes. The crowdsourcing project web-GIS *Birdwatching* (2016) is developed by the Siberian Environmental Center (Novosibirsk). This is an open database: any registered user can upload or download data. Users can upload and store their data on bird species occurrences, and also create vector layers to the map system. Loading data is available in CSV format, KML / KMZ, ESRI shape and MapInfo files, and as doc files (reports). The system supports custom requests. Data collected through *Birdwatching* were used in at least 12 publications in Russian and international journals. This system provides a universal tool for biodiversity monitoring. *Birdwatching* supported work in the Red Data book of the Samara region (The red data book of Samara region 2016), Altai Krai (The red data book of Altai Krai 2016), Krasnoyarsk Krai (only as resource in internal network), projects *Rare plants of Siberia* (2016), *Nestboxing* (2016), *Small Wild Cats of Eurasia* (2016), *Wetland Mammals of Eurasia* (2016). Thus, *Birdwatching* unified data standards and software (Wildlife monitoring 2016) and summarize more than 35,000 observations (Table 1a).

Another category of regional biodiversity data sources is spatial Information systems, such as the one on animal and plant species of Khanty–Mansi Autonomous Okrug, developed by *NextGIS* Ltd.

(UgraBio. Information system of biodiversity of Ugra 2016). The information system is designed for management tasks such as checking for presence of red-listed species in specific areas and to support preliminary scientific inquiries, such as modeling ranges of rare species and help assessment of the degree of rarity of a particular species. The main objective of the application is to show species locations and allow to visualization and quick editing of data. Besides locations, the system allows automatic creation of species ranges from annotated lists (list of species for a specific area, not necessarily a point) assigned to grid cells. Currently, the database includes information about occurrences of Protozoa, Fungi, Plantae and Animalia. Data can be added in the system or downloaded by registered users.

It is noteworthy that while all three open systems support the upload and download of data by registered users, the rules and licenses for the citation of the data is not always described. Each of these information systems use their own data standards. Apparently, these standards have been developed based on the characteristics of the target taxa, and the specific project goals. Use of different standards complicates the interoperability of the systems.

Globally, the Darwin Core standard, DwC (Wieczorek et al. 2012) is a leading global standard for biodiversity data. This standard is followed by the major international biodiversity information systems such as GBIF, EoL, ORNIS (2016) and many others. To the best of our knowledge, only one Russian database is created using the DwC standard, the database *Lobaria* *pulmonaria* occurrences in Russia (Shashkov and Ivanova 2012; *Lobaria pulmonaria* in Russia. Information system. 2016). This online database documents the rare lichen *Lobaria pulmonaria* in Russia. The database is comprised by the data from the literature, herbarium collections, open databases, the authors own field data, and personal communications of researches and is aimed at supporting modeling the population dynamics of *Lobaria pulmonaria*. The detailed descriptions are available for field recordings, but missing for many herbarium and literature-based records. The database is implemented based on an open object-relational database management system PostgreSQL (2016). For a detailed description of the *Lobaria pulmonaria* occurrences about 60 DwC terms were selected. In addition, a number of non-DwC terms were suggested for detailed description of *Lobaria pulmonaria* findings (Fig. 1). Both DwC and non-DwC terms were structured into vocabularies and work tables. Five vocabularies were formed: three for administrative division – countries, regions and administrative districts, the other two – a text description of accuracy of georeferencing and list of host tree species. A few tables were so “updatable vocabularies” in which new records are added in the course of working with the database: name of the datasets, collections and bibliographic references. Detailed description of the occurrences was combined into three logical parts: (1) description of the location, (2) description of the habitat, (3) description of the host tree and *Lobaria pulmonaria* population. If exact (e.g. with GPS navigator) georeferencing was possible, one location (point) corresponded to one biotope (habitat) and to one or more occurrences (Fig. 2a). If the georeferencing was not exact, typically based on a text description without geographic coordinates, location may correspond to multiple habitats (Fig. 2b).

The corresponding dataset was published through gbif.org (Table 2, doi:10.15468/uennht). The dataset is dynamically connected to the source database through a SQL query, the way that greatly simplifies the work with the data in comparison with CSV file loading. For the publication of the data contained in the database the Integrated Publishing Toolkit (IPT) installation of the Institute of Mathematical Problems of Biology RAS (IMPB) was used (Russian GBIF IPT 2016). Through such a setup, all data from the resource database Lobaria.ru (occurrence map and viewing of attributive information of findings) are also available through gbif.org and all updates in the database are rapidly reflected in the dataset on the global portal, which also allows for data downloads and issues digital object identifiers (DOIs) for each download. In 2016, the online version of the system contained data on more than 1,200 occurrences of *Lobaria pulmonaria*.

Despite some progress in promoting of GBIF and data mobilization in Russia, a national system of biodiversity is still lacking, but the existing resource on *Lobaria pulmonaria* distribution can be a prototype of a database component of this system. Only a minor redesign of *Lobaria pulmonaria* database structure would allow scaling up for distribution data on other taxonomic or ecological groups. In principle, the schema of three logical blocks *location* – *habitat* – *occurrence* is already applicable to other taxa. For compatibility with GBIF it will be necessary to use GBIF Backbone Taxonomy (GBIF Secretariat: GBIF Backbone Taxonomy 2016).

