

ГЕОДАННЫЕ ГЕРБАРИЯ МГУ: СТРУКТУРА ДАННЫХ И МЕТОДИКА ГЕОПРИВЯЗКИ

© 2023 г. А. П. Серёгин^{1*}, В. Н. Пашкина¹, И. Н. Поспелов²

¹Московский государственный университет имени М.В. Ломоносова
Ленинские горы, 1, Москва, 119991, Россия

²Институт проблем экологии и эволюции имени А.Н. Северцова РАН
Ленинский пр., 33, Москва, 119071, Россия

*e-mail: botanik.seregin@gmail.com

Поступила в редакцию 13.03.2023 г.

После доработки 08.04.2023 г.

Принята к публикации 18.04.2023 г.

Оцифровка Гербария Московского университета (MW) была начата весной 2015 г., с октября 2016 г. мы стали размещать изображения образцов в открытый доступ на портале Цифрового гербария МГУ (<https://plant.depo.msu.ru/>). Постепенно на нашем портале стали появляться и расширенные метаданные: полные расшифровки текста этикеток и геопривязки, т.е. значения координат места сбора каждого образца. На 14.01.2023 две трети образцов Гербария МГУ (693168 штук, или 66%) имеют электронные метки на карте, в то время как 359593 образца их пока не имеют. В статье дана характеристика массива геоданных Гербария Московского университета как в географическом аспекте (покрытие геоданных для разных регионов), так и в аспекте точности геопривязок. Приведены общие методические указания по геопривязке гербарных коллекций.

Ключевые слова: гербарий, пространственные данные, география растений

DOI: 10.31857/S0006813623050083, EDN: YZKILC

Оцифровка Гербария Московского университета (MW) была начата весной 2015 года с участием коммерческого партнера в рамках комплексной программы “Научные основы создания Национального банка-депозитария живых систем” по гранту РНФ. В течение 2015 года было отсканировано 502 тыс. образцов из отделов Восточной Европы, Сибири и Дальнего Востока – примерно половина фондовой коллекции Гербария МГУ.

С октября 2016 г. мы стали размещать изображения образцов в открытый доступ на портале Цифрового гербария МГУ (<https://plant.depo.msu.ru/>) (Seregin, 2018, 2023b). Изображения образцов выкладывались в открытый доступ с минимумом ключевых метаданных, таких как таксономия, район гербария и идентификаторы образцов, используя удачный опыт оцифровки Музея естественной истории в Париже (Le Bras et al., 2017).

Постепенно на нашем портале стали появляться и расширенные метаданные: полные расшифровки текста этикеток и геопривязки, т.е. значения координат места сбора каждого образца. К концу 2018 года было полностью завершено сканирование фондовой коллекции с разрешением 300 dpi (для типовых образцов 600 dpi). К этому моменту база содержала 971732 образца, 323015

геопривязок и 135812 расшифрованных этикеток. Кроме того, для всех образцов были получены результаты автоматического оптического распознавания символов (OCR) (Drinkwater et al., 2014), что позволило осуществлять быстрый поиск образцов по ключевым словам, имеющимся в печатных элементах этикеток. Эта работа была начата в марте 2018 года, и с тех пор OCR в фоновом режиме осуществляется с помощью программного продукта Tesseract для всех образцов Цифрового гербария МГУ.

С ноября 2017 года массив данных Гербария МГУ стал целиком индексироваться в Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>) (Global..., 2023). Это позволило как полностью интегрировать электронные сведения о наших образцах в международный агрегатор открытых данных о биоразнообразии, так и использовать инструменты GBIF в ежедневной кураторской, научной и изыскательской работе. На 14.01.2023 день данные Гербария Московского университета в GBIF были процитированы 857 раз (Seregin, 2023a).

Возможности Цифрового гербария МГУ привлекли внимание не только пользователей коллекций, но и кураторов других гербариев. С апреля 2019 года он стал консорциумом (Seregin, 2020),

Таблица 1. Основные показатели вклада участников консорциума Цифрового гербария МГУ (на 14.01.2023)
Table 1. Contributions from the participants of the Moscow Digital Herbarium consortium (as of 14 January, 2023)

Гербарий Herbarium	Дата подключения/ Date of the first publication	Число образцов/ Number of specimens	Число изображений/ Number of images	Число видов/ Number of species	Число геопривязок/ Number of georeferences	Число этикеток/ Number of captured labels
MW (Гербарий Московского государственного университета, г. Москва) / Moscow University Herbarium	16.10.2016	1052761	1031628	39579	693168	449389
МНА (Гербарий Главного ботанического сада РАН, г. Москва)/Herbarium of the Main Botanical Garden RAS, Moscow	01.04.2019	116362	116157	3962	87742	50053
IRKU (Гербарий Иркутского государственного университета)/Herbarium of Irkutsk State University	11.09.2020	42265	42154	1112	29050	30901
KUZ (Гербарий Кузбасского ботанического сада СО РАН, г. Кемерово)/Herbarium of the Kuzbass Botanical Garden SB RAS, Kemerovo	14.05.2020	19014	19104	1445	19004	19009
TUL (Гербарий Тульского государственного педагогического университета, г. Тула)/Herbarium of Leo Tolstoy Tula State Pedagogical University	25.12.2019	9808	9832	1208	9369	9807
TULGU (Гербарий Тульского государственного университета, г. Тула)/Herbarium of Tula State University	15.07.2021	5054	5054	818	5045	5054
KULPOL (Гербарий Музея-заповедника "Куликово поле", г. Тула)/Herbarium of the Museum-Reserve "Kulikovo Field", Tula	15.07.2021	3657	3656	594	3522	3657
ТКМ (Гербарий Тульского областного краеведческого музея, г. Тула) / Herbarium of the Tula Regional Museum of Local Lore	15.07.2021	2873	2855	878	1468	2866

Таблица 1. Окончание

Гербарий Herbarium	Дата подключения/ Date of the first publication	Число образцов/ Number of specimens	Число изображений/ Number of images	Число видов/ Number of species	Число геопривязок/ Number of georeferences	Число этикеток/ Number of captured labels
МАГ (Гербарий Института биологических проблем Севера ДВО РАН, г. Магадан)/Herbarium of the Institute of Biological Problems of the North FEB RAS, Magadan	22.10.2020	2604	2604	106	2537	2598
Цифровой гербарий МГУ (итог по консорциуму)/ Moscow Digital Herbarium (consortium total)		1 254 398	1 233 044	40 072	850 905	573 334

в котором организации-участники из Москвы, Иркутска, Кемерово, Тулы и Магадана публикуют свои коллекции в электронном виде по единым стандартам. Уже девять российских гербариев сделали свои коллекции доступными через платформу Цифрового гербария МГУ (табл. 1).

По результатам оцифровки фондов мы получили точные данные по объему и составу коллекций. Физическая коллекция Гербария Московского университета после ежегодного пополнения в 2022 году насчитывает 1 109 606 образцов, занимая по этому показателю 62-е место в мире и 2-е место в России (Thiers, 2023). Оцифрованы все образцы, кроме 62 тыс. образцов немонтированных исторических коллекций и коллекции лишайников, современный объем которой установить не представляется возможным. Отсканированные коллекции сосудистых растений и мохообразных представляют 40016 видов согласно номенклатуре "Catalogue of Life" (<https://www.catalogueoflife.org/>) (Catalogue..., 2023).

