ДЕПОЗИТАРИЙ ЖИВЫХ СИСТЕМ «НОЕВ КОВЧЕГ»

Moscow Digital Herbarium: A Global Approach Through The Regional Actions

Dr Sci Alexey P. Seregin
Lomonosov Moscow State University

Congresso S.B.I. (IPSC) - Padova, 4-7 September 2019

Moscow University Herbarium (MW)

- Founded ca. 1780
- The second largest herbarium in Russia
- Six staff members (science & technical)
- Part of the *National Depository Bank of Live Systems* initiative (launched by the Moscow University)
- No international loans due to custom regulations

Specimen Total

- **1,047,009** specimens (as of Nov 2018) including:

- 955,640 vascular plants
- 81,369 bryophytes
- ca. 10,000 lichens

Stats update is available once a year.

Specimen Ranking

***Stipa* specimens from Eastern Europe**

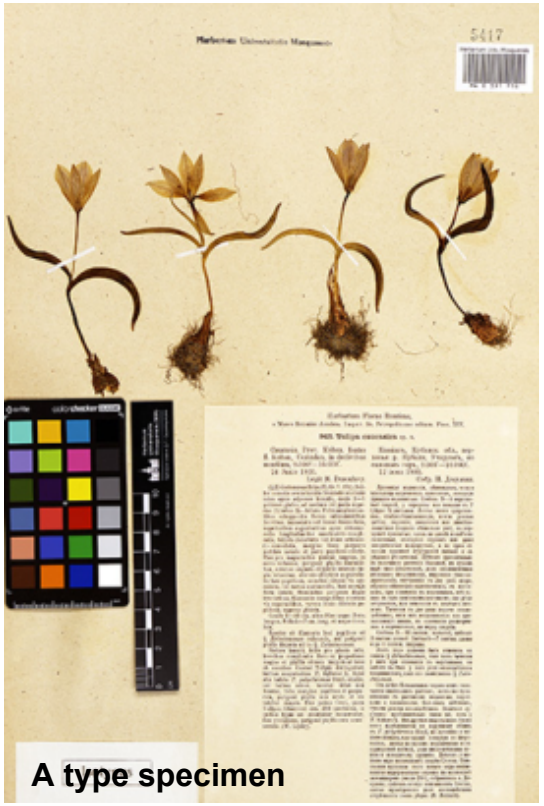
- 63rd in the world
- 24th among university herbaria

- 2nd in Russia (after Komarov Institute)

Taxonomic Outline

- 37,857 species (excl. historic collections) including:
- 35,634 vascular plants
- 2,223 bryophytes

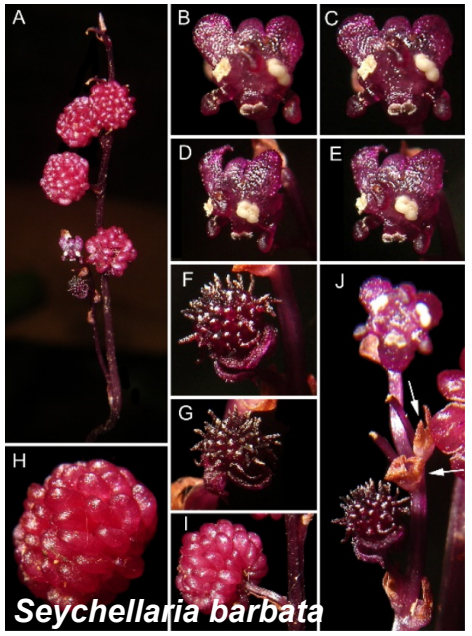
Types of ca. 3,000 taxa are curated separately (4,821 specimens).