A vast amount of biodiversity data in Russia is not digitized and is typically restricted in access. There is a significant overlap in the topics of the individual databases combined with the differences in data format resulting in a blockade of data exchange between different resources. A general lack of maintenance of the project database results in a very short lifetime of these potentially very valuable data products. Commercial web hosting can be considered good practice, the most advanced systems in our review use such approach for data hosting, improving maintenance and software upgrades at the resource. At the same time, several systems demonstrate successful examples of open multi-user databases.

1.2 Taxonomic databases

Among taxonomic databases, “Flora of vascular plants in the Central European Russia” (was developed in the IMPB) and the family of information systems of the Zoological Institute RAS based on the ZOOCOD standard are the best known ones (Table 1b).

The database “Flora of vascular plants in the Central European Russia” (Zaugol’nova and Khanina 1996) was developed for generalization and standardization of taxonomic lists used in different regions of central Russia. Species checklist are included for vascular plants from Moscow, Smolensk, Tver’, Yaroslavl’, Vladimir, Kostroma, Ivanovo, Ryazan', Tula, Kaluga and Bryansk regions. For the packaging of systematic data in the relational structure a special code was developed. Each species has a unique code consisting of nine numbers: the first three represent the family, the next three are used for the genus within the family, and the last three letters of the code represent the species within the genus. This nine-letter code is associated with a table of synonyms and reference tables on the ecology of individual species. More than 120 literature sources were used for the creation of the database. The web interface of the database was developed in 2004 and included checklist of species according to Cherepanov (1995) and synonyms and biological and ecological characteristic of more than 2,300 plant species (Flora of vascular plants in the Central European Russia 2016). The species checklist is available through gbif.org (Table 2, doi:10.15468/96gqtn).

The local taxonomic standard (ZOOCOD) is developed in the Zoological Institute RAS. Construction of the ZOOCOD is detailed by Lobanov and Smirnov (1997). Each specimen has a unique code that describes its systematic position. The classifier concept was developed to demonstrate any hierarchy taxa detail in relational databases. All the taxonomic information systems of the Zoological Institute RAS are based on ZOOCOD. The main sources of the Zoological Institute RAS are the *ZOOlogical INTegrated retrieval system, ZooInt* (Smirnov et al. 1997), the Russian Information system *Biodiversity of Animals*, *ZooDiv* (Biodiversity of animals. Russian information system 2016), information system *Biodiversity in Russia* (2016), and the *taxonomy and collections Interactive database of world insect fauna* (2016). These and other developments of the Zoological Institute RAS summarize data about taxonomy, biology, bibliography of different groups of animals, protists, prokaryotes, fungi and partially plants. For example *ZooDIV* unites 32 systematic databases (>90,000 species) (Biodiversity of animals. Russian information system 2016).

The ZOOCOD standard has been successfully used outside of the Zoological Institute RAS: in the Botanical Institute RAS, the Institute of Ecology and Evolution RAS, at Moscow State University, at Nizhny Novgorod State University and others (Lobanov and Smirnov 1997; Information Retrieval System for Fauna and Flora in Protected Natural Areas of the Russian Federation 2016; Vertebrate Animals of Russia 2016).

Internationally, the Catalogue of Life, COL (2016) is one of the most common basic taxonomic sources for the development of biodiversity databases. However, the Catalogue of Life is not yet complete and covers only 84% of world diversity. Many species recorded in Russia now are missing from the COL database, especially endemics of Russia and the former USSR. The integration of Russian species checklists, already summarized in a relational databases into the Catalogue of Life (and as a result – into the GBIF Taxonomic Backbone) would significantly expand its cover of species diversity and would provide the critical taxonomic foundation for the development of Russian biodiversity databases.

1.3 Digitized collections

The majority of Russian botanical and zoological collections are not digitized. However, a number of digitization projects in the country's largest collections has been launched recently (Table 1c). The project of the Moscow State University (MSU) *Noah's Ark* aims at the creation of multifunctional network storage of biological material. This project includes the digitization of the MSU herbarium (MW). Currently, more than 150,000 plant samples from Siberia and the Russian Far East are digitally available in the internet (The information-analytical system of Depositary of biomaterial of the resource center of Moscow State University 2016). Moreover, the MW herbarium has now more than 1 million of digitized specimens of vascular plants (A. Seregin, pers. comm.). A small part of herbarium of the Komarov Botanical Institute of the RAS (LE) is also digitized (Catalog of specimens found in collections Komarov Botanical Institute RAS 2016). Herbarium data on vascular plants from Russia, Canada, China, Kazakhstan, North Korea, the US, South Korea; herbarium of fungi; collection of basidiomycetes cultures and herbarium and collection of algae are discoverable through the internet. A collection of algae is also digitized in the Institute of Biology, Komi Scientific Center (Collection of microalgae strains in the Institute of Biology of Komi Scientific Centre (SYKOA) 2016) and in the Institute of Physical-Chemical and Biological Problems in Soil Science (Web site of Algal Collection of Soil Science Institute (ACSSI) 2016), part of this information is published through gbif.org (Table 2, doi:10.15468/nt9emp, 10.15468/cm3n7s). Data of fungi, hepatics, lichens and mosses from KPABG collection is available through CRIS (Melechin et al. 2013). Some data on the moss herbarium specimens from different Russian collections are available on the *Arctoa* web site (Arctoa. Project 'Flora of mosses of Russia' 2016). Last year, the work on digitization of the collection of the Zoological Institute RAS of the funds have started. Today specimens of Pogonophora, Coleoptera, Lepidoptera, Flea, Ophiuroidea, Reptilia, Mammalia is available online (Digitized Research Collections of the Zoological Institute RAS 2016). Generalized data of labels and their original images as well as images of exhibits are available (Zoological Institute of Russian Academy of Science 2016).

