



грант РНФ № 14-50-00029

# Цифровой гербарий МГУ – крупнейшая российская база данных о биоразнообразии

А.П. Серегин, в.н.с. Гербария МГУ

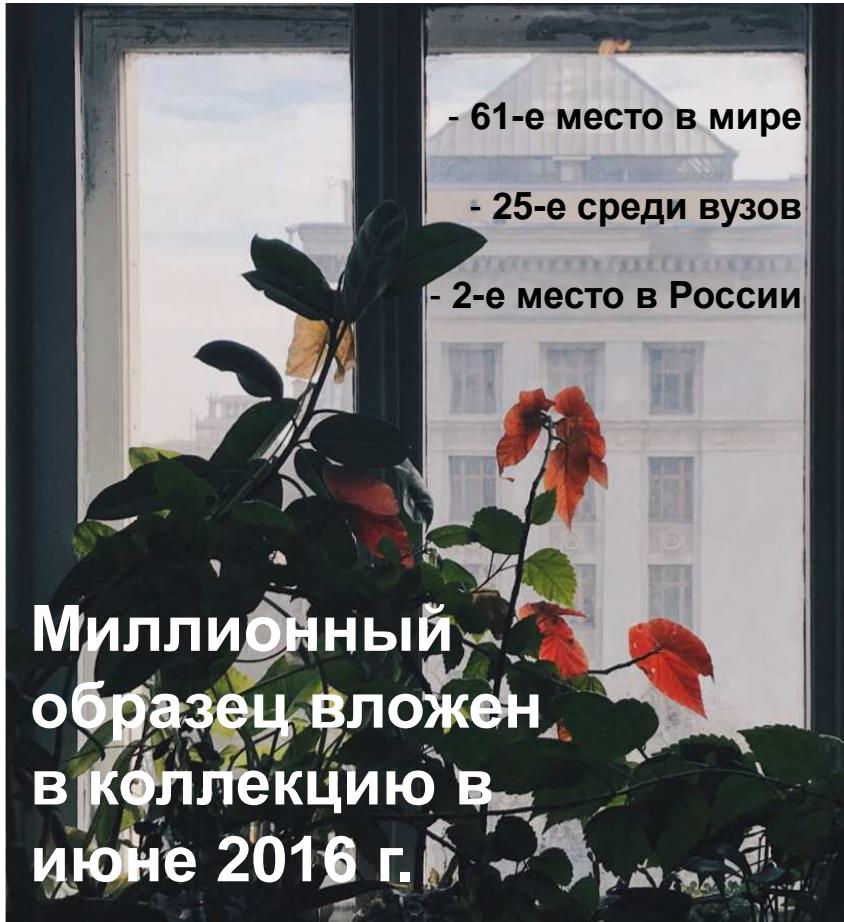
Generated 9 hours ago © OpenStreetMap contributors, © OpenMapTiles, GBIF

Any year

1739-2017



# Фонды Гербария МГУ



**Финальная статистика за 2018 г.**

Восточная Европа	368 617	Типовые образцы	4 676
Сибирь и Дальний Восток	165 476	Herbarium Alchemillarum	3 952
Кавказ	100 668	Ботанические сады	3 164
Средняя Азия	97 141	Австралия и Океания	3 337
Зарубежная Европа	40 481	Антарктика	34
Крым	32 748	Мохообразные	79 111
Монголия	27 341	Прочие коллекции	75 229
Зарубежная Азия	24 681	<b>итого</b>	<b>1 044 751</b>
Америка	12 236		
Африка	9 057		



# Цифровой гербарий МГУ

- **Цифровой гербарий МГУ** появился в результате работ по сканированию фондов гербария Московского университета.
- **Источник финансирования:** грант РНФ «Научные основы создания национального банка-депозитария живых систем» (№ 14-50-00029).
- **Участие** на разных этапах работы принимали свыше 100 человек.



# Подготовка к сканированию: 2015 год

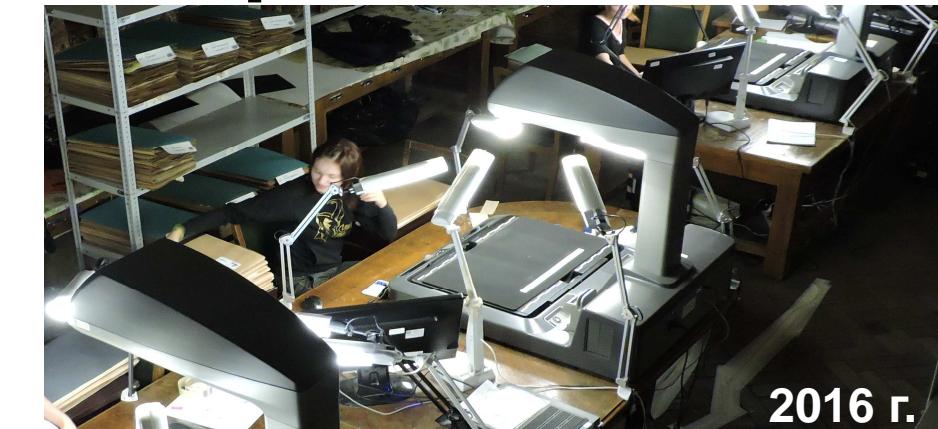


В процессе подготовки гербария к сканированию участвовало 63 человека.

30,5% работ провели пятеро штатных сотрудников гербария.



