Skvortsovia: 4(3): 100–111 (2018)

Skvortsovia
http://skvortsovia.uran.ru/
ISSN 2309-6497 (Print)
ISSN 2309-6500 (Online)

Article

Digitization of vascular plant herbarium collections at the Central Siberian Botanical Garden, Novosibirsk, Russia

Nataliya K. Kovtonyuk*, Irina V. Han and Evgeniya A. Gatilova

Russian Academy of Sciences, Siberian Branch: Central Siberian Botanical garden, Zolotodolinskaya St. 101, Novosibirsk 630090, Russian Federation. *Email: nkovtonyuk@csbg.nsc.ru

Received: 25 September 2018 | Accepted by A.P.Dyachenko: 25 November 2018 | Published on line: 03 December 2018

Abstract

The herbaria at the Central Siberian Botanical Garden Siberian Branch of the Russian Academy of Sciences (CSBG) were started in 1946. There are two herbarium collections with their own codes and registrations in the Index Herbariorum: M.G.Popov Herbarium (NSK) and I.M.Krasnoborov Herbarium (NS). Approximately 800000 herbarium specimens of vascular plants, mosses, lichens and fungi collected in Siberia, the Russian Far East, Europe, Asia and North America are stored in NSK and NS. Digitization of the herbarium specimens of vascular plants at the resolution of 600 dpi began in 2014 using a special scanner, *Herbscan*, starting with type specimens. In 2017 a new research group Unique Scientific Unit – Herbarium (USU-Herbarium) was organized in the CSBG for the digitization and management of herbarium collections. Herbarium specimens are being digitized using three customized scanners: *Herbscan* and two *ObjectScan* 1600, according to the international standards, at 600 dpi, with a barcode, 24-color scale and spatial scale bar. Images and metadata are stored in the CSBG Database generated by *ScanWizard Botany* and *MiVapp Botany* software (*Microtek*, Taiwan).

Keywords: biodiversity, Central Siberian Botanical Garden, digitization, digital herbarium collections, vascular plants

Introduction

Botanical Science must adapt to the digital revolution and concede that information about plants should be available online. This process of virtualization of botanical sources, such as botanical publications and herbarium collections, is developing very quickly worldwide. Many herbaria have websites with their virtual collections available online, yet in Russia only a few institutions have such a facility, including the Central Siberian Botanical Garden (CSBG).

Herbarium collections are the most important source of scientific information on distribution of plants in the past and present, also allowing for simulations of future distribution dynamics. Only the herbarium sample reliably confirms the presence of the plant at a specific point of space and time. Herbarium collections and the data they hold are valuable for more traditional studies of taxonomy and systematics, but also for research in the

fields of ecology, bioengineering, conservation, food security, and ethnobotany (Kovtonyuk, 2015, 2017; Seregin, 2016, 2018; Kislov *et al*., 2017; Kovtonyuk *et al*., 2018; James *et al*., 2018; NicLughadha *et al*., 2018). The value and universality of herbarium specimens are recognized in most countries, where national and large regional herbaria are actively developing and improving. The herbaria at the CSBG were organized in 1946. There are two major herbarium collections with their own codes and registrations in the Index Herbariorum founded in CSBG: M.G.Popov's Herbarium (NSK) and I.M.Krasnoborov's Herbarium (NS) (Doronkin *et al*., 2015). Nearly 800000 herbarium specimens of vascular plants, mosses, lichens and fungi collected in Siberia, the Russian Far East, Europe, Asia and North America are stored in NSK and NS (Fig.1).

Figure 1. (A) M.G.Popov Herbarium (NSK) and (B) I.M.Krasnoborov Herbarium (NS) in CSBG, Novosibirsk, Russia. Photo by N.Kovtonyuk

Materials and methods

Digitization of the vascular plants in the NSK and NS collections initially was undertaken with a customized *HerbScan* unit, which consisted of a flatbed scanner (*Epson Expression* 10000XL) modified for inverted use (Fig. 2). Currently digitization of samples is carried out using *Herbscan* and two *ObjectScan* 1600 scanners. Each workstation consists of an *ObjectScan* 1600 scanner, *ScanWizard_Botany* software (reg. number I41-018966) and *MiVapp_Botany* archive management system (reg. number I41-018969) software (*Microtek*, Taiwan). This integrated workstation is characterized by on-top scan design for full-frame focus, a maximum of 1600 dpi (equal to 1 Gigabyte pixels), color CCD, Optical Character Recognition (OCR) for specimen label and ID barcode, and image archive and privileged-account cloud management system. Specimens were scanned using international standards: with the resolution of 600 dpi, a barcode for each specimen, and a 24-color scale and scale

bar. http://www.snsb.info/SNSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf)