The majority of Russian digitized collections use local data standards. The original label is not always available. In our opinion, the digitized collection of the Zoological Institute RAS is most similar to modern data standards. Developers use the DwC and similar terms for description of specimens. Original labels are also available.

Digitization of Russian botanical and zoological collections is a very important activity for the global assessment of species diversity and distribution. According to portal *Genetic and biological (zoological and botanical) collections of the Russian Federation* (2016) 148 herbarium collections from 102 cities were present in Russia in 2004. The collection of vascular plants of the LE herbarium contains more than 6 million specimens. Many Russian universities and scientific organizations have their own herbarium collections. The Herbarium of Tomsk State University (500,000 specimens), the Herbarium of Institute of Biology Komi Scientific Centre (180,000 specimens of vascular plants, 40,500 specimens of mosses, 18,000 specimens of lichens), the KPABG Herbarium (100,000 specimens), the Herbarium of Institute of Biology of Inland Waters (>33,000 specimens of water and coastal water plants) are among the most significant collections. Almost all Russian nature reserves and some regional museums also have their own herbarium collections. The Zoological Institute RAS has one of the largest zoological collections in the world, with more than 60 million storage units (Zoological Institute of Russian Academy of Science 2016). The research collection of the Zoological Museum currently includes more than 8 million units (Zoological Museum of Moscow University 2016). Digitizing these data would greatly extend our knowledge of the Russian biodiversity, and would revitalize collection-based research.

2. Russian data available on GBIF.org and Russian GBIF community activities

More than 1.6 million species occurrences records from Russia have been published through gbif.org (1,035 datasets). About 95% of this data were published by institutions outside Russia, most data from UK, USA and Estonia. The first dataset from Russia was published through gbif.org by the Zoological Institute RAS in 2011 (Table 2, doi:10.15468/c9g3nw). Since 2014, Russian publishers made available about 140,000 species occurrences, not only for Russia, but even dozens of countries and territories. At the moment of writing, about 97,000 records for Russia in 15 occurrence datasets and 1 checklist dataset were published through gbif.org by a few Russian institutions (Table 2). Most of the data has been published by the largest Russian data holders: the Zoological Institute RAS, A.N. Severtsov Institute of Ecology and Evolution RAS, and Moscow State University.

Data mobilization through GBIF is carried out through four Russian IPT installations, and half of the datasets are published through an IPT installation hosted by the IMPB. Even though the institute is not a large data holder, this is currently the most active technical support hub for data publishing through gbif.org for Russian Institutes. This IPT installation (Russian GBIF IPT 2016) is associated to the information web-site gbif.ru (2016). This resource contains information about the structure and functioning of the gbif.org portal, Darwin Core standards specification in Russian, information about events connected with GBIF. Gbif.ru is a base for collection and generalization of metadata information about Russian resources on biodiversity.

Gbif.ru is a very important source for informing the Russian research community about the use of data standards and data mobilization. Important activities include workshops, which were organized by the IMPB in 2015 and 2016. The publication of Russian-language articles about modern data standards (Ivanova and Shashkov 2014; Grebennikov 2016) and a mostly completed IPT translation into Russian (mainly by the staff of the Institute of Biology of Komi Republic) will help mobilization of biodiversity data in Russia through gbif.org.

Conclusion

Considerable experience in biodiversity informatics has accumulated in Russia, but a nation-wide portal on biodiversity is lacking. This and the earlier review (Ivanova and Shashkov 2014), as well as the review of the information systems used in Russian nature reserves (Grebennikov 2016) suggests that the creation of a national portal is necessary and should be based on the international Darwin Core standard. Creation of a national GBIF node in Russia, depends on formal participation of the Russian Federation in GBIF through signature of the GBIF Memorandum of Understanding (2010) and would support data authorship protecting and contribute to national and global biodiversity science.

Acknowledgments

We would like to thank Dmitry Schigel and Steffi Ickert-Bond for helpful comments which improved the manuscript, we also thank Ilya Filippov, Victor Chepinoga and Aleksey Vaganov for advices and Tim Hirsch for revision of the manuscript. Natalya V. Ivanova is partly supported by the Russian Foundation of Basic Research, project #16-34-00866.