# Сканирование фондов



# Параметры изображений



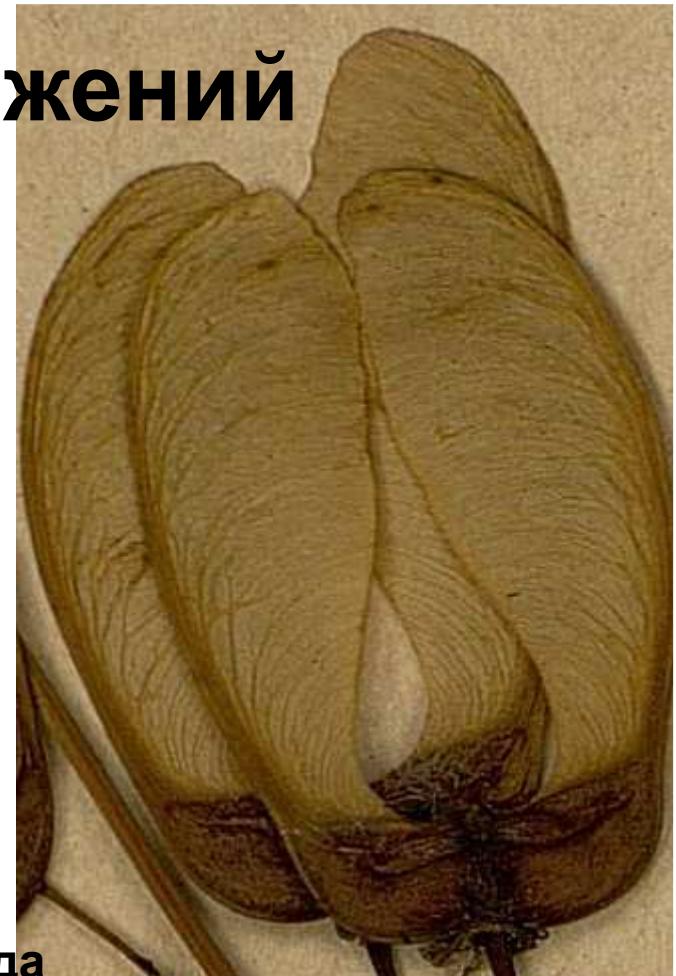
Основное:  
**TIFF 300 dpi**

вес файла: 50-60 МВ  
массив: 27,5 ТВ

Копия:  
**JPG 300 dpi 80%**

вес файла: 2-2,5 МВ  
массив: 1-1,5 ТВ

Масштабная линейка – да  
Цветовая мишень – нет



# Главная страница портала

<http://plant.depo.msu.ru/>

 ДЕПОЗИТАРИЙ  
живых систем  
«НОЕВ КОВЧЕГ»

Микроорганизмы и грибы    Растения    Животные    Биоматериалы человека    Био. информация

RU EN    Вход в систему

О системе    Коллекции    Контакты    Ссылки    Инфраструктура    Цитировать

Сейчас в базе данных (гербарий, образцы ДНК, фотографии растений в природе):  
Образцов: 974285    Изображений: 968031    Видов: 37782    Геопривязок: 361160    Этикеток + OCR: 153706 + 320340

**Национальный банк-депозитарий живых систем**

**Цифровой гербарий МГУ**

Проект Московского университета "Ноев ковчег" посвящен созданию многофункционального сетевого хранилища биологического материала.

Планируется работа с материалом всех возможных типов - от отдельных биологических молекул до целых живых организмов.

Создание депозитария позволит сохранить биоразнообразие нашей планеты и создать новые способы полезного использования биологического материала.

**Атлас флоры России**

Хочешь принять участие в создании "Атласа флоры России"? Загружай свои фотографии растений в природе и точку съемки на [iNaturalist](#), где они станут частью нашего нового проекта "Флора России | Flora of Russia".

  
0 1 2 3 4 5 6 7 8 9 10  
cm

All rights reserved

 Herbarium  
Московского  
Университета  
Herbarium  
Universitatis  
Moscuensis (MW)

# Четыре точки доступа

- <https://plant.depo.msu.ru/> (операционная версия)
- <https://plant.depo.msu.ru/open/> (открытая версия,  
оптимизированная для мобильных устройств)
- <https://www.gbif.org/> (еженедельная выгрузка)
- <https://yandex.ru/images/> (индексация картинок роботами  
Яндекса)





# Отсканировали. Что дальше?



**ДЕПОЗИТАРИЙ**  
живых систем  
«НОЕВ КОВЧЕГ»

Микроорганизмы и грибы
Растения
Животные
Биоматериалы человека
Био. информация

RU EN

[Вход в систему](#)

---

[О системе](#)
[Коллекции](#)
[Контакты](#)
[Ссылки](#)
[Инфраструктура](#)
[Цитировать](#)

Сейчас в базе данных (гербарий, образцы ДНК, фотографии растений в природе).

Образцов: [974285](#) Изображений: [968031](#) Видов: [37785](#) Геопривязок: [361160](#) Этикеток + OCR: [153706](#) + [20340](#)

**Поиск**

Все поля

Выше рода

Род / Вид (лат., рус.)

[Расширенный поиск](#)

[Поиск по русским названиям](#)

[Поиск по этикеткам и OCR](#)

[Таксономический браузер](#)

## Национальный банк-депозитарий живых систем

### Цифровой гербарий МГУ

Проект Московского университета "Ноев ковчег" посвящен созданию многофункционального сетевого хранилища биологического материала.

Планируется работа с материалом всех возможных типов - от отдельных биологических молекул до целых живых организмов.

Создание депозитария позволит сохранить биоразнообразие нашей планеты и создать новые способы полезного использования биологического материала.

### Атлас флоры России

Хочешь принять участие в создании "Атласа флоры России"? Загружай свои фотографии растений в природе и точку съемки на [iNaturalist](#), где они станут частью нашего нового проекта "Флора России | Flora of Russia".

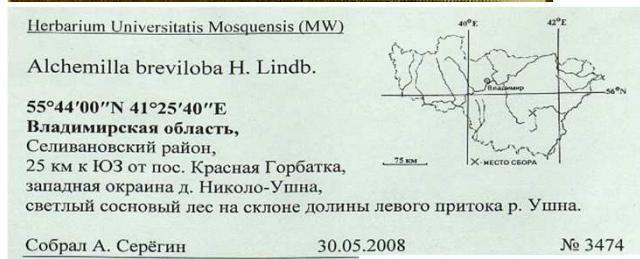
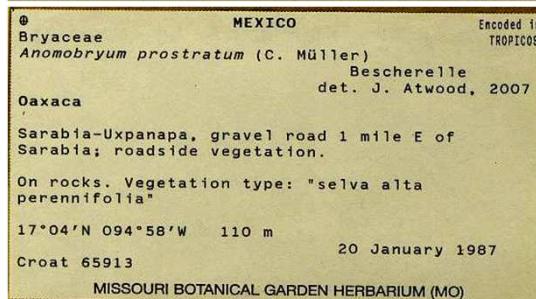


All rights reserved

Herbarium  
Universitatis  
Moskovensis (MW)  
Гербарий имени Д.П. Сырнинцева

# Источники данных БД этикеток

- Ручной ввод
- Открытые источники
- Файлы с текстом



## Источники данных

## • Ручной ввод

# Источники данных

- Открытые источники

The screenshot shows a web browser displaying the Tropicos.org website. The URL in the address bar is tropicos.org/Specimen/128308. The page title is "Tropicos®". The main content is about a specimen collection by Elsa M. Zardini, number 29733. The "Details" tab is selected. Key information includes:

- Current Determination:** *Trichilia catigua* A. Juss. Family: Meliaceae Determined By: T. Pennington (K), 2005
- Previous Determinations:** !*Trichilia* P. Browne Family: Meliaceae Determined By:
- Collection Information:**
  - Collectors: Elsa M. Zardini & P. Aquino
  - Collector Team: Elsa M. Zardini  
P. Aquino
  - Collection Number: 29733
  - Collection Date: 13 January 1992
  - Location: Paraguay, Paraguarí
  - Keywords: Acahay Massif
  - Coordinate: 25°52'S 057°08'W (-25.8666600, -57.1333300)
  - Elevation:
  - Herbaria: MO, PY
  - Description: Tree 12 m. Corolla yellow.
  - Locality: Acahay Massif. Easternmost Peak. Northern part. Scrub on rocks.
  - Footer: MISSOURI BOTANICAL GARDEN (MO)  
INVENTARIO BIOLOGICO NACIONAL (PY)
- Barcodes/Accession Numbers:**

Institution Barcode	Accession Object Kind	Object Subkind
MO	MO-1325239	04834639 sheet

At the bottom right, there is a scale bar from 0 to 10 cm and a logo for Herbarium Universitatis Mosquensis (MW) with the text "Гербарий имени Д.П. Сиродюкова".