Plant Taxonomy in the Moscow University

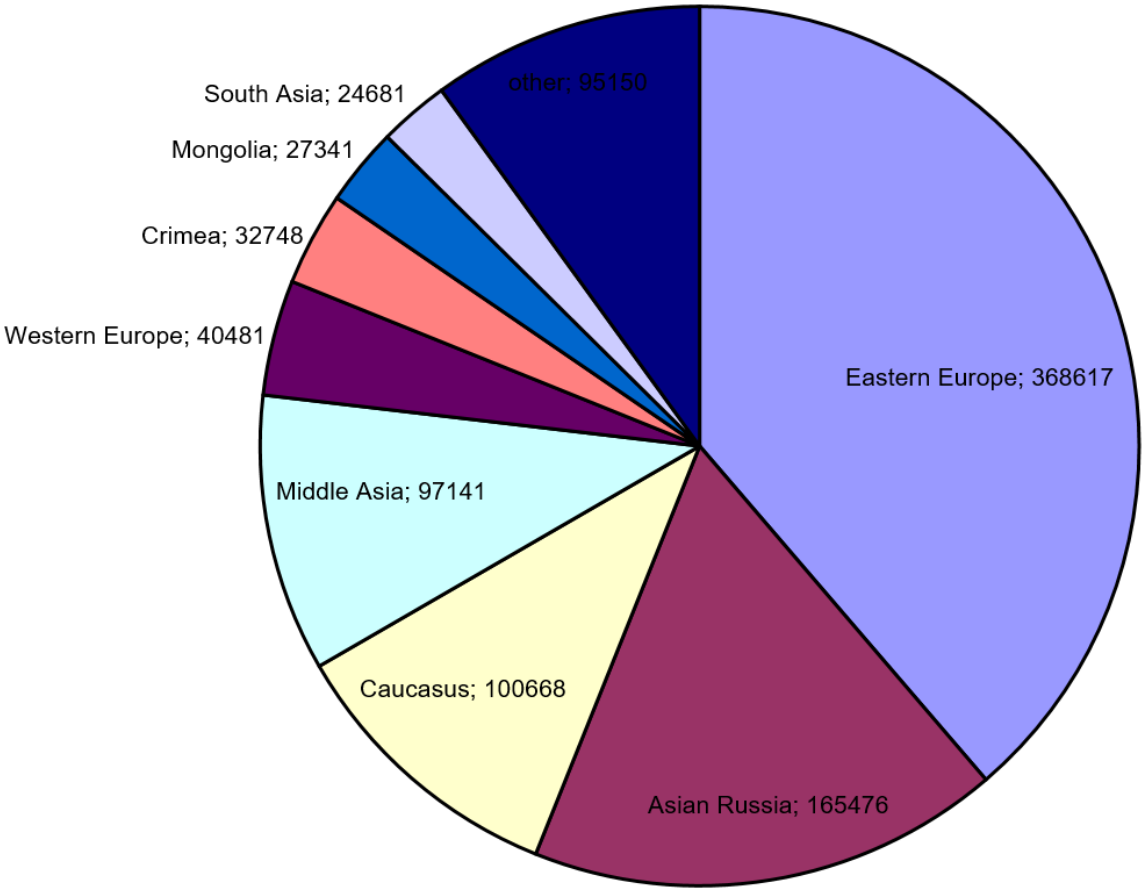
60 new species of vascular plants were described by the Moscow University staff members in 2012-2016:

- **Apiaceae**: 15 species (M. Pimenov, E. Kljuykov, D. Lyskov et al.)
- **Amaranthaceae s. l.**: 11 species (A. Sukhorukov)
- **Asparagaceae**: 10 species (N. Vislobokov et al.)
- **Amaryllidaceae s. l.**: 9 species (A. Seregin)
- **Apocynaceae**: 2 species (Y. Alexeev)
- **Cucurbitaceae**: 2 species (M. Nuraliev)
- **Polygonaceae**: 2 species (O. Yurtseva)
- **Thismiaceae**: 2 species (M. Nuraliev)
- **Molluginaceae**: 2 species (A. Sukhorukov)
- **Asteraceae, Centrolepidaceae, Lophiocarpaceae, Orchidaceae, Triuridaceae**: 1 species



Seychellaria barbata

Geographical Scope: Vascular Plants



Moscow University Project

- Moscow University have received budget for foundation of the ***National Depository Bank of Live Systems***.

- 4-year project (2015-2018).

- Five branches: 1) Plants; 2) Animals; 3) Fungi & Microbes; 4) Human Material; 5) Bioinformatics.

- Nine labs studying plants, including my lab (Moscow University Herbarium).