References:

Abdrahimov, I.S., Petruhin, I.A., and Henkina, A.A. 2011. Knowledge base 'Baikal Siberia Flora' in the system Libretto. Proceedings of IMEI IGU. Mathematics and informatics. **1**: 15–22. (In Russian)

Alien species in the territory of Russia. 2016. Available from: http://www.sevin.ru/invasive/dbases.html [accessed September 12, 2016]. (In Russian)

Arctoa. Project 'Flora of mosses of Russia'. 2016. Available from: http://arctoa.ru/Flora/general.php [accessed September 12, 2016]. (In Russian)

Biodiversity in Russia (BIODIV). 2016. Available from: https://www.zin.ru/BioDiv/index_e.html [accessed September 12, 2016].

Biodiversity Information Standards (TDWG). 2016. Available from: http://www.tdwg.org/ [accessed September 12, 2016].

Biodiversity of Altai-Sayan Ecoregion. 2016. Available from: http://bioaltai-sayan.ru/regnum/eng/index.html [accessed September 12, 2016].

Biodiversity of animals. Russian information system (ZooDiv) 2016. Available from: https://www.zin.ru/ZOODIV/Index.html [accessed September 12, 2016]. (In Russian)

Catalog of specimens found in collections Komarov Botanical Institute RAS. 2016. Available from: http://www.binran.ru/collections/index_e.html [accessed September 12, 2016]. (In Russian)

Chernenkova, T., Basova, E., and Koroleva, N. 2012. Database Biodiversity of Murmansk Region. Biodiversity and Ecology. **4**: 425–425. (In Russian)

Climate Change Impacts in the Russian Arctic, Searching for Ways for Adaptation. 2009. Compillation of basic recommendations prepared by WWF-Russia. WWF-Russia, Moscow.

Collection of microalgae strains in the Institute of Biology of Komi Scientific Centre (SYKOA). 2016. Available from: http://ib.komisc.ru/sykoa/eng/ [accessed September 12, 2016].

Cherepanov, S.K. 1995. Vascular plants of Russia and neighboring countries. Mir i sem'ja, SPb. (in Russian)

Dal’ke, I.V., Chadin, I.F., Madi, E.G., and Zackozhy, I.G. 2014. The collection and view of data on the distribution of invasive plant species on the basis of a programming interface (API) service YandexMap. *In* Biodiagnostics of condition of natural and natural-technical systems: proceedings of a conference, 2. VyatGU Press, Kirov. pp. 98–101. (In Russian)

Digitized Research Collections of the Zoological Institute RAS. 2016. Available from: http://www.zin.ru/collections/collections_en.html [accessed September 12, 2016].

Encyclopedia of Life. 2016. Available from http://eol.org/ [accessed September 12, 2016].

Flora of vascular plants in the Central European Russia. 2016. Available from: http://www.impb.ru/eco/index.php?l=ru [accessed September 12, 2016]. (In Russian)

Genetic and biological (zoological and botanical) collections of the Russian Federation. 2016. Available from: http://www.sevin.ru/collections/ [accessed September 12, 2016]. (In Russian)

Global Biodiversity Information Facility (GBIF). 2016. Available from: http://www.gbif.org/ [accessed September 12, 2016].

GBIF Secretariat: GBIF Backbone Taxonomy. 2016. doi:10.15468/39omei Available from: http://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c [accessed September 12, 2016].

GBIF Memorandum of Understanding. 2010. Copenhagen: Global Biodiversity Information Facility, 20 pp. Available from: http://www.gbif.org/resource/80661 [accessed September 12, 2016].

Golub, V.B., Sorokin, A.N., Ivakhnova, T.L., Starichkova, K.A., Nikolaychuk, L.F., and Bondareva, V.V. 2009. Lower Volga valley phytosociological database. Proceedings of the Samara Scientific Center of the Russian Academy of Sciences. **11**(1(4)): 577–582. (In Russian)

Grebennikov, K.A. 2016. Study of biodiversity of nature reserves of the Russia in the digital age: experience and perspectives. Nature Conservation Research. **1**(2): 1–10.

Information Retrieval System for Fauna and Flora in Protected Natural Areas of the Russian Federation. 2016. Available from: http://www.sevin.ru/natreserves/ [accessed September 12, 2016]. (In Russian)

Information System ZInsecta. 2016. Available from: https://www.zin.ru/projects/zinsecta/eng/ZInsecta.asp [accessed September 12, 2016].

Integrated Digitized Biocollections. 2016. Available from: https://www.idigbio.org/ [accessed September 12, 2016].