# Источники данных

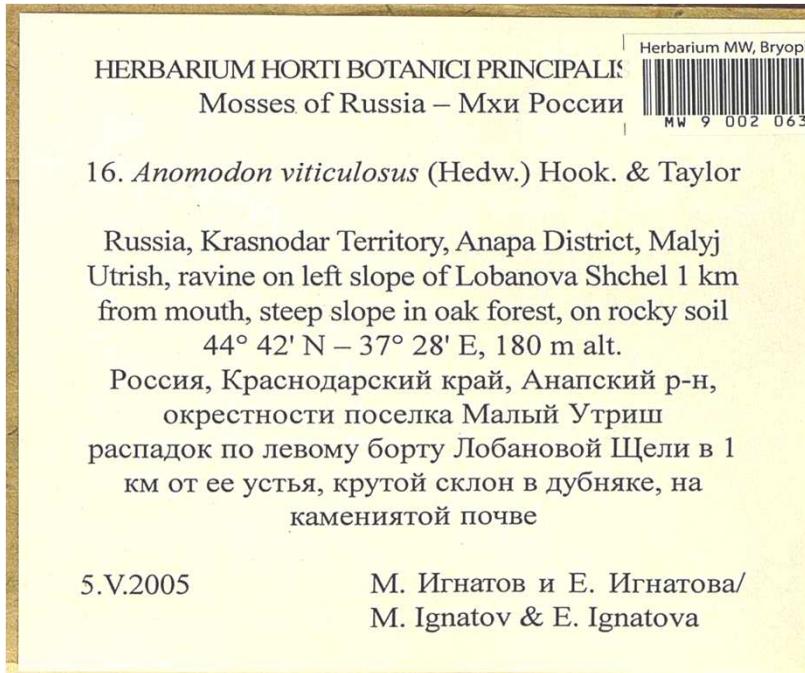
- Файлы с текстом этикеток

<p><u>Herbarium Universitatis Mosquensis (MW)</u></p> <p><i>Eleocharis acicularis</i> (L.) Roem. et Schult.</p> <p>55°57'50"N 42°06'40"E Владимирская область, Вязниковский район, 26 км к Ю от г. Вязники, 2 км к Ю от с. Сергиевы Горки, правый берег водохранилища на рч. Сураж, илистое мелководье и берег.</p> <p>Собрал А. Серёгин 3.07.2009 № 3885 Сеточное картирование флоры Владимирской области: Л21 Atlas Florae Europeae: 38VLH2-10</p> <p><u>Herbarium Universitatis Mosquensis (MW)</u></p> <p><i>Potamogeton berchtoldii</i> Fieb.</p> <p>55°58'20"N 42°06'20"E Владимирская область, Вязниковский район, 25 км к Ю от г. Вязники, 1 км к Ю от с. Сергиевы Горки, правый берег водохранилища на рч. Сураж у дамбы, плавало у берега.</p> <p>Собрал А. Серёгин 3.07.2009 № 3886 Сеточное картирование флоры Владимирской области: Л21 Atlas Florae Europeae: 38VLH2-10</p>	<p><u>Herbarium Universitatis Mosquensis (MW)</u></p> <p><i>Beckmania eruciformis</i> (L.) Host</p> <p>55°58'20"N 42°06'20"E Владимирская область, Вязниковский район, 25 км к Ю от г. Вязники, 1 км к Ю от с. Сергиевы Горки, кювет вдоль дороги от г. Вязники до шоссе Р76.</p> <p>Собрал А. Серёгин 3.07.2009 № 3887 Сеточное картирование флоры Владимирской области: Л21 Atlas Florae Europeae: 38VLH2-10</p> <p><u>Herbarium Universitatis Mosquensis (MW)</u></p> <p><i>Alisma plantago-aquatica</i> L.</p> <p>55°58'20"N 42°06'20"E Владимирская область, Вязниковский район, 25 км к Ю от г. Вязники, 1 км к Ю от с. Сергиевы Горки, кювет вдоль дороги от г. Вязники до шоссе Р76.</p> <p>Собрал А. Серёгин 3.07.2009 № 3888 Сеточное картирование флоры Владимирской области: Л21 Atlas Florae Europeae: 38VLH2-10</p>
---	---