Отдельный интерес в методологическом плане и в плане полученных результатов представляет развитие блока пространственных данных. Данная статья логически продолжает публикацию 2017 года (Seregin, 2017), охватывая последние шесть лет работы Гербария Московского университета как центра документации и изучения разнообразия растений международного значения. Задачи у настоящей статьи две: (1) обозначить протоколы ввода и обработки пространственных данных в Цифровом гербарии МГУ и (2) охарактеризовать созданный нами массив геоданных Гербария Московского университета. Все приве-

денные цифры включают как сосудистые растения, так и мохообразные, если не указано иное. Отдельно подчеркнем, что речь идет не обо всех коллекциях участников консорциума Цифрового гербария МГУ (табл. 1), а только о Гербарии Московского университета (MW).

МАТЕРИАЛЫ И МЕТОДЫ

Категории геопривязок

Не у каждого образца из Гербария Московского университета, который отсканирован и размещен в открытом доступе, имеется геопривязка. Это процесс, который занимает время, и работа по полному покрытию коллекций геоданными ведется постоянно.

Все геоданные находятся в открытом доступе как на платформе Цифрового гербария МГУ, так и в GBIF (откуда их можно скачивать). Принципиально все геопривязки, которые у нас имеются, делятся на четыре категории:

- взятые напрямую из этикеток;
- установленные вручную для конкретного образца (или группы образцов с одинаковым текстом этикетки);
- установленные автоматически в результате работы алгоритма ИСТРА;
- образцы, у которых геопривязка невозможна в принципе.

Рассмотрим каждую из этих категорий, охарактеризовав источники и методы получения пространственных данных.

(1) **Геопривязки взяты напрямую с этикеток** (128957 шт., или 18.6%). Этикетки свежих образцов, как правило, содержат в явном виде географические координаты места сбора. Они вводятся в табличном виде операторами ввода этикеточных данных. В поле “Оператор геопривязки” при этом выбирается стандартный параметр “Коллектор”. С декабря 2022 г. такой ввод осуществляется волонтерами в модуле “Помогатор” (<https://plant.depo.msu.ru/pomogator>) (Pomogator, 2023) с автоматическим контролем ошибок на основании двойного независимого ввода данных.

Доля коллекций, которые непосредственно в этикетках содержат координаты места сбора, неуклонно растет. Стремление коллекторов, которые передают свои сборы в Гербарий МГУ, к точности и надежности документации находок можно продемонстрировать такими цифрами: в 2000 г., когда портативные GPS-приемники только стали появляться, 39.5% собранных образцов имели координаты в этикетках; в 2005 г. таковых уже было 55.2%; в 2010 г. эта доля выросла до 60.3%; а к 2015 г. дошла до 78.0%, сохраняясь примерно на том же уровне до самого последнего времени.

Формат записи географических координат на этикетках, поступающих на монтировку, самый разнообразный: 55.755831° (градусы), 55°45.35' или 55°45' (градусы и минуты), 55°45'20.99" или 55°45'21" (градусы, минуты и секунды), а также редкие случаи более экзотической записи форматов и непредсказуемых ошибок форматирования координат. При этом коллекторы используют как округление значений, так и, напротив, недопустимо большую точность, которая превосходит погрешность самих приборов. Среда Цифрового гербария МГУ работает со всеми форматами записи, преобразуя в вид, указанный в настройках пользователя. В GBIF данные импортируются строго в градусах в виде десятичной дроби с точностью до шестого знака после запятой.

Одна из особенностей непосредственного использования геоданных с этикеток заключается в том, что при вводе нами не осуществляется формальная оценка точности данных — это не соответствует международным стандартам ввода данных, но является общепринятой практикой (Marcer et al., 2021). Иными словами, цифры берутся как есть и отправляются на ручную геопривязку (см. ниже) только в том случае, если имеется явная ошибка указания координат коллектором. Ведется постепенная проверка координат, взятых непосредственно с этикеток.

(2) **Геопривязки установлены вручную** для конкретного образца или группы образцов с одинаковым текстом этикетки (240834 шт., или 34.7%). Ручная геопривязка — это установление места сбора конкретного образца с использованием

корпуса исторических и современных картографических и справочных материалов, которые широко представлены в открытом доступе в интернете. Кроме того, для каждой ручной геопривязки устанавливается значение точности геопривязки (см. ниже). В целом наши стандарты и решения соответствуют общим методическим рекомендациям (Chapman, Wiczorek, 2020).

Массив ручных геопривязок — результат работы свыше 60 операторов, которые в разное время в качестве основной работы или в ходе решения научно-исследовательских задач осуществляли геопривязку образцов из фондов Гербария МГУ. Три автора настоящей статьи сделали 47.5% ручных геопривязок (В.Н. Пашкина — 62837 шт., И.Н. Пospelov — 27069 шт., А.П. Серегин — 24514 шт.).

Для осуществления этого этапа ключевыми исходными данными являются, как правило, не изображения образцов, а расшифрованные и введенные в систему транскрипции тексты этикеток. В Цифровом гербарии МГУ используется строгий протокол ввода этикеточных данных, который позволяет жестко унифицировать процедуры ввода различными операторами с минимальными разночтениями. Так, например, используется единое текстовое поле “*Вся география и экология (текст этикетки)*”, а страна и административные единицы устанавливаются, как правило, автоматически на основе введенных координат.

Операторы ручной геопривязки в качестве исходных данных работают с табличными расшифровками текста этикеток, ограничиваясь выборкой образцов по какому-нибудь району или отделу гербария. Это позволяет быстро выявлять повторяющиеся топонимы и восстанавливать маршруты конкретных исследователей, что уменьшает время ввода и повышает качество взаимной геопривязки близко расположенных пунктов работ. Если работа ведется в табличном редакторе, то на разных этапах производится сортировка массива по различным параметрам. Так, сортировка по дате сбора позволяет устанавливать у анонимных сборов фамилию коллектора, опираясь на место сбора и почерк на этикетке. В целом, сортировка по тексту этикетки ускоряет работу, позволяя выявить одни и те же локации, где в разные годы работали разные коллекторы.

Кроме того, именно на этапе геопривязки операторы зачастую выполняют окончательное редактирование (чистку) массива введенного текста многочисленных этикеток.

(3) **Геопривязки установлены автоматически** в результате работы алгоритма ИСТРА (314215 шт., или 45.3%). Алгоритм ИСТРА (Интеллектуальная Система Топонимического Распознавания И Атрибутирования) — разработка О.Н. Платко, программиста “Проекта Скулачева”, архитектора и

главного разработчика Цифрового гербария МГУ. На сегодняшний день ИСТРА в фоновом режиме объединяет все образцы из базы в группы по паре “коллектор–дата” с дополнительной нормировкой по географическим районам гербария. На ручную геопривязку отправляется один образец из группы, а все образцы, собранные коллектором в течение одного дня, привязываются к одному и тому же месту сбора. Такие автоматически сгруппированные и привязанные к карте образцы получают особую пометку в базе, и постепенно их место сбора может уточняться за счет ручной геопривязки.