Ivanova, N.V., and Shashkov, M.P. 2014. Approaches to Development of Common Information Resource on Species Biodiversity. Mathematical biology and bioinformatics. **9**(2): 396–405. doi: 10.17537/2014.9.396. (In Russian)

Knyazeva, S.G., Milutin, L.I., Muratova, E.N., Larionova, A.Y., and Oreshkova, N.V. 2007. Database of intraspecific variation of Siberian conifers. Boreal Coniferous **58**(2-3): 201–206. (In Russian)

Koropachinsky, I.Y., Shokin, Y.I., Noisy, V.K., Ermakov, N.B., Kolchanov, N.A., and Fedotov, A.M. 1999. Electronic atlas "Biodiversity of flora and fauna of Siberia". *In* Digital Libraries: Advanced Methods and Technologies, Digital Collections: proceedings of a conference. St. Petersburg, pp. 187–192. (In Russian)

Kryshen', A.M., Polevoi, A.V., Gnatyuk, E.P., Kravchenko, A.V., and Kuznetsov, O.L. 2009. Database of habitats (biotopes) of Republic of Karelia'. Transactions of Karelian Research Centre of the Russian Academy of Sciences. **4**: 3–10. (In Russian)

Lobanov, A.L., and Smirnov, I.S. 1997. The principles of arrangement and using of classifiers of animals in the standard ZOOCOD. *In* Databases and computer graphics in zoological investigations: proceedings of the conference. Zoological Institute, St. Petersburg. pp. 66–75. (In Russian)

Lobaria pulmonaria in Russia. Information system. 2016. Available from: http://lobaria.ru/ [accessed September 12, 2016]. (In Russian)

Melechin, A.V., Davydov, D.A., Shalygin, S.S., and Borovichev, E.A. 2013. Open information system on biodiversity cyanoprokaryotes and lichens CRIS (Cryptogamic Russian Information System). Bulletin of Moscow Society of Naturalists. Biological series. **118**(6): 51–56. (In Russian)

Morozova, O.V., and Borisov, M.M. 2010. Web-Oriented Geoinformation System on Alien Plant Species of European Russia. Russian Journal of Biological Invasions. **2**: 47–55. (In Russian)

Nestboxing. 2016. Available from: http://nestboxing.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

ORNIS. 2016. Available from: http://www.ornisnet.org/ [accessed September 12, 2016].

Panyukova, E.V., Pestov S.V., and Madi, E.G. 2014. Information system on insects of the ‘GNUS’ complex from the north-eastern European Russia. Parasitology. **48**(1): 71–75. (In Russian)

PostgreSQL: The world’s most advanced open source database. 2016. Available from: https://www.postgresql.org/ [accessed September 12, 2016].

Protected areas of Russia. 2016. Available from: http://oopt.aari.ru/ [accessed September 12, 2016]. (In Russian)

Rare plants of Siberia. 2016. Available from: http://sibrareplants.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

Russian GBIF IPT. 2016. Available from: http://www.gbif.org/installation/7743c35d-adbd-4ca6-a185-193a7a8d2684 [accessed September 12, 2016].

Russian website of Global Biodiversity Information Facility. 2016. Available from: http://gbif.ru/ [accessed September 12, 2016]. (In Russian)

Shashkov, M.P., and Ivanova, N.V. 2012. Web-based information system on distribution of rare lichen *Lobaria pulmonaria* (L.) Hoffm. in Russia. Mathematical biology and bioinformatics. **7**(1): 334–344. doi: 10.17537/2012.7.334. (In Russian)

Small Wild Cats of Eurasia. 2016. Available from: http://wildcats.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

Smirnov, I., Lobanov, A., Alimov, A., Dianov, M., Golikov, A., and Stepanjants, S. 1997. Integrated information system ZOOINT: the present state and prospects *In:* ADBIS'97. Proceedings of the First East-European Symposium on Advances in Databases and Information Systems, 2. Nevsky Dialect, St. Petersburg. pp. 112–114.

The black Data Book of Russian Flora. 2016. Available from: http://www.bookblack.ru/ [accessed September 12, 2016]. (In Russian)

The Catalogue of Life. 2016. Available from: http://www.catalogueoflife.org/ [accessed September 12, 2016]. (In Russian)

The information-analytical system of Depositary of biomaterial of the resource center of Moscow State University. Plants. 2016. Available from: https://plant.mitotech.ru/ [accessed September 12, 2016]. (In Russian)

The information system 'Biodiversity in Russia'. 2016. Available from: https://www.zin.ru/BioDiv/index_e.html [accessed September 12, 2016]. (In Russian)

The National Strategy for biodiversity conservation in Russia. 2002. Moscow.

The red data book of Altai Krai. 2016. Available from: http://altayredbook.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

The red data book of Samara region. 2016. Available from: http://samarafauna.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

UgraBio. Information system of biodiversity of Ugra. 2016. Available from: http://ugrabio.ru/ [accessed September 12, 2016]. (In Russian)

Vertebrate Animals of Russia. 2016. Available from: http://www.sevin.ru/vertebrates/ [accessed September 12, 2016]. (In Russian)

Web-GIS Birdwatching. 2016. Available from: http://rrrcn.ru/en/birdwatching/web-gis [accessed September 12, 2016].

Web site of Algal Collection of Soil Science Institute (ACSSI). 2016. Available from: http://acssi.org/index.php [accessed September 12, 2016]. (In Russian)

Wetland Mammals of Eurasia. 2016. Available from: http://wetlandmammals.wildlifemonitoring.ru/?lang=en [accessed September 12, 2016].

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Dorring, M., Gilovanni, R., Robertson, T., and Vieglais, D. 2012. Darwin core: an evolving community-developed biodiversity data standard. PLoS ONE. **7**(1):1–8. doi: 10.1371/journal.pone.0029715.