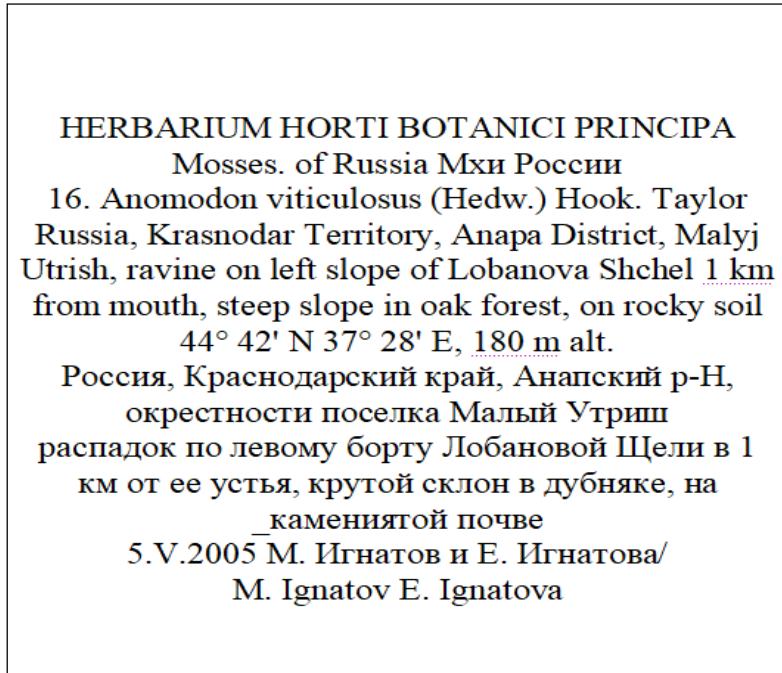


# OCR как источник текстовых данных

Tesseract  
OCR



Исходная этикетка



Результат OCR

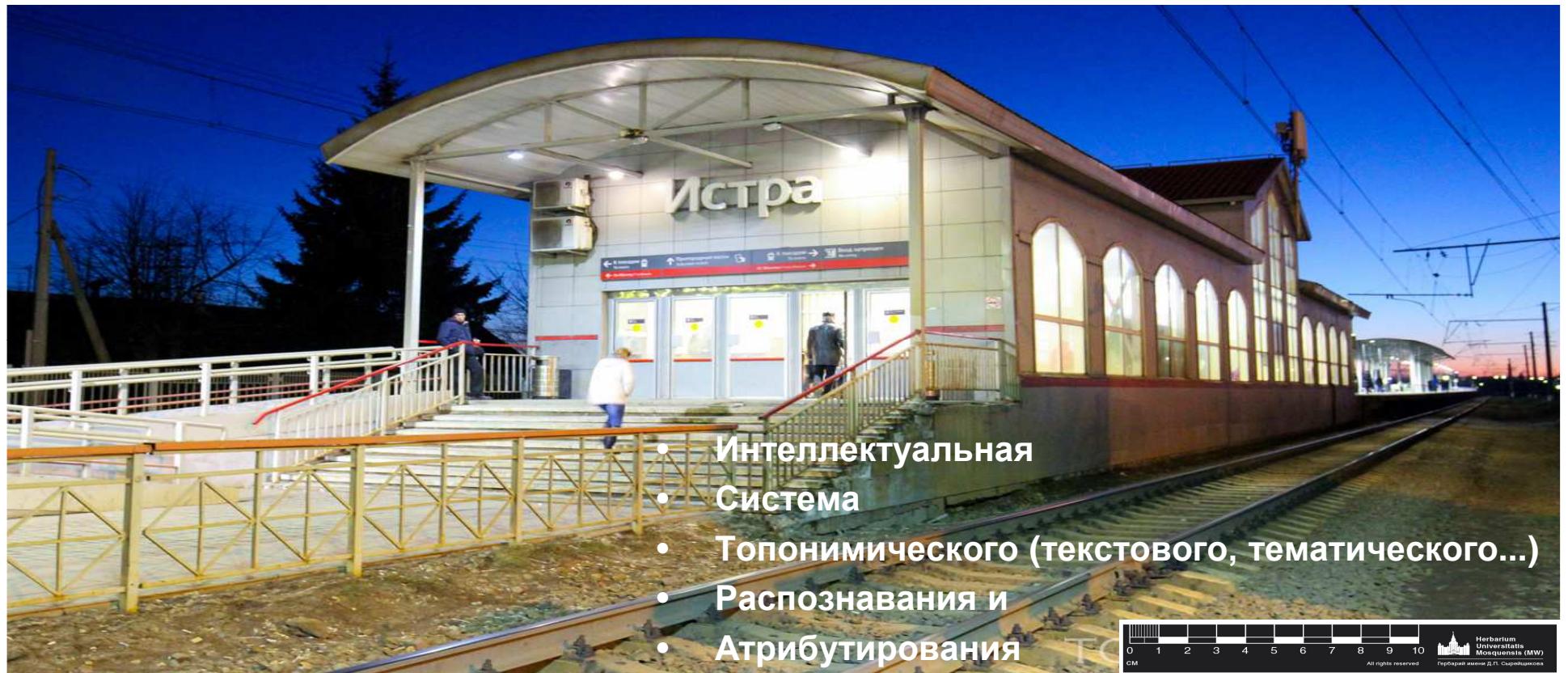


# Источники для БД геопривязок

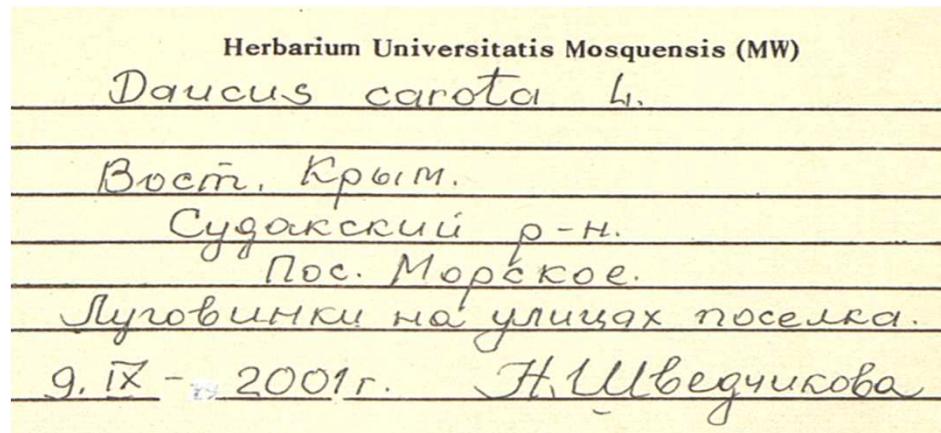
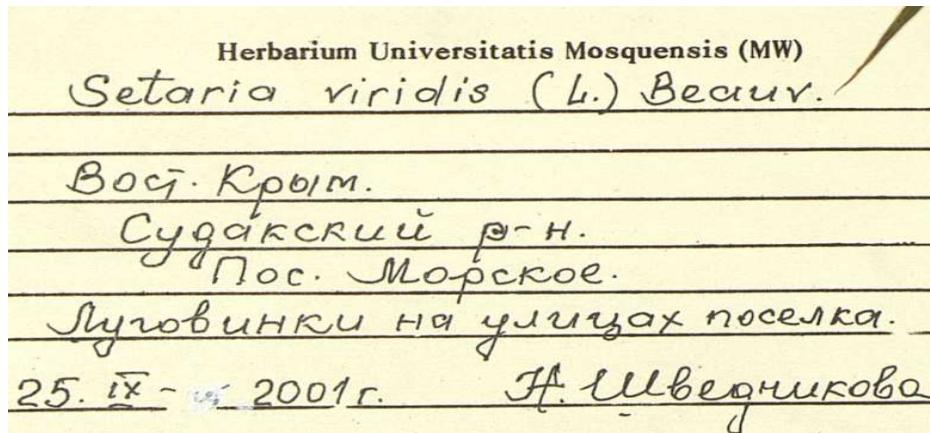
- Координаты непосредственно с гербарных этикеток.
- Привлечение знатоков местности для геопривязки.
- Привлечение волонтеров для поиска «простых» мест по online ресурсам.
- Разработка алгоритма ИСТРА для автоматической привязки образцов к карте.