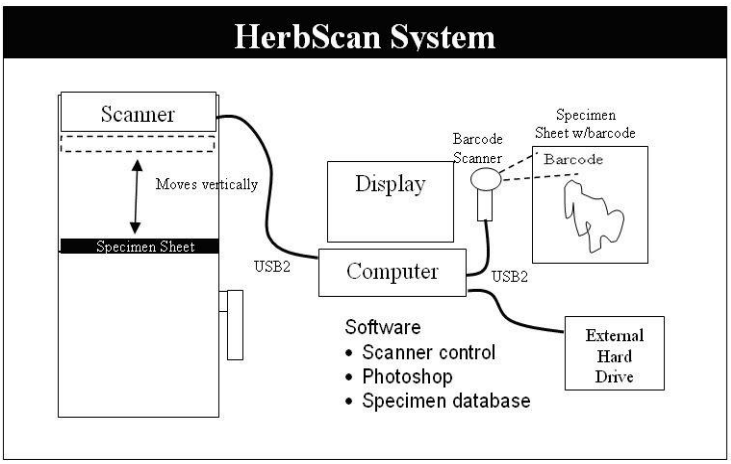

Figure 2. Scheme of the *HerbScan* system operation for scanning herbarium specimens

Figure 2 depicts the stages of the scanning process. Upon scanning an herbarium specimen, the scan is stored in the connected PC. A barcode is attached to the specimen and stored electronically, using a barcode scanner. In addition to storage in the PC, the scanned image is copied to an external hard drive.

Following recommendations of the Global Plant Initiative (GPI), a training course *Specimen Digitization and Quality Assurance* was organized at Royal Botanic Gardens, Kew (London, UK) on 13–15 May, 2014 by Dr. Jonathan Krieger and his team. Dr. Nataliya Kovtonyuk attended this course and received a Certificate of Attendance (Fig. 3–4). This course was a prerequisite for starting digitization of the herbarium collections in the CSBG. Special emphasis was laid on providing online high-resolution (600 dpi) images and metadata for all type specimens (Kovtonyuk, 2015, 2017a) and creation of a "typo-theca," a collection of type specimen images and copies of pertaining protologues (Fig. 5–6), as many protologues published in Russian have not been available in digital format. Images and metadata for 895 type specimens preserved at the NSK and NS are currently available on the *Virtual Herbaria* web site at the University of Vienna, Austria (http://herbarium.univie.ac.at/database) and some of them in JSTOR (https://plants.jstor.org).

Figure 3. Nataliya Kovtonyuk selects herbarium specimens for scanning. Photo by Irina Belyaeva

The value of digital images of type specimens has recently grown, since it grants ready access to authentic material necessary for research and typification, thus providing for correct application of scientific names, as stated in Art. 7.1 of the International Code of

Figure 4. Certificate of Attendance. Photo by Nataliya Kovtonyuk

Nomenclature for Algae, Fungi, and Plants (ICN) (Turland *et al*., 2018) and demonstrated in recent publications (Kovtonyuk and Belyaeva, 2015; Kuzovkina *et al*., 2016a; Kuzovkina *et al*., 2016b; Kovtonyuk, 2017b; Belyaeva *et al*., 2018). Type specimens are kept in the CSBG

Figure 5. L.Lukmanova digitizes type specimens on *Herbscan*. Photo by Nataliya Kovtonvuk

Figure 6. Typo-theca in NSK. Photo by Nataliya Kovtonyuk

in designated cabinets for better preservation and control of handling.

A new research group *USU-Herbarium* was organized in the CSBG in 2017 (registered at ckp-rf.ru as 'USU 440537') with the purpose of digitization and management of herbarium specimens in the NS and NSK collections.