Zaugol'nova, L.B., and Khanina, L.G. 1996. Experience in the development and use of databases in the forest phytocenology. Lesovedenie. **1**: 76–83. (In Russian)

Zeltyn, A.C., and Insarov, G.E. 1993. Database of background monitoring of epiphytic lichens. *In* The problems of ecosystem environmental monitoring and ecosystem modeling: proceedings of the conference, 15. Gidrometeoizdat Press, St. Petersburg. pp. 247–263. (In Russian)

Zoological Institute of Russian Academy of Science. 2016. Available from: http://www.zin.ru/ [accessed September 12, 2016]. (In Russian)

Zoological museum of Moscow University. 2016. Available from: http://zmmu.msu.ru/ [accessed September 12, 2016]. (In Russian)

Figure Captions

Fig. 1. General scheme of the database *Lobaria pulmonaria* occurrences in Russia.

Fig. 2: The relation of the logical parts of the data set with exact (A) and inexact (B) georeferencing

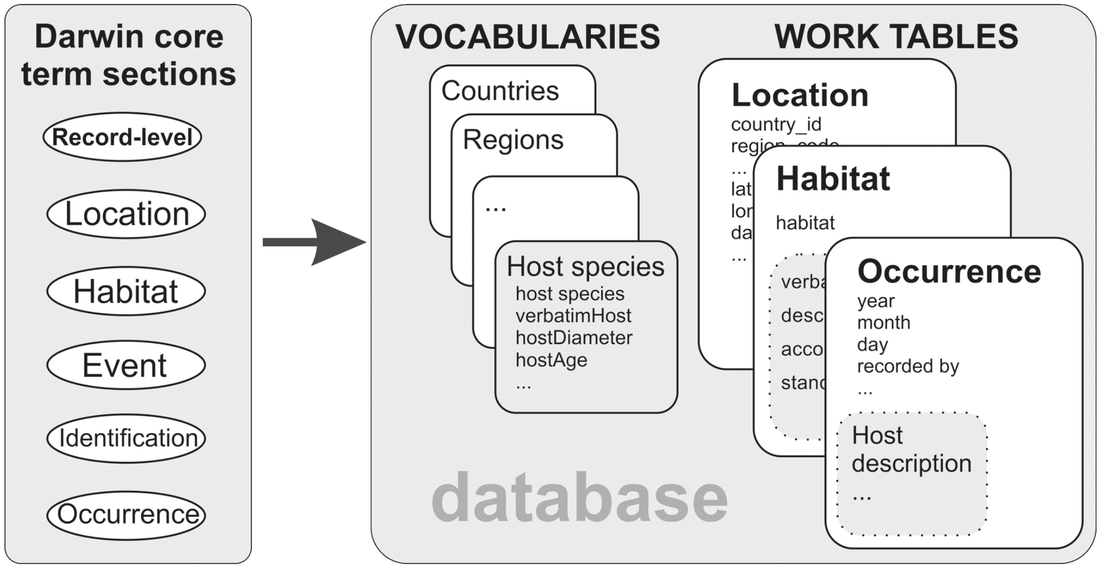

Fig. 1. General scheme of the database *Lobaria pulmonaria* occurrences in Russia.

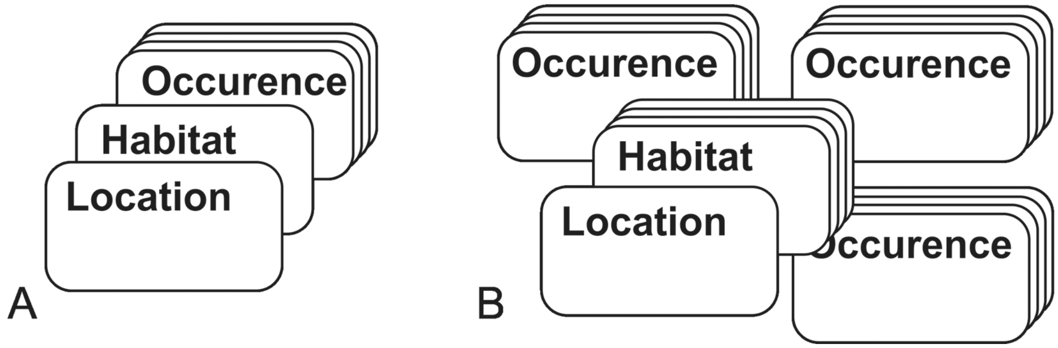

Fig. 2: The relation ratio of the logical parts of the data set with exact (A) and inexact (B) georeferencing

Table 1a. Occurrence databases characteristics.