Для обеспечения работы алгоритма ИСТРА еще в июле 2018 года силами коммерческого партнера был осуществлен массовый ввод пары “коллектор–дата” для 80% образцов Гербария Московского университета. По сути, эти данные были введены для всех образцов, в которых единственная дата сбора и фамилия коллектора были указаны однозначно и полно, однако без нормоконтроля, т.е. однозначного сопоставления указанной фамилии коллектора и конкретного человека, что является трудной задачей (Güntsch et al., 2021). На следующий день система ИСТРА автоматически привязала к карте 63 тыс. образцов и общее число геопривязок составило 217971 шт. — каждый четвертый оцифрованный образец Гербария МГУ получил в тот момент метку на карте.

Подход с вводом пары “коллектор–дата” в качестве обязательных метаданных позволил, с одной стороны, добиться группировки массива по описанным принципам, а с другой стороны — систематизировать дальнейшую выдачу образцов на ввод полного текста этикеток. Так, операторам ввода текстовых данных отправляется теперь строго отсортированная выборка образцов, день за днем представляющая сборы одного и того же исследователя. Операторы заметно улучшили качество ввода текстовых данных, поскольку у них появилась возможность разобраться с почерком одного и того же человека, вводить повторяющиеся топонимы, а зачастую — копировать текст этикеток, которые выдаются на ввод в хронологическом порядке.

Система ИСТРА работает не только с образцами Гербария Московского университета, а охватывает все коллекции участников консорциума. Это позволяет создать единую сеть данных, полностью включающую как оригинальные коллекции каждого гербария, так и дублиеты, поступающие из других коллекций. Кроме того, любая геопривязка, внесенная в систему любым участником консорциума, пополняет общую библиотеку референсных геоданных, необходимых для работы системы ИСТРА. Инструменты и протоколы автоматического и полуавтоматического поиска и исправления ошибок (как внутри плат-

формы, так и в GBIF) позволяют поддерживать чистоту массивов данных от взаимных нестыковок.

(4) Кроме того, имеется относительно небольшая группа образцов, у которых **установить геопривязку принципиально невозможно**. Это связано с отсутствием указаний места сбора на образце или лишь с самыми общими указаниями. Таких образцов 10 460 штук (или 1.5%). Вот реальные примеры текста этикеток образцов из европейской части России, которые получили пометку о невозможности геопривязки:

“Deserto”;

“Gouv. Samara. Steppenabhang”;

“Астраханская обл. Полузакрепленные пески”;

“В лесах, в кустарниках”;

“В лесу, под пологом”;

“Встречается по всей ИПО, как спец[иализированный] сорняк оз[имой] ржи, реже в друг[их] культурах”;

“Жел. дорога. Часто”;

“Иваново-Вознесенская губерния”;

“Костромская обл.”;

“Курская губ.” и т.д.

Специальным случаем является ситуация, когда на этикетке место сбора отсутствует, однако в явном виде указаны дата сбора и коллектор. Такой образец входит в соответствующую группу образцов по паре “коллектор–дата” и получает автоматическую геопривязку с помощью алгоритма ИСТРА (см. выше).

Оценка точности геопривязок

Мы используем в качестве обязательного параметра при ручной геопривязке переменную “Точность”. Это радиус круга (в километрах) с центром в предположительном месте сбора образца. Такой круг охватывает все возможные места сбора образца, исходя из той неопределенности, что заложена в текстовом описании места сбора в этикетке. На переменную “Точность” влияет и число используемых румбов направлений (4, 8 или 16), и точность определения расстояния от условной точки (например, населенного пункта), и характер работы с различными картами в прошлом, и взаимное расположение точечных, линейных и площадных объектов, упомянутых в этикетке.

Точность установления координат в результате ручной геопривязки может быть как очень высокой (например, если сохранилось конкретное здание, местоположение которого легко установить по спутниковым снимкам), так и чрезвычайно низкой. Например, у образца MW0210476 с этикеткой “Кольский п-ов, Терский берег Мурманская область, Терский р-он. Сухая лишайни-

ково-кустарничковая тундра на морской террасе” точность геопривязки составляет 260 км. Используя инструменты GBIF, пользователи имеют возможность отсечь неточные геопривязки, указывая значения этой переменной.

В целом случаи, когда место сбора указано в тексте этикетки неточно, нередки. Характер этой неопределенности таков, что предполагаемое место сбора представляет собой не одну, а массив равновероятных точек. Здесь используется следующий подход: параметр “Точность” здесь — это минимальный радиус круга, в пределах которого реальное место сбора находится с вероятностью, близкой к 1, а в качестве геопривязки выступают координаты центра данного круга. Например, для образца одного из гибридов ив MW0307869 (“Канин п-ов, южн. часть. Оз. Яжемское. Низкий берег”) в качестве геопривязки использованы координаты центра озера Яжемского, а точность равна 2 км (радиус озера плюс 250 м), хотя центр озера, безусловно, не является местообитанием этой ивы.

В целом указываемая точность тем ниже (радиус больше), чем больше размер населенного пункта, чем менее населена данная территория, чем дальше расстояние до условной точки, указанной в этикетке. Для старых сборов приходится увеличивать радиус, в связи с неточностью самих старых карт, с изменением величины и положения населенных пунктов за последние два века и другими событиями, такими как прокладка и исчезновение дорог, осушение местности, создание водохранилищ и т.д. Мы стараемся сверяться с двумя картами, чтобы свести к минимуму возможную погрешность. Основные подходы к оценке точности, используемые в Цифровом гербарии МГУ, размещены в открытом доступе в качестве стандартной инструкции (Seregin, 2022).

В консорциуме Цифрового гербария МГУ каждый из гербариев самостоятельно отвечает за объем и качество внесенных геоданных. Редактор портала ведет мониторинг возможных ошибок, связанных с качеством ввода текстовых метаданных, элементов курирования коллекции и геопривязок места сбора. В текстовых метаданных ошибки, влияющие на качество геоданных, могут содержаться в дате сбора, фамилии и инициалах коллектора, расшифровке текста этикетки, а также в цифрах координат, введенных с этикеток. Источником ошибок среди элементов курирования коллекции чаще всего является неверно указанный район гербария. Сами геопривязки, поступившие от операторов, могут содержать опечатки, ошибки установления топонимов, неточности интерпретации этикеток, сбой форматов координат, и проч.

Проверка введенных геопривязок осуществляется также пользователями Цифрового гербария

МГУ путем краудсорсинга. В паспорте каждого образца имеется ссылка “Сообщить об ошибке”, где любой пользователь без предварительной регистрации может сообщить нам о той или иной проблеме с вводом образца. Мы регулярно получаем сообщения как об ошибках в геопривязках, так и о неточностях трактовок текста этикеток или некорректной работе алгоритма ИСТРА. Поступившие от пользователей данные используются для улучшения базы.

Красноярский край: пример монографической работы с геоданными

Отдел Сибири и Дальнего Востока в Гербарии Московского университета был оцифрован в первую очередь и уже в октябре 2016 г. к нему был открыт онлайн-доступ.