Digitisation as a World Trend

Leiden (Netherlands)

Paris (France)

New York (USA)

Equipment or Services?

or

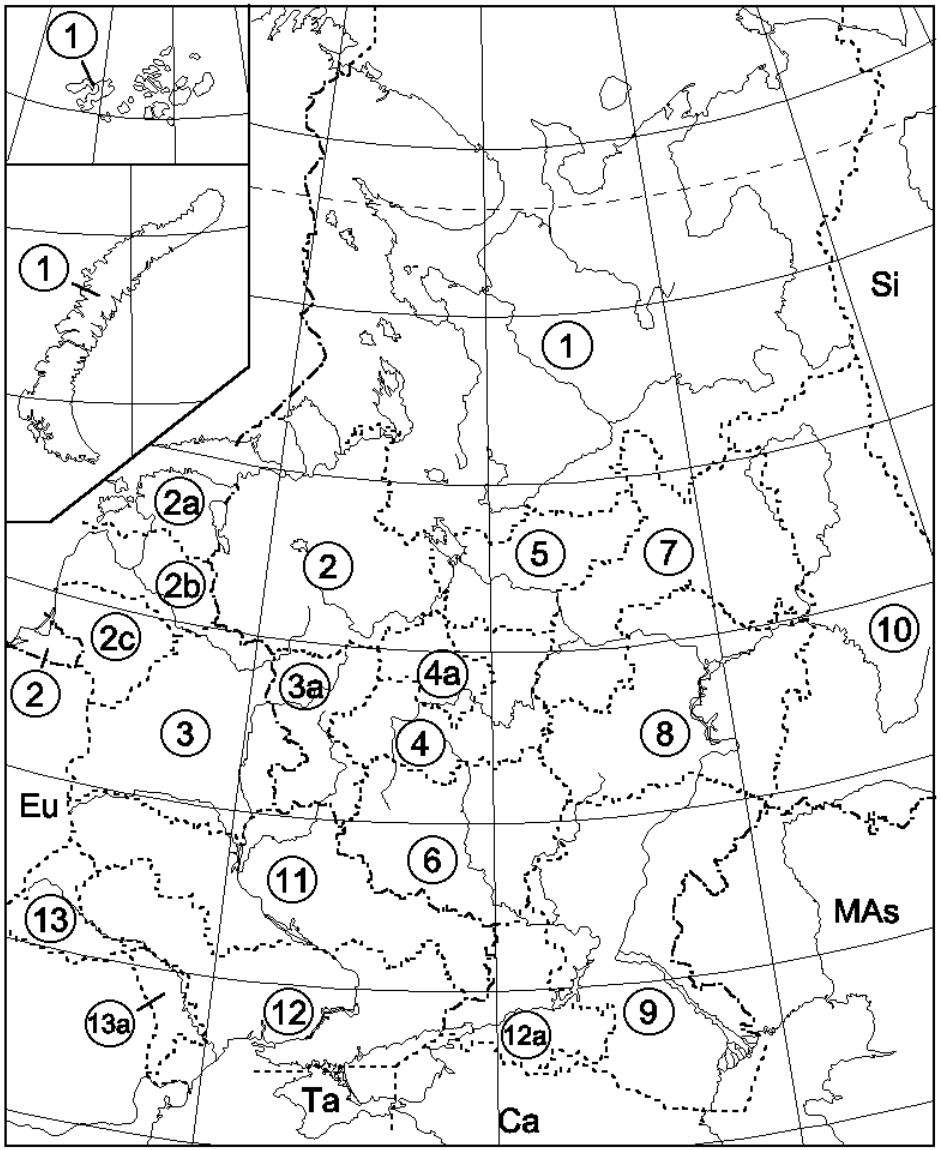
Digitisation:
Imaging as the First Step

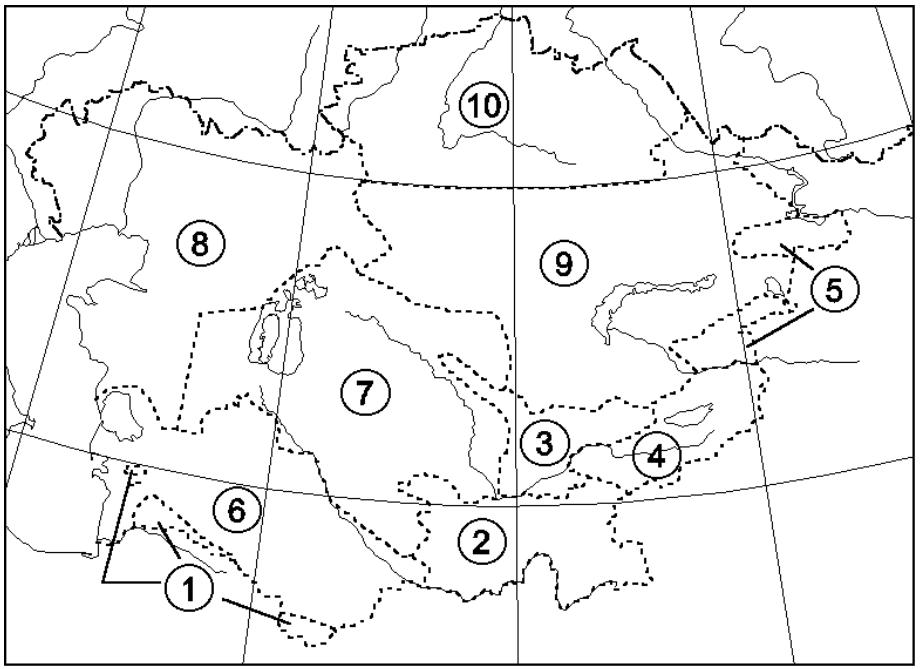
- Only imaging at 300 dpi TIFF (JPG copy)
- Basic metadata from folders (ID, species name, area code)
- No label data
- Free full online access

Typical folders with species names & area codes


Second phase of digitisation (September 2016)

Area Codes from Folders





↑ Area codes of Middle Asia

←Area codes of Eastern Europe

Metadata Production

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Fam	Index	×	Genus	×	Species	Authors	ssp/va	Infraspecific epit	Synonym	Msp	M1	M2	M3	M4	M5	M6	M7	M8
26856	26903	238a	6351		Limonium		bicolor	(Bunge) Kuntze												
26857	26904	238a	6351		Limonium		bungei	(Claus) Gamajun.												
26858	26905	238a	6351		Limonium		californicum	(Boiss.) Heller												
26859	26906	238a	6351		Limonium		cancellatum	(Bernh. ex Bertol.) Kuntze												
26860	26907	238a	6351		Limonium		carnosum	(Boiss.) Kuntze			Statice carnosa Boiss.									
26861	26908	238a	6351		Limonium		caspium	(Willd.) Gams											4	15
26862	26909	238a	6351		Limonium		chinensis	Girard.												