Figure 7. Workflow at the workstation *ObjectScan* 1600 (http://www.microtek.com/products.php?KindID=12&ID=373)

The digitization process is shown in Figs.7–9 and includes the following 6 steps:

1. Mounting of dry plant material onto an herbarium sheet;

2. Check of the identification and nomenclature by a specialist;

3. Barcoding the specimen: printing a barcode on the thermal printer and affixing it to the herbarium sheet;

4. Placing the herbarium sheet, 24-color scale and scale bar (Fig. 8a) on the scanner platform and image capturing;

5. Metadata generation, label OCR by *ScanWizard Botany* and verification of the label text by experts;

6. Archive management by *MiVapp-Botany*.

Figure 8. I.Deyun scans herbarium specimens of NSK collections using *ObjectScan* 1600. 8a. The scale bar. Photos by Nataliya Kovtonyuk

Figure 9. Evgeniya Gatilova scans herbarium specimens of the NS collections using *ObjectScan*1600. Photo by Nataliya Kovtonyuk

Specimen label information is recognized and automatically saved under herbarium code and specimen serial number in XML format through *ScanWizard-Botany*. *MiVapp-Botany* is a web-server system and a specimen image authentication database aiming for an efficient and integrated multi-functional platform. After hierarchical login-based image quality and metadata profile validation by experts (Fig. 10), *MiVapp-Botany* can quickly update the system and make verified specimens accessible for users.

Overview

Data on the employment of some type specimens stored in NSK in previous taxonomic research were published by Kovtonyuk and Belyaeva (2015) and Han (2016).

Currently about 15000 herbarium specimens from NS and NSK have been digitized at 600 dpi by special scanners *HerbScan* and *ObjectScan* and are available for researchers worldwide. Images and metadata of herbarium specimens are stored in the CSBG Database (http://84.237.85.99:8081), generated by *ScanWizard Botany* and *MiVapp Botany* (Kovtonyuk *et al*., 2018).

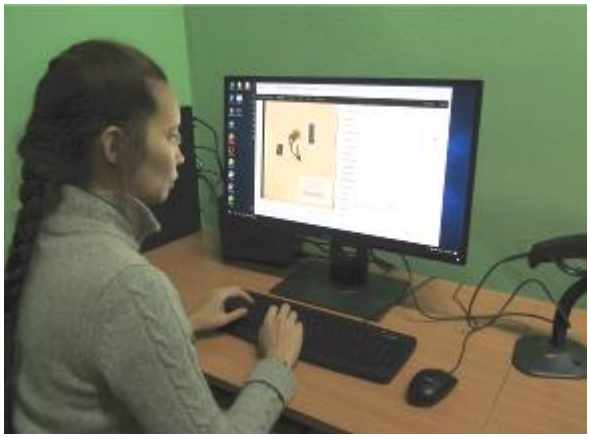

Figure 10. Verification of herbarium specimens by expert Irina Han. Photo by Nataliya Kovtonyuk

The database is structured in such a way that, while using a specimen record, the user can access a high-resolution image along with the following information: specimen ID (= barcode), family name, scientific name (genus, species, author of taxon), collector's name and collection date and the country and administrative region. Alternatively, a query may be submitted using habitat characteristics as key words, for example, "pine forest" ("сосновый лес" in Russian), "meadows" ("заливные луга"), etc. Images of specimens, label data and maps are available for downloads (the latter as attached images) (Figs. 11–13).