И.Н. Поспелов поставил перед собой задачу выявления всех сборов сосудистых растений из Таймырского района Красноярского края. В процессе работы выяснилось, что для полного выявления коллекций с Таймыра необходим просмотр фондов всего района Центральная Сибирь (S3) Гербария МГУ, включавшего на тот момент около 19 500 образцов. Примерно треть (около 6300 образцов) составляли сборы И.Н. и Е.Б. Поспеловых и участников совместных экспедиций. Для этих коллекций с середины 1990-х годов велась база данных с координатами всех собранных и переданных в МГУ образцов, то есть достаточно было связать ID образца в MW (его роль выполняет штрихкод) с уникальным коллекторским номером, чтобы отправить данные в базу Цифрового гербария МГУ.

Остальные образцы по Центральной Сибири были привязаны И.Н. Поспеловым “с нуля” в течение 2017 года (Pospelov, 2018), хотя примерно в 10% случаев координаты района сбора разной степени точности на этикетках все же имелись. Например, в некоторых образцах начала XX века на этикетках использовались значения долготы от Пулковского меридиана. В целом, район S3 Гербария МГУ охватывает Красноярский край без горных массивов Алтае-Саянской горной страны, расположенных на юге края.

В этом случае первым условием привязки образца являлись отличные знания географической номенклатуры и истории ботанических исследований региона. Оператор геопривязки хорошо знал природные условия региона и имел опыт полевых работ здесь. Это позволяло “поставить себя на место коллектора”, например, оценить реальность суточных перемещений коллектора в соответствии с условиями местности, что существенно повышало качество привязки. Во многих случаях коллекторы ныне здравствуют, и при необходимости могли при личном обращении уточнить место

или хотя бы район сбора (Pospelov, Pospelova, 2019).

После нахождения искомого района (по топониму) осуществлялся переход по координатам в проект в среде ArcGIS. Координаты уточнялись по описанию места сбора и при помощи специального скрипта одним кликом переносились в табличный редактор, там же вводились коллекторы, дата сбора и экспертная точность привязки в км.

После привязки первых 1000 образцов определились районы работ основных коллекторов гербария, что позволило в проекте ArcGIS установить закладки на эти районы (около 25 шт.), что позволило выполнять до 500 привязок в день. Как показала практика, после подобного “погружения” в коллекцию большинство сборов может быть привязано с точностью 0.5–5 км (Pospelov, 2020). В ряде случаев имела бы смысл привязка не к точке с радиусом точности, а к площадному или линейному объекту (участок течения реки, прохождения дороги, окрестностям населенного пункта и т.д.), однако эта практика пока не получила широкого распространения в международных базах данных.

РЕЗУЛЬТАТЫ

На 14.01.2023 две трети образцов Гербария МГУ (693168 шт., или 66%) имеют электронные метки на карте, в то время как 359593 образца их пока не имеют. Характеристику массива геоданных Гербария Московского университета необходимо дать как в географическом аспекте (покрытие геоданных для разных регионов), так и в аспекте точности геопривязок.

Основная работа по формированию массива геоданных в Гербарии Московского университета (MW) ведется для “Атласа флоры России”, который основан на множестве различных источников. Исходя из этого, именно территория России является приоритетом с точки зрения геопривязки образцов (табл. 2). Больше 80% образцов привязано к карте из России (80.8%), Вьетнама (85.2%), Канады (96.2%), а также не попавших в таблицу Афганистана, Парагвая, Кипра и Великобритании.

Распределение образцов сосудистых растений с территории России по районам, принятым для курирования фондов Гербария Московского университета, дано в табл. 3. В целом мы ввели геопривязки для 81.6% образцов, собранных в нашей стране. Наибольший объем как физических образцов, так и геопривязок имеется в Гербарии МГУ по Европейской России – Москве и Подмосковию (район E4a), Центральному Нечерноземью (E4), Центральному Черноземью (E6), Европейскому Северу (E1) и Крыму (KRYM). По азиатской части России более полно пред-

ставлены коллекции по югу Дальнего Востока (S6) и Алтае-Саянскому региону (S2), а по Кавказу – в основном, по Карачаево-Черкесии (K1b).

За редким исключением доля образцов, которые мы смогли привязать к карте, не превышает 90% по каждому району. Остальные 10%, как правило, приходится на образцы, у которых геопривязка невозможна (старые неотыскиваемые топонимы, образцы без этикеток и проч.). Осторожно можно предположить, что нам удастся привязать к карте еще около 50–55 тыс. российских образцов из числа имеющихся в фондах Гербария Московского университета, а остальное, к сожалению, представляет собой “мертвый объем” (разумеется, только с географической точки зрения).

Каждый год мы включаем в фонды Гербария Московского университета около 15 тыс. образцов – это средняя цифра ежегодных поступлений за последние десять лет. Большинство из них происходят с территории России. С 2016 года мы полностью включили в свои фонды несколько малых коллекций: гербарий лаборатории устойчивости лесных экосистем Костромского университета (около 1000 листов – Костромская область, 2016 г.); гербарий Бадхызского заповедника (около 1000 листов – юго-восточная Туркмения, 2017 г.); гербарий Института проблем почвоведения РАН в г. Пушкино (около 1000 листов – преимущественно Якутия, 2018 г.); гербарий Музея природы в г. Владимир (около 500 листов – Владимирская область, 2019 г.); гербарий Института лесоведения РАН (около 28000 листов – бывший СССР, КНР, 2022 г.).

Постепенно эти образцы становятся доступными в Цифровом гербарии МГУ не только в виде отсканированных изображений, но и с метками на карте. Физическое объединение всех гербариев в один произошло в столь разных странах как Узбекистан, Нидерланды и Израиль. По-видимому, процесс вливания малых гербариев в состав крупных коллекций продолжится в ближайшем будущем и в России.

ОБСУЖДЕНИЕ

Считается, что геопривязка образцов завершает процесс их оцифровки и включения в научный оборот после этапов (1) подготовки, (2) сканирования, (3) обработки изображений и (4) ввода данных этикеток (Nelson et al., 2012). Однако наш опыт показывает, что предварительный ввод ключевых метаданных (пара “коллектор–дата”) и геопривязка образцов с использованием этого параметра заметно упрощает дальнейший ввод этикеточных данных и делает его эффективнее. Саму работу по геопривязке коллекций мы ведем с использованием современных информационных технологий и различных источников (текстовых,

Таблица 2. Распределение образцов Гербария Московского университета по странам и характеру геоданных (на 14.01.2023). Таблица отсортирована по общему числу геопривязок
Table 2. Distribution of specimens from the Moscow University Herbarium by country and sources of geodata (as of 14 Jan 2023). The table is sorted by the total number of georeferences