Name	Data standard	Number of records	Availability of the primary data	Dynamic map (updatable)	Data of creation / Updates	Developer
Alien species in the territory of Russia (2016)	Local data standard	1448 plant species, 54 mammals species	Only species checklists	No	2003 / No	Institute of Ecology and Evolution RAS
Information Retrieval System for Fauna and Flora in Protected Natural Areas of the Russian Federation (2016)	ZOOCOD standard	333 fish species, 26 amphibians, 56 reptiles, 702 birds, 254 mammal, 883 moss, 2015 lichen, 328 Hepaticae and Anthocerotae, 7920 vascular plant	Only species checklists	No	2003 / No	Institute of Ecology and Evolution RAS
Protected areas of Russia (2016)	Local data standard	Unknown	No	Yes	2012 / Yes	
Vertebrate Animals of Russia (2016)	ZOOCOD standard	295 species and subspecies of freshwater fish, 29 amphibians, 84 reptiles, 730 birds and 310 mammal species	No	No	2003 / No	Institute of Ecology and Evolution RAS
Web-Oriented Geoinformation System on Alien Plant Species of European Russia (Morozova and Borisov 2010)	Local data standard	About 25000 records for 2000 species	Yes	Yes	2010 / Yes	Institute of Geography RAS
The Black Data Book of Russian Flora (2016)	Local data standard	52 plant species	No	No	Unknown / No	Unknown
Occurrences of the invasive plant species *Heracleum sosnowskyi* Manden. in the Komi Republic territory	Local data standard	10894 occurrences	Through GBIF.org (table 2, doi 10.15468/zo2svq)	Yes	2014 / Yes	Institute of Biology of Komi Scientific Centre of the Ural Branch of the

(European North-East Russia)						Russian Academy of Sciences
Information system on insects of the «gnus» complex From the north-eastern European Russia (Panyukova et al. 2014)	Local data standard	173 species	No	No	2014 / No	Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences
Baikal Siberia Flora (Abdrahimov et al. 2011)	Local data standard	3026 records	No	No	2010 / No	Irkutsk State University
Biodiversity of flora and fauna of Siberia (Koropachinsky et al. 1999)	Local data standard	Unknown	No	No	1999 / No	Institute of Computational Technologies of SB RAS (ICT SB RAS)
Biodiversity of Altai-Sayan Ecoregion (2016)	Local data standard	1879 animal species, 3162 plant species	No	No	2007 / No	South Siberian Botanical Garden of Altai State University
Cryptogamic Russian Information System, CRIS (Melechin et al. 2013)	Local data standard	157 species of Fungi, 1902 species (23233 samples) of hepatics, 1568 species (10270 samples) of lichens and 1438 species (19004 samples) of mosses	Yes	Yes	2012 / Yes	Polar-Alpine Botanical Garden-Institute of N.A. Avrorin KSC RAS
Web-GIS Birdwatching (2016)	Local data standard*	30716 observations and 23262 occurrences	Yes	Yes	2012 /Yes	Siberian Environmental Center
Nestboxing (2016)	Local data standard*	1400 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center
Small Wild Cats of Eurasia (2016)	Local data standard*	390 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center

Rare plants of Siberia (2016)	Local data standard*	41 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center
The red data book of Samara region (2016)	Local data standard*	50 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center
The red data book of Altai Krai (2016)	Local data standard*	2060 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center
Wetland Mammals of Eurasia (2016)	Local data standard*	1250 observations	Yes	Yes	Unknown /Yes	Siberian Environmental Center
UgraBio. Information system of biodiversity of Ugra (2016)	Local data standard	4241 records	Yes	Yes	2014 / Yes	NextGIS Ltd.
Lobaria pulmonaria in Russia. Information system.(Shashkov and Ivanova 2012)	Darwin Core	1200 records	Through GBIF.org (table 2, doi 10.15468/uennht)	Yes	2012 / Yes	Institute of Mathematical Problems of Biology RAS

* *unified data standard*

Table 1b. Taxonomic database characteristics.

Name	Data standard	Number of records	Availability of the primary data	Data of creation / Updates	Developer
Flora of vascular plants in the Central European Russia	Local data standard	2365 species	Species checklist through GBIF.org (table 2, doi 10.15468/96gqtn)	2004 / Only interface modification	Institute of Mathematical Problems of Biology RAS
Information System ZooInt (Smirnov et al. 1997)	ZOOCOD standard	>30 data bases	Only viewable. Species checklist as single file is not available	1992 / No	Zoological Institute RAS
Biodiversity in Russia, BIODIV (2016)	ZOOCOD standard	Unknown. The object of the research are prokaryotes, protists, fungi, plants, and animals inhabiting Russia and neighbouring territories	Only viewable. Species checklist as single file is not available	2001 / No	Zoological Institute RAS
Information System ZInsecta (2016)	ZOOCOD standard	Unknown	Only viewable. Species checklist as single file is not available	2002 / No	Zoological Institute RAS
Biodiversity of Animals. Russian information system, ZooDiv (2016)	ZOOCOD standard	More than 45,000 taxa, >90,000 species	Only viewable. Species checklist as single file is not available	2007 / No	Zoological Institute RAS

Table 1c. Digitized collections characteristics.