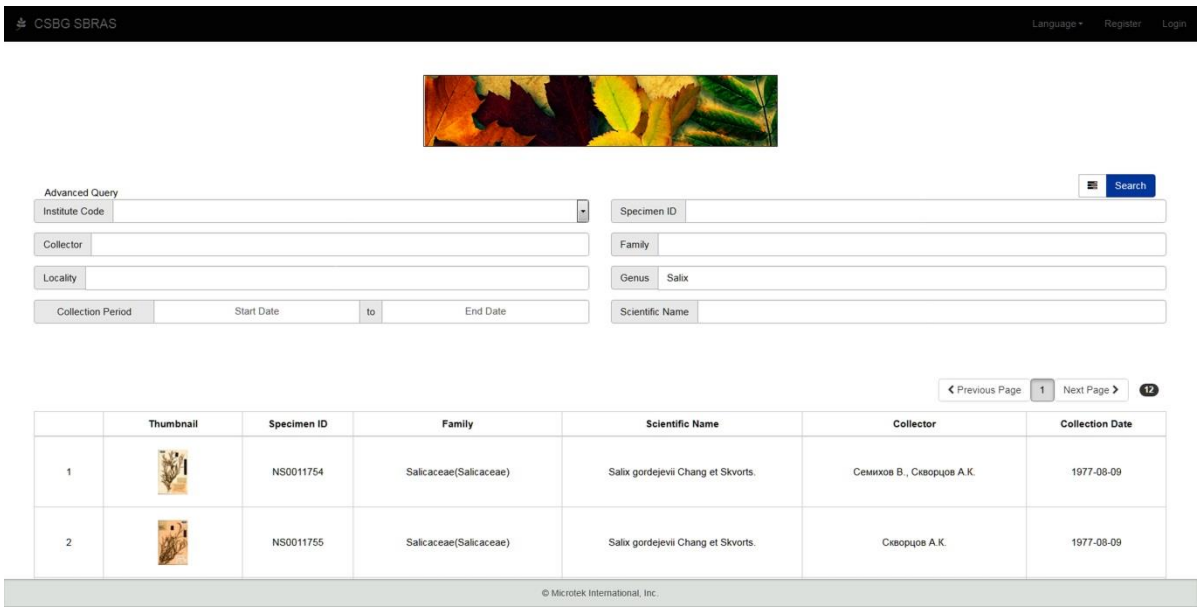

Figure 11. Genus *Salix* in CSBG Digital Herbarium database

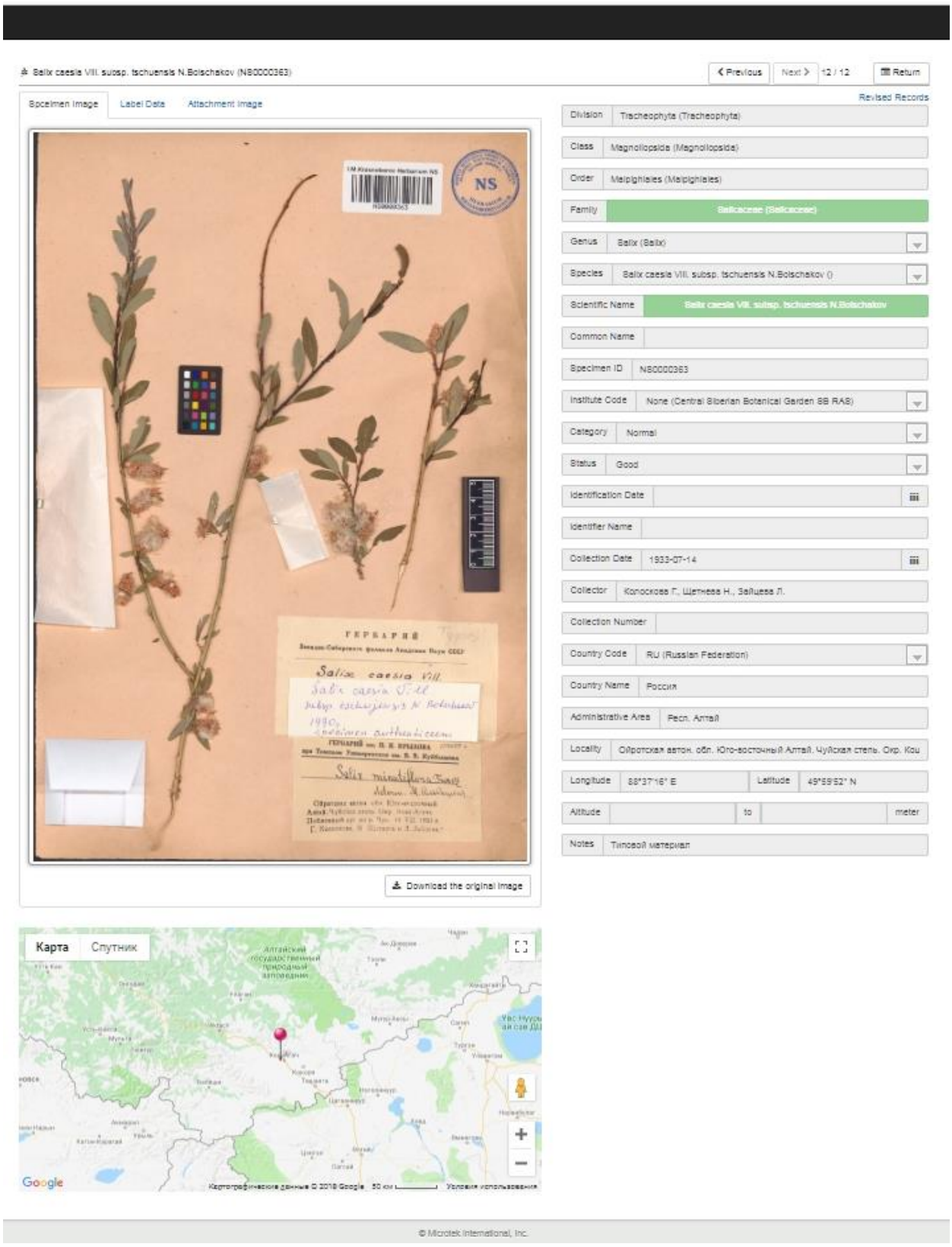

Figure 12. Example of a record in CSBG Digital Herbarium

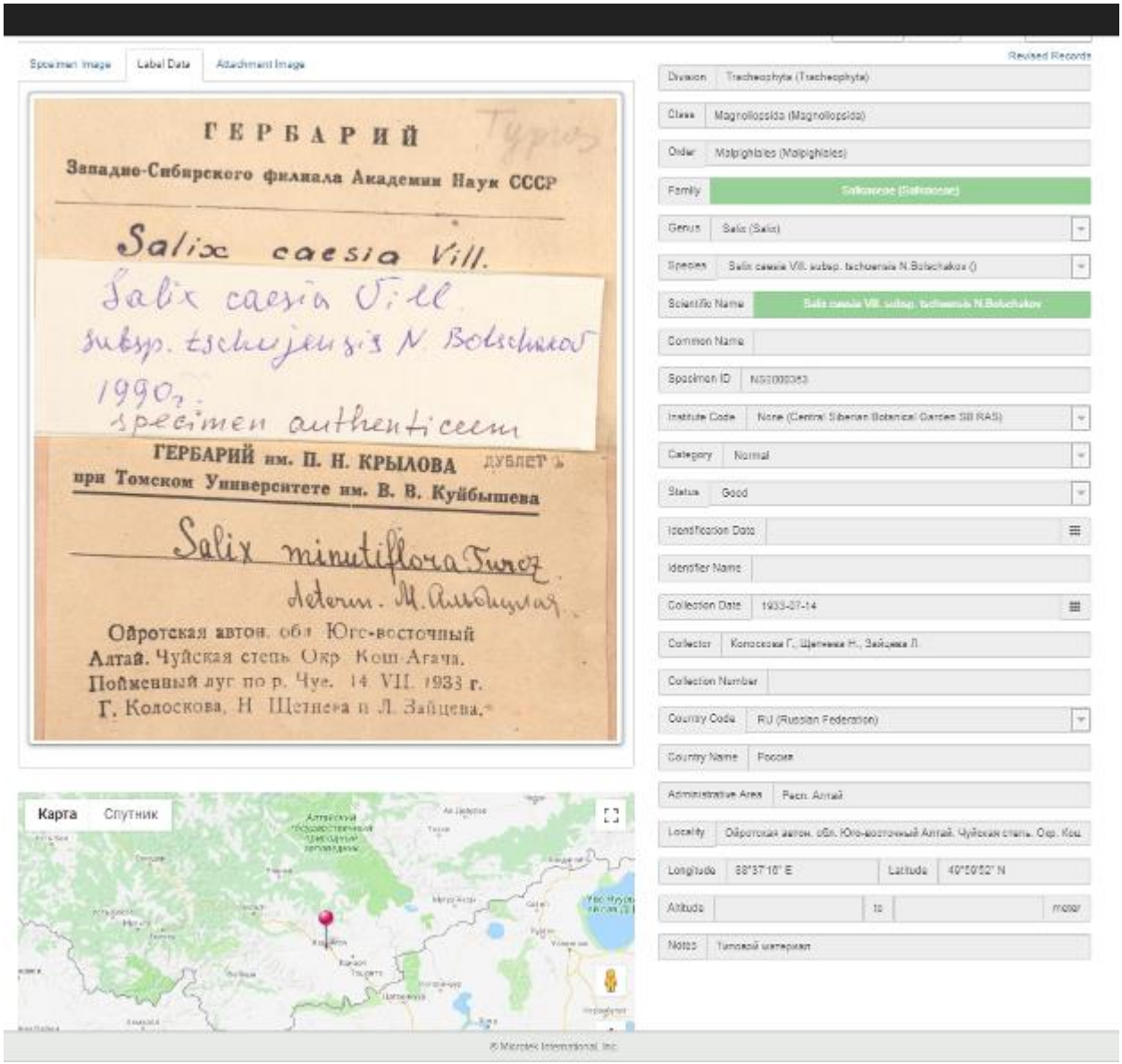

Fig.13. Herbarium label, map and metadata for *Salix caesia* Vill. ssp. *tschujensis* N.Bolschakov (NS0000363)

In our internal database each image is provided with the following information: barcode, type status, genus name, species name, author's name, infraspecific name, family name, collector's name, field collection number, date (yyyy-mm-dd), country, administrative region, latitude (degrees and minutes), longitude (degrees and minutes), label text, name of identifier, annotation and accepted name in the *Catalogue of Life* (CoL) with a link to this database. Data from herbarium labels are entered in 26 fields in the Calc tables of the *LibreOffice* software package. Parsing information from herbarium labels allows the placement of queries, exporting electronic tables to other media, such as domestic bio-resource networks or international databases, GBIF resources (gbif.org) and processing statistical data (Kovtonyuk *et al*., 2018b). The CSBG Herbarium taxonomic database is