№ п/п	Страна/Country	Число образцов/ Number of specimens	Есть геопривязки/ Specimens with geodata	Нет геопривязок/ Specimens with no geodata	Точные геопривязки/ Manual georeferences	Геопривязки по ИСТРЕ/ Automatic georeferences	Невозможно привязать/ Georeferencing impossible
1	Россия (с Крымом)/ Russia (with Crimea)	714628	577147	137481	328354	240200	9688
2	Монголия/Mongolia	27890	16827	11063	3326	13501	0
3	Казахстан/ Kazakhstan	53308	14248	39060	1474	12774	18
4	Киргизия/ Kyrgyzstan	18583	8093	10490	2036	6057	22
5	Украина/Ukraine	31570	7619	23951	1067	6551	12
6	Азербайджан/ Azerbaijan	15403	7325	8078	1501	5776	5
7	Армения/Armenia	11147	5571	5576	1208	4338	1
8	Вьетнам/Vietnam	5974	5088	886	3868	807	0
9	Узбекистан/ Uzbekistan	11047	3928	7119	644	3284	1
10	Южная Осетия/ South Ossetia	5437	3411	2026	859	2520	0
11	Туркмения/Turkmenistan	10841	3305	7536	223	3082	1
12	Канада/Canada	3432	3301	131	2222	1079	71
13	Таджикистан/ Tajikistan	10327	2962	7365	477	2483	0
14	Грузия/Georgia	10608	2573	8035	607	1950	16
15	Эфиопия/Ethiopia	2958	2209	749	1164	1045	10
16	Латвия/Latvia	4044	2200	1844	1087	1103	3
17	КНР/China	5662	2185	3477	1738	447	0
18	Белоруссия / Belarus	7229	2005	5224	1263	741	0
19	Турция / Turkey	3670	1349	2321	843	506	0
20	Иран / Iran	2619	1284	1335	1273	11	1

картографических, архивных), которые находятся как в открытом доступе, так и в доступных хранилищах. Еще несколько лет назад наши возможности в этом плане были гораздо скромнее.

Так, для геопривязки сборов европейской части России конца XIX — начала XX вв. используются карты Шуберта, Менде, а также военные трехверстовки отдельных губерний, доступные на сайтах “RetroMap: Старые карты России и зарубежья” (<http://retromap.ru/>) (RetroMap, 2023) и “ЭтоМесто” (<http://www.etomesto.ru/>) (EtoMesto, 2023). На этих ресурсах растровые отсканированные карты имеют привязку к современным картам и координатной сетке с возможностью наложения слоев или параллельной работы на двух картах. Также на сайте “ЭтоМесто” размещены карты РККА масштабом до 1 : 100 000, позволяющие установить места сбора образцов начала — середины 20 века. Крайне полезны также карты Генштаба СССР, где отмечено множество точечных объектов (мельницы, плотины, башни, броды и т.д.). В редких случаях используются также космоснимки и результаты аэрофотосъемки — как исторические, так и современные.

Определенную сложность представляет геопривязка образцов, место сбора которых указано с точностью до лесного квартала. Причины этого — регулярные изменения в лесоустройстве отдельных лесхозов, а также отсутствие поквартальных архивных схем лесничеств в открытом доступе. Иногда номера кварталов отмечены на картах ГГЦ масштаба 1 : 25 000 с сайта “Маршруты.Ру” (<https://maps.marshruty.ru/>) (Marshruty..., 2023). Геопривязка сборов с территории заказников, национальных парков и заповедников осуществляется с использованием карт данных ООПТ. При этом важно, чтобы схема кварталов была актуальна на дату сбора — в этом отношении ценен сайт “ООПТ России” (<http://www.oopt.aari.ru/>) (ООПТ, 2023).

Информация о местоположении урочищ, не обозначенных на картах, зачастую кажется утраченной. Тем не менее, порой удается ее установить. Сведения о таких топонимах можно в наиболее затруднительных случаях получить, обратившись в библиотеку или краеведческий музей соответствующего района. Много ценной информации содержится также на краеведческих сайтах и форумах: например, на краеведческом портале “Родная Вятка” (<http://rodnaaya-vyatka.ru/>) (Rodnaaya..., 2023) и ряде других. На таких ресурсах оцифрованы списки населенных мест по уездам Российской Империи, с которыми необходимо сверяться при ручной геопривязке. Делается это из-за возможного наличия на искомой территории нескольких населенных пунктов с одинаковым названием. В процессе работы учитываются

также изменения границ и названий административных единиц.

Кроме того, один и тот же географический объект в тексте этикетки и на картах разных лет может иметь разное название (которое менялось с течением времени, либо приводится на другом языке). Например, расположенное в Хибинских горах озеро Гольцовое также фигурирует под названиями оз. Пай-кун’явр или Пай-Куныявр. Так, работа по ручной геопривязке регулярно приобредает элементы исторического исследования.

В Центральной Сибири для первичной привязки использовались топографические карты масштаба 1 : 200 000 советского периода (состояние местности на 1960—1980-е гг.). Непосредственно поиск мест сбора осуществлялся в картографических веб-сервисах — в основном “Яндекс.Карты” (<https://yandex.ru/maps/>) (Yandex, 2023). Если же в тексте этикетки использовались устаревшие топонимы или названия ныне отсутствующих населенных пунктов и иных объектов, производился поиск в общих поисковых системах (Google, Яндекс). Почти во всех случаях удавалось найти хотя бы косвенное упоминание объекта, анализ которого позволял идентифицировать и сам объект (хотя и здесь не обходилось без казусов).

Для сборов XIX — начала XX века с этой территории использовались карты из “Атласа Азиатской России” (Glinka et al., 1914) хотя и имеющие довольно мелкий масштаб (например, 1 : 3 360 000, 80 верст в дюйме для Енисейской губернии), но содержащие для большей части края населенные пункты, ныне забытые даже старожилыми районов. При необходимости изучались отчеты экспедиций, публикации коллекторов и даже биографические сведения о них. Возможность и качество привязки часто не зависит даже от времени сбора — так, сборы XVIII — начала XIX века из окрестностей Красноярска привязаны с довольно высокой точностью, т.к. выполнены на известных горных останцах в пригородах.

Очень важную помощь в работе с привязкой образцов по Центральной Сибири оказали также монография “История открытия и освоения Северного Морского пути”, в особенности тома 3 и 4 (Belov, 1959a, b), посвященные советскому периоду освоения Арктики, где описываются или как минимум упоминаются почти забытые в настоящий момент экспедиционные исследования севера России. Кроме того, недавно была издана монография Н.Н. Тупицыной и др. (Tupitsyna et al., 2016), содержащая почти полный обзор флористических исследований на территории Красноярского края.

Обширный объем привязанных к карте данных Гербария Московского университета был использован для создания предварительных сеточных

Таблица 3. Распределение образцов сосудистых растений Гербария Московского университета, собранных в России, по районам гербария (на 14.01.2023)
Table 3. Distribution of specimens from the Moscow University Herbarium collected in Russia by curatorial areas (as of 14 January, 2023)