Name	Data standard	Number of digitized samples	Type of data formalization	Data of creation / Updates	Developer
The information-analytical system of Depositary of biomaterial of the resource center of Moscow State University. Plants. (2016)	Local data standard	501892 samples, 9516 species	Digitized samples (images 300 dpi) and formalized attribute data	2015 / Yes	Moscow State University
Catalog of specimens found in collections Komarov Botanical Institute RAS (2016)	Local data standard	1320 specimens	Digitized samples (images) and formalized attribute data	Unknown / No	Komarov Botanical Institute of the RAS
Digitized Research Collections of the Zoological Institute RAS (2016)	Local Darwin Core-similar data standard	440 specimens	Digitized samples (images), original label (image) and formalized attribute data	Yes	Zoological Institute RAS
Web site of Algal Collection of Soil Science Institute, ACSSI (2016)	Local data standard	294 strains	Strain photos, taxonomy, attributive data	2016 / Yes	Institute of Physical-Chemical and Biological Problems in Soil Science
Collection of microalgae strains in the Institute of Biology of Komi Scientific Centre, SYKOA (2016)	Local data standard	About 250 strains	Strain photos, attributive data	2010 / Unknown	Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences
Arctoa. Project ‘Flora of mosses of Russia’ (2016)	Local data standard	Unknown	Formalized attribute data about herbarium spesimens	Unknown / Unknown	Moscow State University
Genetic and biological (zoological and botanical) collections of the Russian Federation. Herbarium collections. (2016)	Local data standard	146 herbarium collections	Information about herbarium collection.	2003 / No	Institute of Ecology and Evolution RAS

Table 2. Data from Russia available from GBIF.

Dataset title	Count	Publication date	Publisher	Served by
Amphibian specimens doi:10.15468/c9g3nw	0	was deleted on 18.08.2016	Zoological Institute, Russian Academy of Sciences, St. Petersburg	Zoological Institute RAS
A grid-based database on vascular plant distribution in the Meshchera National Park, Vladimir Oblast, Russia doi:10.15468/ahunho	22 625	11.09.2014	Moscow University Herbarium (MW)	MSU http
Avena wild species collection. (N.I.Vavilov Institute of Plant Genetic Resources (VIR)) doi:10.15468/xtcciv	1 639	19.12.2015	N. I. Vavilov Institute of Plant Genetic Resources (VIR)	IPT VIR N.I. Vavilov
Rare vascular plant species in North-West of Kostroma region, Russia doi:10.15468/tnlga7	106	01.09.2015	Institute of Mathematical Problems of Biology, Russian Academy of Sciences	Russian GBIF IPT
Kpabg_lichens doi:10.15468/nctfm2	10 730	02.03.2016	Polar-Alpine Botanical Garden-Institute of N.A. Avrorin KSC RAS	Russian GBIF IPT
Kpabg_cyano doi:10.15468/80tu83	3 201	04.03.2016	Polar-Alpine Botanical Garden-Institute of N.A. Avrorin KSC RAS	Russian GBIF IPT
Kpabg_hepatics doi:10.15468/yxt7co	23 421	24.03.2016	Polar-Alpine Botanical Garden-Institute of N.A. Avrorin KSC RAS	Russian GBIF IPT
Avena wild species VIR Herbarium. N.I.Vavilov Institute of Plant Genetic Resources (VIR) doi:10.15468/cjzloe	311	23.12.2015	N. I. Vavilov Institute of Plant Genetic Resources (VIR)	IPT VIR N.I. Vavilov
Database of finds of rare lichen species *Lobaria pulmonaria* in Russia doi:10.15468/uennht	1 186	16.04.2016	Institute of Mathematical Problems of Biology, Russian Academy of Sciences	Russian GBIF IPT
Ophiuroidea collections of the Zoological Institute Russian Academy of Sciences doi:10.15468/ej3i4f	8 715	11.07.2016	Zoological Institute, Russian Academy of Sciences, St. Petersburg	ZIN RAS and BIN RAS IPT

Amphibians of the Former USSR doi:10.15468/wxz3yj	52 474	03.08.2016	A.N. Severtsov Institute of Ecology and Evolution, RUSSIAN ACADEMY OF SCIENCES	Russian GBIF IPT
Occurrences of the invasive plant species *Heracleum sosnowskyi* Manden. in the Komi Republic territory (European North-East Russia) doi:10.15468/zo2svq	10894	19.10.2016	Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences	Institute of Biology (Syktyvkar, Russia) Integrated Publishing Toolkit (IPT) Installation
List of Spiders of Prioksko-Terrasnyi Biosphere Reserve doi:10.15468/3cbyt7	787	15.11.2016	Prioksko-Terrasnyi Biosphere Reserve	Russian GBIF IPT
Flora of vascular plants in the Central European Russia doi:10.15468/96gqtn	2365	16.11.2016	Institute of Mathematical Problems of Biology, Russian Academy of Sciences	Russian GBIF IPT
Cyanobacteria of Algal Collection of Soil Science Institute (ACSSI) doi:10.15468/nt9emp	37	09.12.2016	Institute of physicochemical and biological problems in soil science of the Russian Academy of Sciences	Russian GBIF IPT
Sarcinoid green algae (Chlorophyta) of Algal Collection of Soil Science Institute (ACSSI) doi:10.15468/cm3n7s	17	09.12.2016	Institute of physicochemical and biological problems in soil science of the Russian Academy of Sciences	Russian GBIF IPT