# Что такое ИСТРА?



# Как работает ИСТРА?

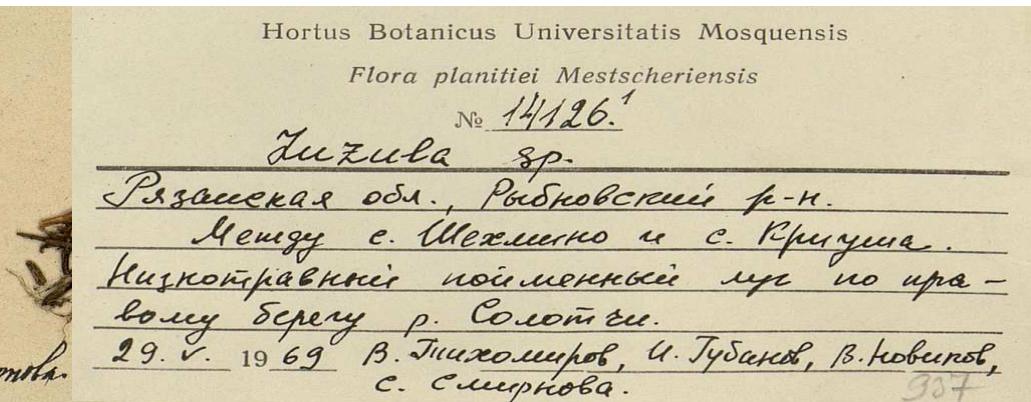
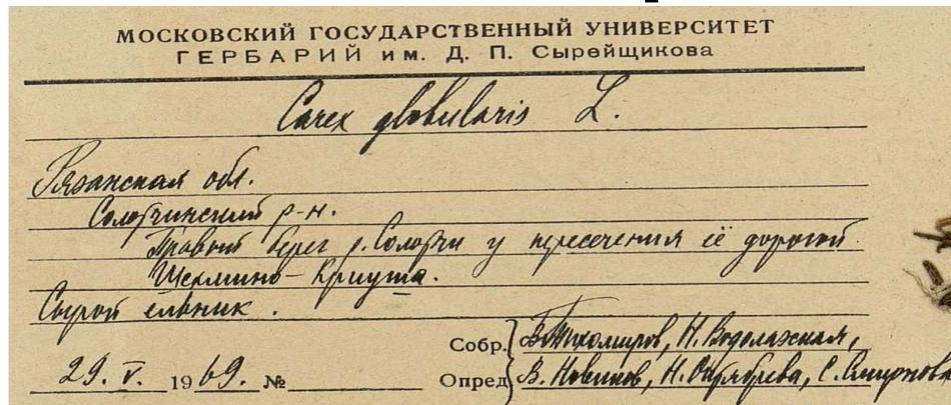


## Алгоритм 1

Много образцов → уникальное место сбора  
(по совпадению поля «Текст этикетки»)



# Как работает ИСТРА?

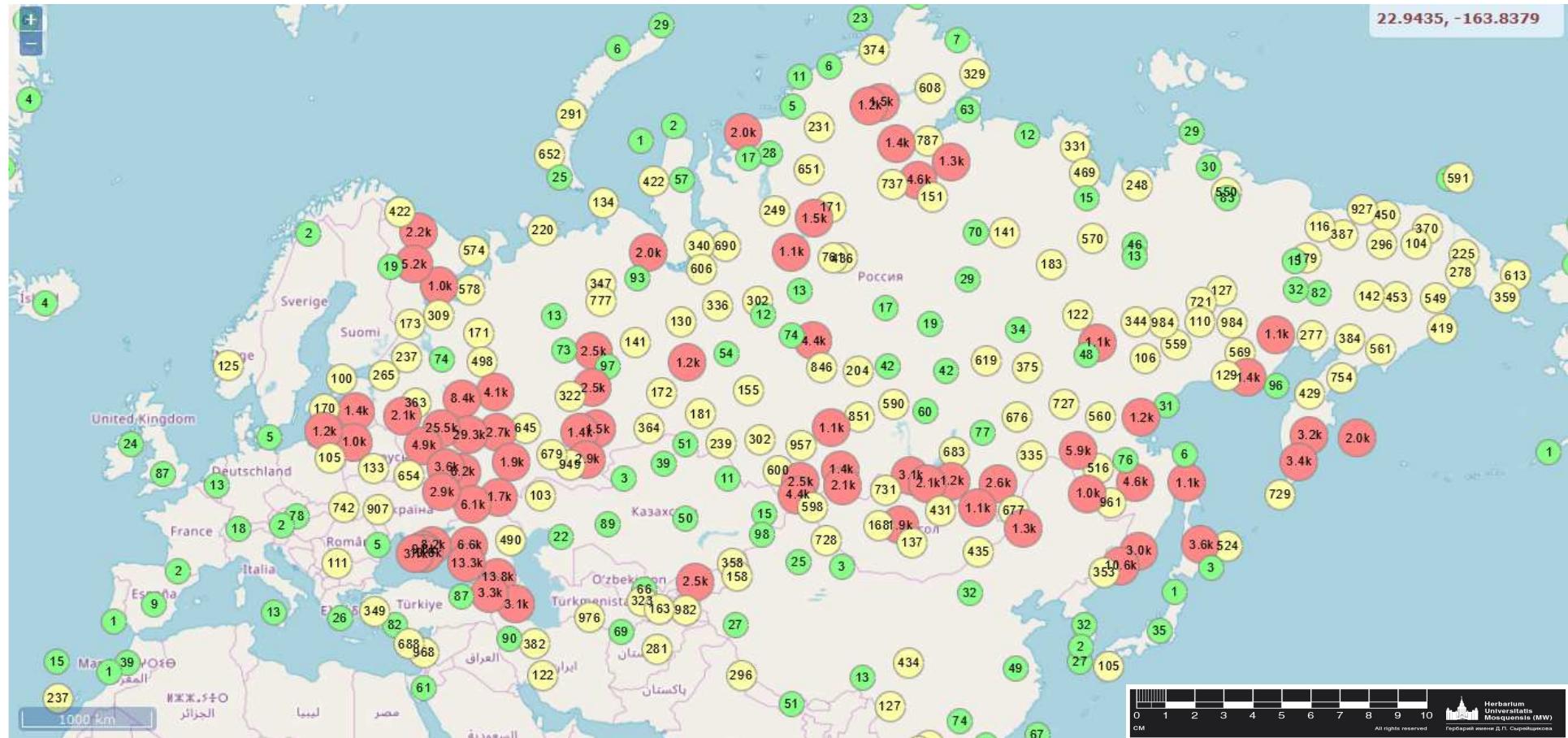


## Алгоритм 2

Один человек в один день  
находится примерно в одном месте  
(по совпадению полей «Дата сбора» и «Коллектор»)



# 361 000 геопривязок!



# Уровень оцифровки на 25 марта 2019 г.

- 968 000 образцов отсканировано (93%)
- 774 930 образцов имеют дату сбора в БД (74%)
- 361 160 образцов имеют геопривязку (35%)
- 153 706 образцов с этикетками в БД (15%)  
+ 320 340 образцов (без этикеток в БД) имеют OCR