Код и название района гербария/ Code and name of the herbarium area	Число образцов/ Number of specimens	Есть геопривязки/ Specimens with geodata	Доля с геопривязкой, %/ Percentage of samples with geodata, %	Нет геопривязок/ Specimens with no geodata	Точные геопривязки/ Manual georeferences	Геопривязки по ИСТРЕ/ Automatic georeferences	Невозможно привязать/ Georeferencing impossible
E1 Северный район/ Northern region	39408	34668	88.0	4740	19538	14698	32
E2 Северо-Западный район/ North-Western region	15626	10259	65.7	5367	7197	2665	41
E3 Западный район/ Western region	6828	5178	75.8	1650	3475	1416	183
E4 Центральный район/ Central region	53936	49583	91.9	4353	23821	24916	1315
E4a Московская область и Москва/Moscow region	77851	69241	88.9	8610	34091	32877	1846
E5 Центральный лесной район/Central forest region	15649	13793	88.1	1856	12290	1157	232
E6 Центральный лесостепной район/Central forest-and-steppe region	51389	35966	70.0	15423	15758	19718	303
E7 Волгоко-Камский район /Volga-Kama region	14429	12440	86.2	1989	6740	5546	66
E8 Средневолжский район/ Middle Volga region	25769	25209	97.8	560	20379	4763	24
E9 Нижневолжский район/ Lower Volga region	25283	22666	89.6	2617	12004	10561	713
E10 Восточный район/ Eastern region	24460	23211	94.9	1249	10561	12451	321
E12a Ростовская область/ Rostov Oblast	8998	7483	83.2	1515	6738	740	1101
K1a Краснодарский край и Адыгея/Krasnodar Krai and Adygea	16642	14050	84.4	2592	9400	4125	1133

Таблица 3. Окончание

Код и название района гербария/ Code and name of the herbarium area	Число образцов/ Number of specimens	Есть геопривязки/ Specimens with geodata	Доля с геопривязкой, %/ Percentage of samples with geodata, %	Нет геопривязок/ Specimens with no geodata	Точные геопривязки/ Manual georeferences	Геопривязки по ИСТРЕ/ Automatic georeferences	Невозможно привязать/ Georeferencing impossible
K1b Ставропольский край, КЧР и КБР / Stavropol Krai, Karachay-Cherkessia and Kabardino-Balkaria	20721	10339	49.9	10382	3614	6568	17
K1c Северная Осетия, Ингушетия и Чечня / North Ossetia, Ingushetia and Chechnya	9264	3948	42.6	5316	774	3117	0
K2 Дагестан / Dagestan	6048	2671	44.2	3377	735	1862	16
K3 Черноморское побере- жье Кавказа / Black Sea Coast	11668	10806	92.6	862	7200	3287	82
KRYM Крым / Crimea	34032	30474	89.5	3558	14619	15375	1447
S1 Западная Сибирь / West- ern Siberia	16230	12492	77.0	3738	7598	4809	136
S2 Алтай и Саяны / Altai and Sayany Mountains	34486	24736	71.7	9750	12508	12070	23
S3 Центральная Сибирь / Central Siberia	23290	21815	93.7	1475	21642	171	0
S4 Прибайкалье и Забайка- лье / Baikal and Transbaikal region	27548	15490	56.2	12058	9033	6140	3
S5 Якутия / Yakutia	17372	11100	63.9	6272	3825	6987	8
S6 Дальний Восток / Russian Far East	35740	30999	86.7	4741	13257	17509	0
S7 Чукотка и Камчатка / Chukotka and Kamchatka	26942	23230	86.2	3712	13908	9144	545
Россия в целом / Russia (total)	639609	521847	81.6	117762	290705	222672	9587

карт “Атласа флоры России” по квадратам 100 × 100 км. Для создания исходного массива данных использованы как предварительно очищенные данные из GBIF, так и из ряда других источников. Эти карты размещены в открытом доступе в Цифровом гербарии МГУ (<https://plant.depo.msu.ru/>) (Seregin, 2023b) и видны в карточке каждого образца, собранного в России. В общей сложности опубликованы карты для 9000 видов. Постепенно мы ведем проверку карт, устраняя ошибки и неточности.

Рост количества геоданных в открытом доступе имеет также важное репутационное значение. По показателю глубины и качества геопривязок мы заметно опережаем многие крупные гербарии мира, размещенные в GBIF. Несмотря на скромные по объему в мировом масштабе коллекции, по абсолютному числу геопривязок Гербарий Московского университета занимает 9-е место в мире (на 31.01.2023), уступая Миссури (МО) (Teisher, Stimmel, 2022), Лейдену (L) (Bijmoer et al., 2023), Нью-Йорку (NY) (Ramirez et al., 2022), Претории (PRE) (Ranwashe et al., 2022), Лунду (LD) (Wittzell, Shah, 2022) и трем австралийским гербариям (CANB, MEL, PERTH).

Создание и развитие Цифрового гербария МГУ со свободным доступом к данным фундаментально изменило доступность коллекций Гербария Московского университета и характер работы с ними – как справочной, так и научно-исследовательской. Так, за 2022 год на портал Цифрового гербария МГУ заходили пользователи с 33870 устройств (без учета роботов). В 2021 г. этот показатель был 23670 единиц, а в 2020 г. – 23180. Кроме того, ежедневно 88 выгрузок данных с портала GBIF включают сведения об образцах из Гербария Московского университета. Во многом, это стало результатом постоянной работы по увеличению объема геоданных, связанных с гербарными образцами.

БЛАГОДАРНОСТИ

Работа А.П. Серегина в части геопривязки коллекций из России выполнена за счет гранта Российского научного фонда (проект № 21-77-20042, <https://rscf.ru/project/21-77-20042/>). Работа И.Н. Поспелова выполнена в рамках гостемы “Фундаментальные проблемы охраны живой природы и рационального использования биоресурсов” (№ 0109-2019-0009). В.Н. Пашкина выражает благодарность сотрудникам Ивanteeвской межпоселенческой центральной библиотеки и лично В.П. Черняевой, сотрудникам МБУК “Трубчевский музей и планетарий” (и лично М.М. Шурубкину).

СПИСОК ЛИТЕРАТУРЫ

- [Belov] Белов М.И. 1959a. История открытия и освоения Северного Морского пути. Т. 3. Советское арктическое мореплавание 1917–1932 гг. Л. 511 с.
- [Belov] Белов М.И. 1959b. История открытия и освоения Северного Морского пути. Т. 4. Научное и хозяйственное освоение Советского Севера. 1933–1945 гг. Л. 616 с.
- Bijmoer R., Scherrenberg M., Creuwels J. 2023. Naturalis Biodiversity Center (NL) – Botany. Naturalis Biodiversity Center. Occurrence dataset <https://doi.org/10.15468/ib5ypt> accessed via GBIF.org on 2023-01-31.
- Catalogue of Life. 2023. <https://www.catalogueoflife.org/>
- Chapman A.D., Wiczorek J.R. 2020. Georeferencing best practices. Version 1.0. Copenhagen, Denmark, GBIF Secretariat. 112 p. <https://doi.org/10.15468/doc-gg7h-s853>
- Drinkwater R.E., Cubey R.W., Haston E.M. 2014. The use of Optical Character Recognition (OCR) in the digitisation of herbarium specimen labels. – *PhytoKeys*. 19(38): 15–30.
- EtoMesto.ru. 2023. <http://www.etomesto.ru/>
- [Glinka et al.] Глинка Г.В., Тхоржевский И.И., Цветков М.А. 1914. Атлас Азиатской России. Издание переселенческого управления главного управления землеустройства и земледелия. СПб. 71 с. <http://kartolog.ru/2009/09/atlas-aziatskoj-rossii-1914-g/?ysclid=ld1jlbntnsj906089553>
- Global Biodiversity Information Facility. 2023. <https://www.gbif.org/>
- Güntsche A., Groom Q., Ernst M., Holetschek J., Plank A., Röpert D., Fichtmüller D., Shorthouse D.P., Hyam R., Dillen M., Trekels M., Haston E., Rainer H. 2021. A botanical demonstration of the potential of linking data using unique identifiers for people. – *PLoS ONE*. 16(12): e0261130. <https://doi.org/10.1371/journal.pone.0261130>
- Le Bras G., Pignal M., Jeanson M.L., Muller S., Aupic C., Carré B., Flament G., Gaudeul M., Gonçalves C., Invernón V.R., Jabbour F., Lerat E., Lowry P.P., Offroy B., Pimparé E.P., Poncy O., Rouhan G., Haevermans T. 2017. The French Muséum national d’histoire naturelle vascular plant herbarium collection dataset. – *Scientific Data*. 4(1): 1–16.
- Marcet A., Groom Q., Haston E., Uribe F. 2021. Natural History collections: georeferencing survey report: current georeferencing practices across institutions worldwide. – *Zenodo*. <https://doi.org/10.5281/zenodo.4644529>
- Marshruty.ru 2023. <https://maps.marshruty.ru/>
- Nelson G., Paul D., Riccardi G., Mast A.R. 2012. Five task clusters that enable efficient and effective digitization of biological collections. – *ZooKeys*. 209: 19–45. <https://doi.org/10.3897/zookeys.209.3135>
- [ООРТ] ООПТ России. 2023. <http://www.oopt.aari.ru/>
- [Pomogator...] Помогатор Цифрового гербария МГУ. 2023. <https://plant.depo.msu.ru/pomogator>
- Pospelov I.N. 2018. The experience of geographical fixing Central Siberia herbarium collections from the Mos-