compatible with the international resource, *Catalogue of Life* (http://www.catalogueoflife.org), where an updated list of taxa is published every month (Roskov *et al.*, 2017). Names used in the herbarium labels have been verified and updated in accordance with the *International Plant Name Index* and POWO (http://powo.science.kew.org).

By the time of the paper publication, scanning was fully completed for Primulaceae (5028 sheets) and partially for a number of other families: Cystopteridaceae (891), Athyriaceae (436), Amaryllidaceae (2551), Orchidaceae (750), Poaceae (632), Fabaceae (418), Boraginaceae (664) and Asteraceae (550). Ninety-two recent herbarium collections representing the genus *Salix*, mostly from the Northern Urals and Taimyr Peninsula have been submitted to the CSBG Herbarium for digitization by the collectors I.N.Pospelov and E.B.Pospelova.

Our plan at the CSBG is to continue with digitization of herbarium specimens of various families as requested by multiple experts. Developing virtual herbaria will not only facilitate the study and research on herbarium specimens, analysis of variability and distribution of taxa, of their composition in certain geographical regions, but also speed up regional projects, such as the long-awaited *Flora of Russia*, along with worldwide biodiversity projects.

Acknowledgements

We would like to thank Sergey Lukyanchikov (CSBG SB RAS) and *Microtek*'s engineers for consultations, Lyalya Lukmanova, Svetlana Krasnikova and Irina Deyun for the technical support of our research; Dr. Yury Rostkov and Dr. Irina Belyaeva for their help and consultations. In 2014–2016 digitization was supported by the Andrew Mellon Foundation. This work was partially financially supported by the Russian Foundation for Basic Research (grant No 15-29-02429) and the base project VI.52.1.5. (state reg. AAAA-A17-117012610055-3). Anonymous reviewers and editors are gratefully acknowledged.

References

Belyaeva, I.V., Epantchintseva, O.V., Govaerts, R.H.A., McGinn, K. , Hunnex, J. and Kuzovkina, Yu.A. 2018. The application of scientific names to plants in cultivation: *Salix vitellina* L. and related taxa (Salicaceae). *Skvortsovia*: 4(2): 42–70.

Doronkin, V.M., Shaulo, D.N., Banaev, E.V. and Naumenko, Yu.V. 2015. The herbarium collections in the Central Siberian Botanical Garden, SB RAS – status and prospects. In:

Novikova, L.A. (ed.) Botanical collections – the National heritage of Russia, Penza, PSU: 228–232.

International Plant Names Index (IPNI), http://www.ipni.org/ipni/ (last accessed on 2018/09/24).