**1 046 000 образец хранится в Гербарии МГУ**

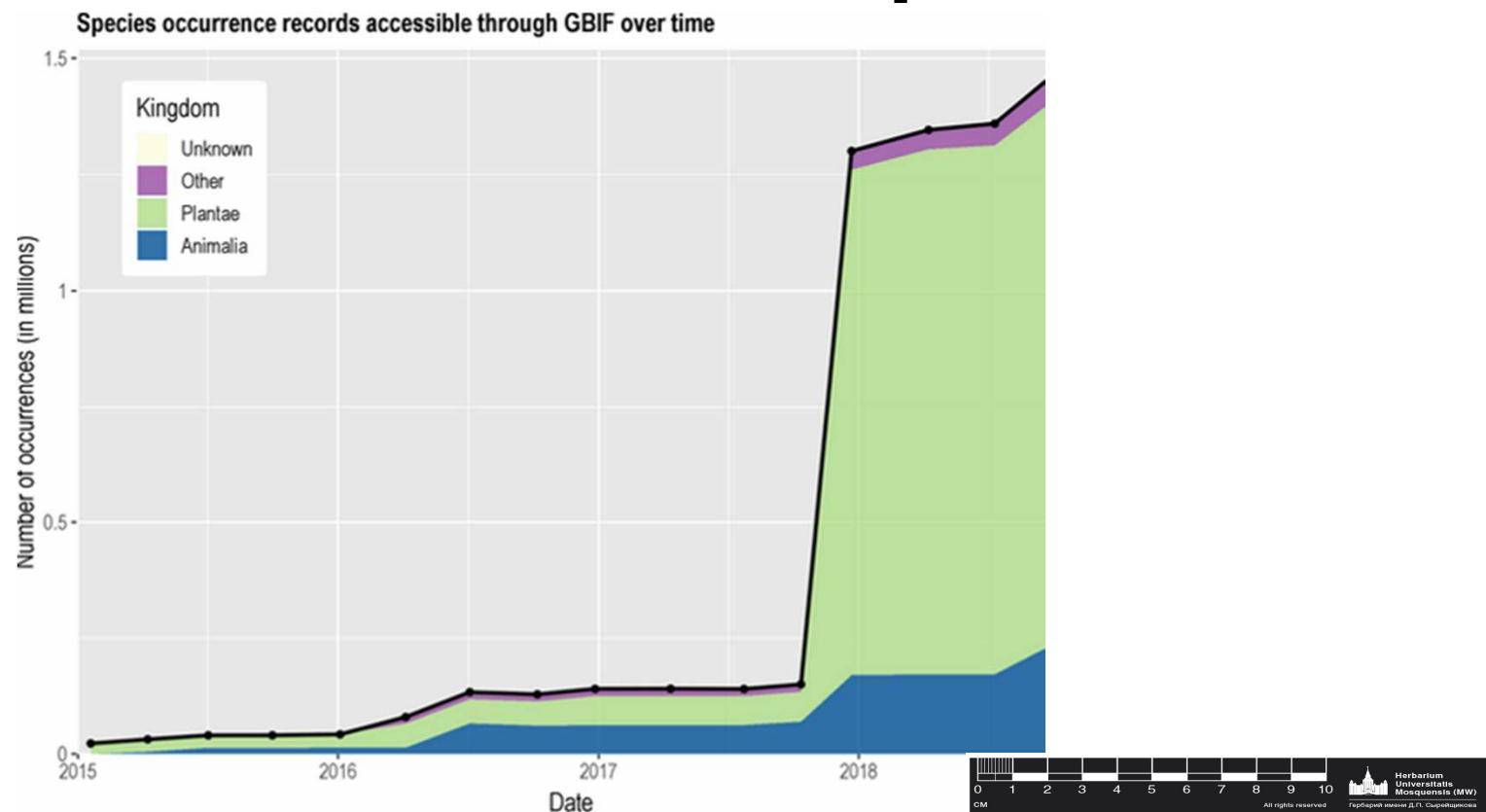


# Три факта о Цифровом гербарии МГУ

1. Восьмой цифровой гербариев мира по числу сканов – 968 031 изображений (на 25.03.2019).
2. Самая большая база данных в России по биоразнообразию.
3. Крупнейший GBIF-донор данных: как среди российских учреждений (59%), так и для территории России (25%).