- cow State University (MW) collection. — In: Abstracts of the conference “Information Technologies in the Research of Biodiversity” (BIT–2018). Irkutsk. P. 105–106.
- [Pospelov] Поспелов И.Н. 2020. Уровни географической локализации пространственных данных в информационных ресурсах о биологическом разнообразии. — В кн.: Информационные технологии в исследовании биоразнообразия: материалы III Национальной науч. конф. с междунар. участием, посвященной 100-летию со дня рождения академика РАН П.Л. Горчаковского. — Екатеринбург. С. 448–451.
- [Pospelov, Pospelova] Поспелов И.Н., Поспелова Е.Б. 2019. О возможностях исследования флористического разнообразия на основании анализа гербарных сборов (Цифровой гербарий MW). — В кн.: Ботанико-географические исследования. Камелинские чтения: сб. науч. тр. Пермь. С. 134–138.
- Ramirez J., Watson K., McMillin L., Gjieli E., Sessa E. 2022. The New York Botanical Garden Herbarium (NY). Version 1.56. The New York Botanical Garden. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/6e8nje>
- Ranwashe F. et al. 2022. Botanical Database of Southern Africa (BODATSA): Botanical Collections. Version 1.16. South African National Biodiversity Institute. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/2aki0q>
- RetroMap. 2023. RetroMap: Старые карты России и зарубежья. <http://retromap.ru/>
- [Rodnaya...] Родная Вятка. 2023. <http://rodnaya-vyatka.ru/>
- [Seregin] Серегин А.П. 2017. Гербарий Московского университета (MW) сегодня: фонды, онлайн доступ и научная работа. — Бот. журн. 102 (3): 281–308.
- Seregin A.P. 2018. The largest digital herbarium in Russia is now available online! — Taxon. 67 (2): 463–467. <https://doi.org/10.12705/672.34>
- Seregin A.P. 2020. Moscow Digital Herbarium: a consortium since 2019 — Taxon. 67 (2): 417–419. <https://doi.org/10.1002/TAX.12228>
- [Seregin] Серегин А.П. 2022. Инструкция по геопривязке образцов Гербария Московского университета (MW). Вариант “Монголия”. Версия 1.0. М. <https://doi.org/10.13140/RG.2.2.14879.84645>
- Seregin A. 2023a. Moscow University Herbarium (MW). Version 1.265. Lomonosov Moscow State University. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/cpnhcc>
- Seregin A.P. (Ed.). 2023b. Moscow Digital Herbarium: Electronic resource. Moscow State University, Moscow. <https://plant.depo.msu.ru/>
- Thiers B.M. 2023. Index Herbariorum. <https://sweetgum.nybg.org/science/ih/>
- [Tupitsyna et al.] Тупицына Н.Н., Шауло Д.Н., Гуреева И.И. 2016. Обзор флористических исследований Средней Сибири. Красноярск. 253 с. <https://doi.org/10.17223/9785946216371/33> (<https://vital.lib.tsu.ru/vital/access/services/Download/vtls:000617746/SOURCE1>)
- Teisher J., Stimmel H. 2022. Tropicos Specimen Data. Missouri Botanical Garden. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/hja69f>
- Witzell H., Shah M. 2022. Lund Botanical Museum (LD). Lund Botanical Museum (LD). Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/c4w4co>
- [Yandex] Яндекс.Карты. 2023. <https://yandex.ru/maps/>

GEODATA OF THE MOSCOW UNIVERSITY HERBARIUM: DATA STRUCTURE AND GEOREFERENCING METHODICS

A. P. Seregin^{a, #}, V. N. Pashkina^a, and I. N. Pospelov^b

^aLomonosov Moscow State University
Leninskiye Gory, 1, Moscow, 119991, Russia

^bSevertsov Institute of Ecology and Evolution of RAS
Leninsky Ave., 33, Moscow, 119071, Russia

[#]e-mail: botanik.seregin@gmail.com

Digitisation of the Moscow University Herbarium (MW) was started in the spring of 2015. Since October 2016, we have been publishing the images of specimens on the portal of the Moscow Digital Herbarium (<https://plant.depo.msu.ru/>) in open access mode. Gradually, we started to publish extended metadata on our portal, like full transcriptions of labels and georeferences, i.e. the coordinates of collection sites for each specimen. As of January, 14, 2023, two thirds of the specimens of the Moscow University Herbarium (693168 specimens, or 66%) are georeferenced, while 359593 specimens are still in work. The article describes the geodata of the Moscow University Herbarium both in the geographical aspect (data coverage for different regions) and in the aspect of georeferencing accuracy. General guidelines for georeferencing herbarium collections are given.

Keywords: herbarium, spatial data, plant geography

ACKNOWLEDGEMENTS

The work of A.P. Seregin under georeferencing of collections from Russia was supported by a grant from the Russian Science Foundation (project No. 21-77-20042, <https://rscf.ru/en/project/21-77-20042/>). The work of I.N. Pospelov was made within the framework of the state assignment “Fundamental Problems of Wildlife Conservation and Rational Use of Bioresources” (No. 0109–2019–0009). V.N. Pashkina expresses her gratitude to the staff of the Ivanteevka Intersettlement Central Library and personally to V.P. Chernyaeva, employees of the Trubchevsk Museum and Planetarium (and personally to M.M. Shurubkin).