James, S. A., Soltis, P.S., Belbin, L., Chapman, A.D., Nelson, G., Paul, D.L. and Collins, M. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences 6(2), e1024. doi:10.1002/aps3.1024.

Han, I.V. 2016. Taxa of Asteraceae, Euphorbiaceae and Scrophulariaceae in collections of types of M.G. Popov Herbarium (NSK). Rast. Mir Asiat. Rossii 3(23): 41–54.

Kislov, D.E., Bakalin, V.A., Pimenova, E.A., Verkholat, V.P. and Krestov, P.V. 2017. An electronic management system for a digital herbarium: development and future prospects. Botanica Pacifica 2: 59–68.

Kovtonyuk, N.K. 2015. Virtual collection of type specimens in M.G. Popov Herbarium (NSK). Rast. Mir Asiat. Rossii 3(19): 88–93.

Kovtonyuk, N.K. 2017. Virtual herbarium collections as a resource for taxonomy and biodiversity study. Rast. Mir Asiat. Rossii, 1(25): 98–104.

Kovtonyuk, N. and Belyaeva, I. 2015. Nomenclatural and taxonomic notes on the names published by M.G.Popov in *Salix* L. and *Populus* L. (Salicaceae). Skvortsovia 2(2): 126–140.

Kovtonyuk, N., Han, I. and Gatilova, E. 2018a. Virtual herbarium collections of the Central Siberian Botanical Garden as a resource for biodiversity study // Abstracts of the conference *Information Technologies in the Research on Biodiversity* (BIT - 2018). Irkutsk, ISDCT SB RAS: 114–115. ISBN 978-5-6041814-0-9 (http://bit.icc.ru/theses.pdf).

Kovtonyuk, N., Han, I. and Gatilova, E. 2018b. Family Primulaceae in Virtual Herbaria of CSBG SB RAS (NSK). Version 1.3. Central Siberian Botanical Garden SB RAS. Occurrence dataset https://doi.org/10.15468/avjnkt accessed via GBIF.org on 2018-11-27.

Kuzovkina, Y., Epantchintseva, O. and Belyaeva, I. 2016a. The application of scientific names to plants in cultivation: *Salix* × *cottetii* Lagger ex A.Kern. (Salicaceae). Skvortsovia 2(3): 32–43.

Kuzovkina, Yu.A., Dodge, M. and Belyaeva, I.V. 2016b. Clarifying affiliations of *Salix graciclistyla* Miq. cultivars and hybrids. HortScience 51(4): 334–341.

Microtek scan the world (http://www.microtek.com/products.php?KindID=12&ID=373 last accessed 2018/09/24).

NicLughadha, E., Walker, B.E., Canteiro, C., Chadburn, H., Davis, A.P., Hargreaves, S., Lucas, E.J., Schuiteman, A., Williams, E., Bachman, S.P., Baines, D., Barker, A.,

Budden, A.P., Carretero, J., Clarkson, J.J., Roberts, A. and Rivers, M.C. 2018. The use and misuse of herbarium specimens in evaluating plant extinction risks. *Phil. Trans. R. Soc. B* 374(1763) DOI: 10.1098/rstb.2017.0402. Available online.

Plants of the World online (POWO, http://powo.science.kew.org last accessed 2018/09/24).

Roskov, Y., Abucay, L., Orrell, T., Nicolson, D., Bailly, N., Kirk, P.M., Bourgoin, T., DeWalt, R.E., Decock, W., De Wever, A., Nieukerken, E. van, Zarucchi, J. and Penev, L. (eds.) Species 2000 & ITIS *Catalogue of Life*. 2017, Annual Checklist. Digital resource at www.catalogueoflife.org/annual-checklist/2017. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405-884X. URL: http://www.catalogueoflife.org (last accessed 24/09/2018).

Seregin, A. 2016. Making the Russian flora visible: fast digitization of the Moscow University Herbarium (MW) in 2015. Taxon 65 (1): 203–209.

Seregin, A. 2018. The largest Digital Herbarium in Russia is now available online. Taxon 67 (2): 463–467.

Turland, N.J., Wiersema, J.H., Barrie, F.R., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T.W., McNeill, J., Monro, A.M., Prado, J., Price, M.J. and Smith, G.F. (eds.) 2018: International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress, Shenzhen, China, July 2017. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books. DOI https://doi.org/10.12705/Code.2018 .