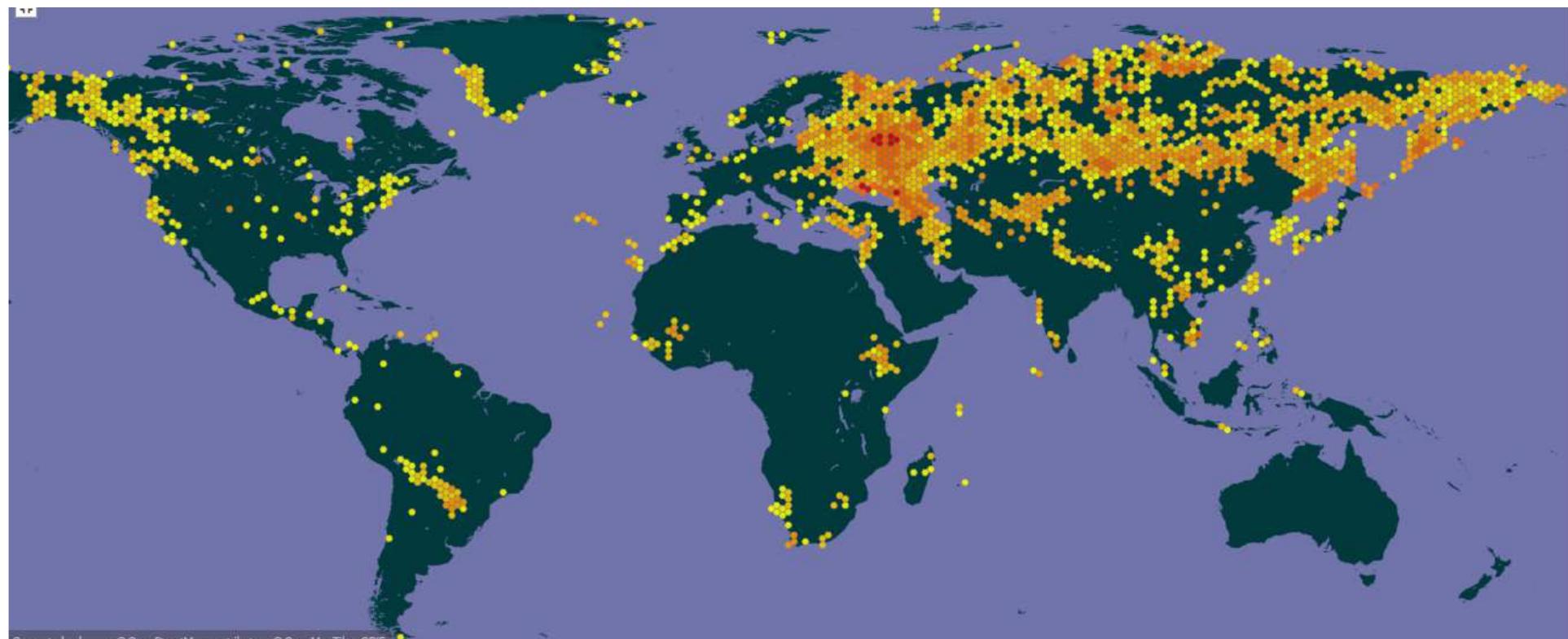


# Участие России в GBIF: большой скачок 2 ноября 2017 г.





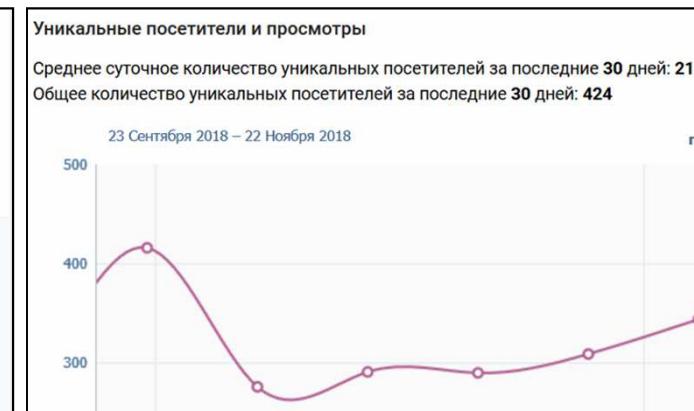
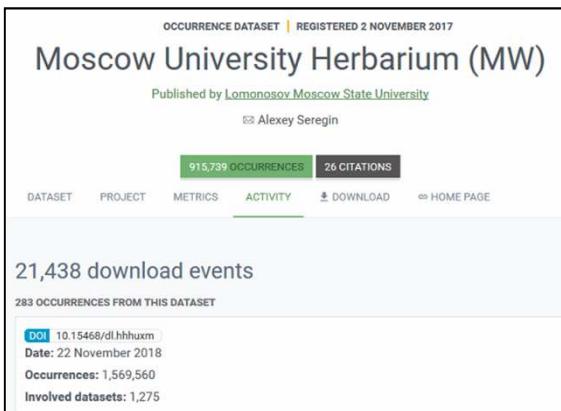
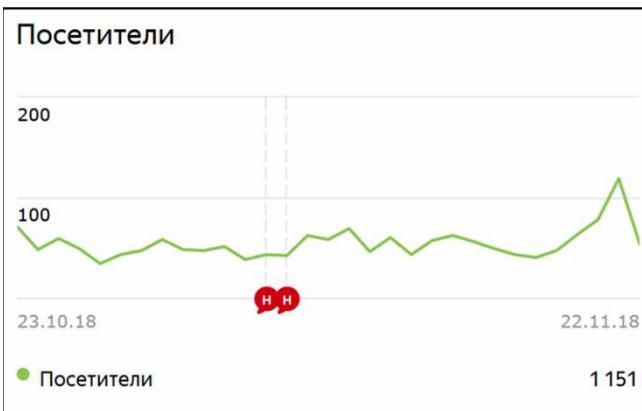
# Moscow University Herbarium (MW) – 13-е место (360 915 точек)



# Востребован ли Цифровой гербарий МГУ?



Российский  
научный  
фонд



1151 посетителей в месяц  
(Яндекс.Метрика для  
[plant.depo.msu.ru](http://plant.depo.msu.ru))

50-60 человек в день

1671 выгрузка в месяц  
(GBIF-выгрузки данных  
Цифрового гербария МГУ)

пересчет 21438 выгрузок за 385  
дней

424 посетителя в месяц  
(Информационный блог  
Цифрового гербария МГУ на  
[vk.com](https://vk.com))

Но: 5879 просмотров новостей через  
ленту в октябре



# Цитирования: Google Scholar

# Цитирования: GBIF

OCCURRENCE DATASET | REGISTERED NOVEMBER 2, 2017

Moscow University Herbarium (MW)

Published by Lomonosov Moscow State University

✉ Alexey Seregin

DATASET PROJECT METRICS ACTIVITY DOWNLOAD HOME PAGE

970,446 OCCURRENCES 34 CITATIONS

BECAUSE YOU ARE TRUSTED CONTACT

Last ingestion a success

HISTORY INGEST NOW LOGS

**M**  
Project ID: 14-50-00029  
Metadata last modified: December 21, 2018  
Data last changed: December 21, 2018  
Hosted by: Lomonosov Moscow State University  
License: CC BY 4.0  
How to cite DOI: 10.15468/cpnihcc

970,446 https://www.gbif.org/dataset/902c8fe7-8f38-45b0-854e-c324fed36303 99.9% With taxon match

31% With coordinates Topic

Dataset Publisher Peer-reviewed Open access

79% With year

Aichi biodiversity targets • Biodiversity indicators • Crop wild relatives • Ex situ conservation • In situ conservation • Sustainable development goals  
Journal article | Open access | Peer-reviewed  
Data used in study DOI: 10.15468/dl.v32kw2

Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Euras... Literature  
Frajman, B., Záveská, E., Gamisch, A., Moser, T. Sc  
Next generation sequencing has revolutionised t

Herbarium Moscow State University (MW)  
Городской институт Д. Г. Соколова

# Наша команда

## Авторы системы

- В.В. Перехватов – руководитель группы
- О.Н. Платко – архитектор системы, разработчик системы
- К.В. Скулачев – разработчик системы
- А.Б. Клюквин – разработчик системы
- А.Е. Соловченко – генеральный заказчик системы
- А.П. Серегин – автор технического задания системы

## Авторы ГИС-модуля

- Т.Е. Самсонов – руководитель группы, разработчик системы
- В.Н. Семин – разработчик системы
- А.Р. Аляутдинов – разработчик системы

## Авторы ключевых метаданных библиотеки изображений

- А.П. Серегин – редактор данных, автор ключевых метаданных
- Н.С. Гамова – автор словаря русских названий, автор ключевых метаданных
- К.В. Дудова – автор ключевых метаданных
- А.Д. Шклярук – автор ключевых метаданных
- А.С. Борцова – автор ключевых метаданных
- А.В. Леонова – автор ключевых метаданных

## Авторы базы данных этикеток

(в порядке убывания числа введенных этикеток, не менее 500 штук)

- сотрудники НПО "Опыт" (корпорация ЭЛАР)
- И.В. Мельник
- А.П. Серегин
- И.Н. Поспелов
- П.Г. Ефимов
- В.Э. Федосов
- Д.В. Бородина
- А.А. Лапин
- В.В. Бурый
- Н.С. Гамова
- С.В. Дудов
- К.В. Дудова
- Т.М. Гаврилова
- М.А. Малашева
- М.С. Нуралиев
- М.О. Иванова
- А.Н. Демидова

## Авторы базы данных геопривязок

(в порядке убывания числа введенных ручных геопривязок, не менее 200 штук)

- И.Н. Поспелов
- Н.В. Иванова
- А.С. Салмин
- А.П. Серегин
- А.А. Лапин
- А.Г. Кузьмин
- И.В. Мельник
- Т.М. Гаврилова
- Е.В. Селезнева
- С.В. Дудов
- К.В. Дудова
- Н.С. Гамова
- М.О. Иванова
- Е.С. Каширина
- М.П. Шашков



# Гербарий МГУ: центр притяжения

Полностью переданы в наши фонды:

- Гербарий лаборатории устойчивости лесных экосистем Костромского университета (2016 г.)
- Гербарий Бадхызского заповедника (2017 г.)
- Гербарий Института почвоведения в Пущино (2018 г.)