REFERENCES

- Belov M.I. 1959a. Istoriya otkrytiya i osvoeniya Severnogo Morskogo puti. T. 3. Sovetskoe arkticheskoe moreplavanie 1917–1932 гг. [The history of the discovery and development of the Northern Sea Route. Vol. 3. Soviet Arctic navigation 1917–1932]. Leningrad. 511 p. (In Russ.).
- Belov M.I. 1959b. Istoriya otkrytiya i osvoeniya Severnogo Morskogo puti. T. 4. Nauchnoe i khozyaystvennoe osvoenie Sovetskogo Severa. 1933–1945 gg. [The history of the discovery and development of the Northern Sea Route. Vol. 4. Scientific and economic development of the Soviet North. 1933–1945.]. Leningrad. 616 p. (In Russ.).
- Bijmoer R., Scherrenberg M., Creuwels J. 2023. Naturalis Biodiversity Center (NL) – Botany. Naturalis Biodiversity Center. Occurrence dataset <https://doi.org/10.15468/ib5ypt> accessed via GBIF.org on 2023-01-31.
- Catalogue of Life. 2023. <https://www.catalogueoflife.org/>
- Chapman A.D., Wiczorek J.R. 2020. Georeferencing best practices. Version 1.0. Copenhagen, Denmark, GBIF Secretariat. 112 p. <https://doi.org/10.15468/doc-gg7h-s853>
- Drinkwater R.E., Cubey R.W., Haston E.M. 2014. The use of Optical Character Recognition (OCR) in the digitisation of herbarium specimen labels. – *PhytoKeys*. 19 (38): 15–30.
- EtoMesto.ru. 2023 (<http://www.etomesto.ru/>)
- Glinka G.V., Tkhorzhevskiy I.I., Tsvetkov M.A., 1914. Atlas Aziatskoy Rossii. Izdanie pereselencheskogo upravleniya glavnogo upravleniya zemleustroystva i zemledeliya. [Atlas of Asian Russia. Publication of the resettlement department of the Main Department of Land Management and Agriculture]. St. Petersburg. 71 p. (In Russ.). <http://kartolog.ru/2009/09/atlas-aziatskoj-rossii-1914-g/?ysclid=ld1jlbtsj906089553>
- Global Biodiversity Information Facility. 2023. <https://www.gbif.org/>
- Güntsche A., Groom Q., Ernst M., Holetschek J., Plank A., Röpert D., Fichtmüller D., Shorthouse D.P., Hyam R., Dillen M., Trekels M., Haston E., Rainer H. 2021. A botanical demonstration of the potential of linking data using unique identifiers for people. – *PLoS ONE*. 16 (12): e0261130. <https://doi.org/10.1371/journal.pone.0261130>
- Le Bras G., Pignal M., Jeanson M.L., Muller S., Aupic C., Carré B., Flament G., Gaudeul M., Gonçalves C., Invernón V.R., Jabbour F., Lerat E., Lowry P.P., Offroy B., Pimparé E.P., Poncy O., Rouhan G., Haevermans T. 2017. The French Muséum national d’histoire naturelle vascular plant herbarium collection dataset. – *Scientific Data*. 4 (1): 1–16.
- Marcer A., Groom Q., Haston E., Uribe F. 2021. Natural History collections: georeferencing survey report: current georeferencing practices across institutions worldwide. – *Zenodo*. <https://doi.org/10.5281/zenodo.4644529>
- Marshruty.ru 2023. <https://maps.marshruty.ru/>
- Nelson G., Paul D., Riccardi G., Mast A.R. 2012. Five task clusters that enable efficient and effective digitization of biological collections. – *ZooKeys*. 209: 19–45. <http://dx.doi.org/10.3897/zookeys.209.3135>
- OOPT Rossii [SPNA of Russia]. 2023. <http://www.oopt.aari.ru/>
- Pomogator of the Moscow Digital Herbarium. 2023. <https://plant.depo.msu.ru/pomogator>
- Pospelov I.N. 2018. The experience of geographical fixing Central Siberia herbarium collections from the Moscow State University (MW) collection. – In: Abstracts of the conference “Information Technologies in the Research of Biodiversity” (BIT–2018). Irkutsk. P. 105–106.
- Pospelov I.N. 2020. Levels of spatial data geographical localization in information resources on biological diversity. – In: Information technology in biodiversity research: III National Scientific Conference with international participation, dedicated to the 100th anniversary of the birth of Russian academician Pavel Gorchakovskii. Conference abstracts. Ekaterinburg. P. 448–451 (In Russ.).
- Pospelov I.N., Pospelova E.B. 2019. About the possibilities of study of floristic diversity based on the analysis of herbarium fees (Digital herbarium MW). – In: Botanical and geographical research. Kamelin readings: a collection of scientific papers. Perm. P. 134–138 (In Russ.).
- Ramirez J., Watson K., McMillin L., Gjieli E., Sessa E. 2022. The New York Botanical Garden Herbarium (NY). Version 1.56. The New York Botanical Garden. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/6e8nje>
- Ranwashe F. 2022. Botanical Database of Southern Africa (BODATSA): Botanical Collections. Version 1.16. South African National Biodiversity Institute. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/2aki0q>
- Rodnaya Vyatka. 2023. <http://rodnaya-vyatka.ru/>
- Seregin A.P. 2017. The Moscow university Herbarium (MW) today: overview of collections, online access and contribution to science. – *Bot. Zhurn.* 102(3): 281–308 (In Russ.).
- Seregin A.P. 2018. The largest digital herbarium in Russia is now available online! – *Taxon*. 67 (2): 463–467. <https://doi.org/10.12705/672.34>

- Seregin A.P. 2020. Moscow Digital Herbarium: a consortium since 2019. — *Taxon*. 67 (2): 417–419. <https://doi.org/10.1002/TAX.12228>
- Seregin A.P. 2022. Instruktsiya po geoprivyazke obraztsov Gerbariya Moskovskogo universiteta (MW). Variant “Mongoliya”. Versiya 1.0. [Instruction for geo-linking samples of the Herbarium of Moscow University (MW). Variant “Mongolia”. Version 1.0]. Moscow. (In Russ.). <https://doi.org/10.13140/RG.2.2.14879.84645>
- Seregin A. 2023a. Moscow University Herbarium (MW). Version 1.265. Lomonosov Moscow State University. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/cpnhcc>
- Seregin A.P. (Ed.). 2023b. Moscow Digital Herbarium: Electronic resource. Moscow State University, Moscow. <https://plant.depo.msu.ru/>
- Thiers B.M. 2023. Index Herbariorum. <https://sweetgum.nybg.org/science/ih/>
- Tupitsyna N.N., Shaulo D.N., Gureeva I.I. 2016. Obzor floristicheskikh issledovaniy Sredney Sibiri. [Overview of floristic studies of Central Siberia]. Krasnoyarsk. 253 p. (In Russ.). <https://doi.org/10.17223/9785946216371/33> (<https://vital.lib.tsu.ru/vital/access/services/Download/vtls:000617746/SOURCE1>)
- Teisher J., Stimmel H. 2022. Tropicos Specimen Data. Missouri Botanical Garden. Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/hja69f>
- Witzell H., Shah M. 2022. Lund Botanical Museum (LD). Occurrence dataset accessed via GBIF.org on 2023-01-31. <https://doi.org/10.15468/c4w4co>
- Yandex. Maps. 2023. <https://yandex.ru/maps/>