Очевидно, что этот процесс продолжится!





22 013 образцов в год (2016 г.)  
19 416 образцов в год (2017 г.)



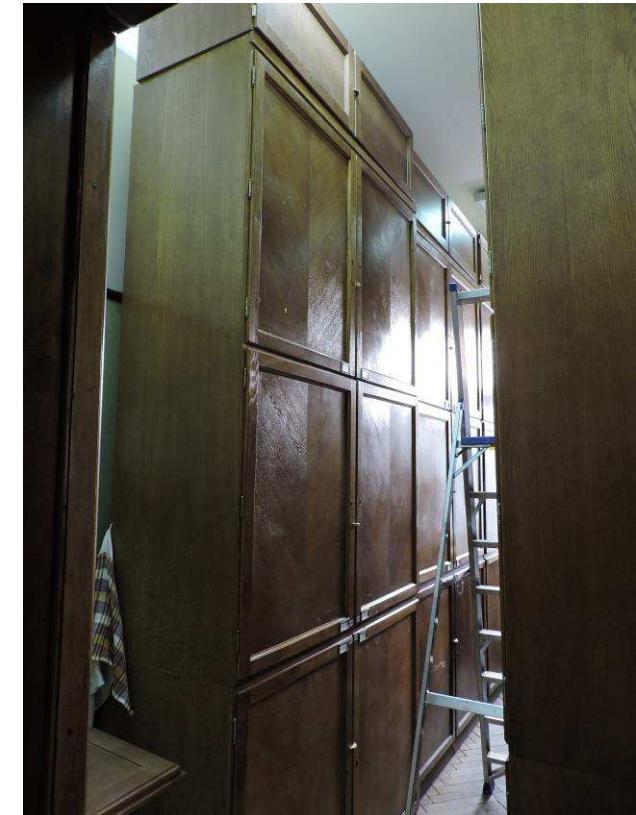
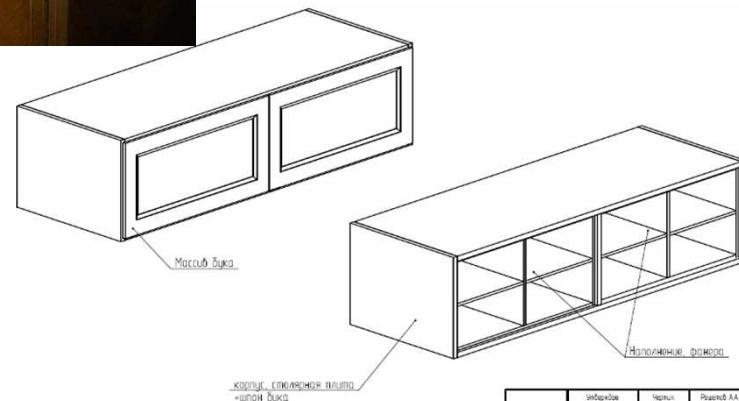
# Хранилище не резиновое



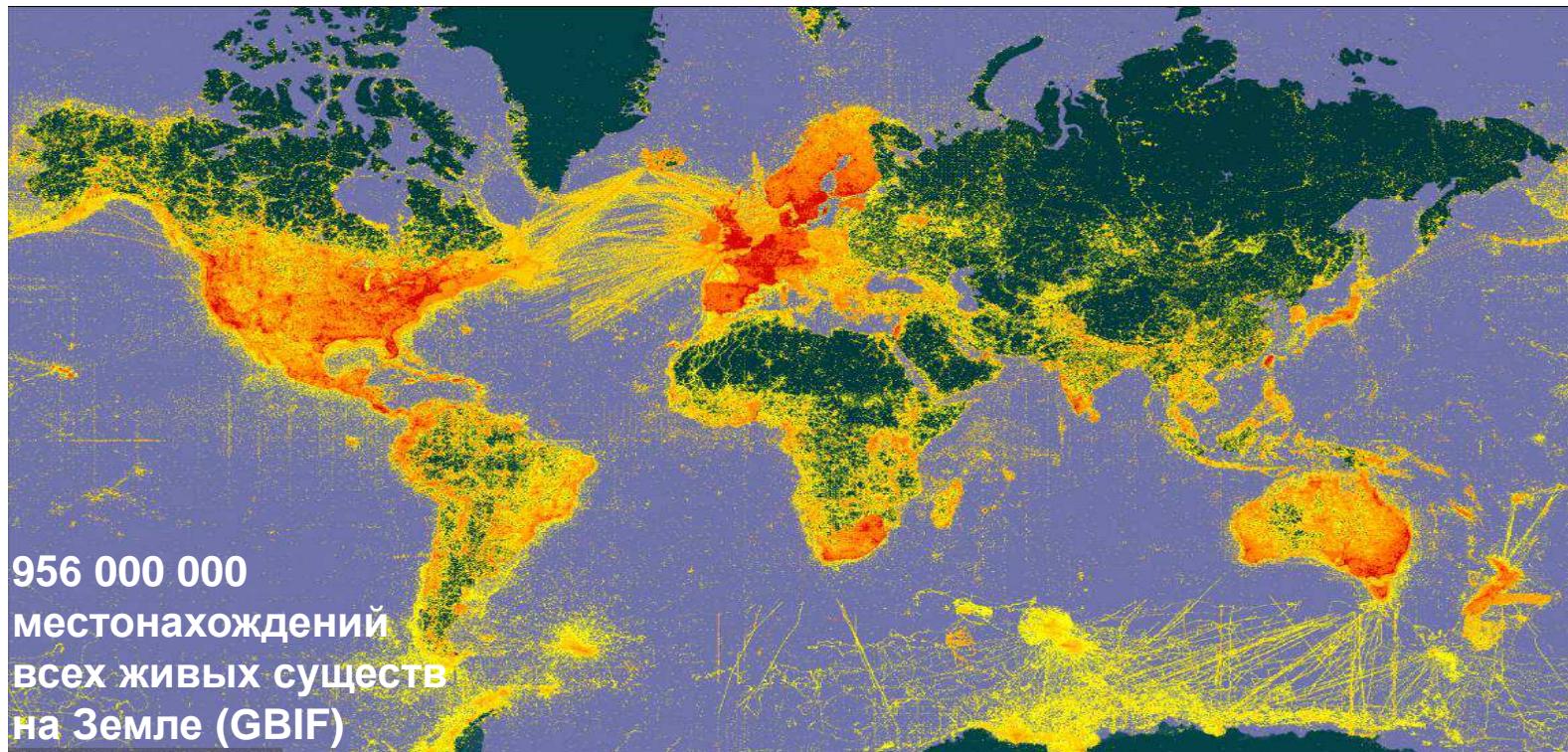
# Рост коллекций: пока вверх



В марте 2017 г.  
по Программе развития  
МГУ ёмкость  
хранилища  
выросла на 15%.  
Этого хватит ненадолго



# Нам предстоит большая работа!



## Заключение