26863	26910	238a	6351		Limonium		chrysocomum	(Kar. et Kir.) Kuntze												
26864	26911	238a	6351		Limonium		congestum	(Ledeb.) Kuntze												
26865	26912	238a	6351		Limonium		coralloides	(Tausch) Lincz.												
26866	26913	238a	6351		Limonium		cordatum	(L.) Mill.												
26867	26914	238a	6351		Limonium		cretaceum	Tscherkasova												3
26868	26915	238a	6351		Limonium		densiflorum	(Guss.) Kuntze												
26869	26916	238a	6351		Limonium		dichroanthum	(Rupr.) Ikonn.-Gal.							11	1			1	
26870	26917	238a	6351		Limonium		diffusum	(Pourr.) Kuntze												
26871	26918	238a	6351		Limonium		donetzicum	Klokov												
26872	26919	238a	6351		Limonium		echioides	(L.) Mill.	subsp	exaristatum (Mu	L. exaristatum (Murb.) P. Fournier									
26873	26920	238a	6351		Limonium		echioides	(L.) Mill.												
26874	26921	238a	6351		Limonium		erythrorhizum	Ikonn.-Gal. ex Lincz.												
26875	26922	238a	6351		Limonium		feralaceum	(L.) Kuntze												
26876	26923	238a	6351		Limonium		ferganense	Ikonn.-Gal.						3	4					
26877	26924	238a	6351		Limonium		fischeri	(Trautv.) Lincz.			L. nudum ...									
26878	26925	238a	6351		Limonium		flexuosum	(L.) Kuntze												
26879	26926	238a	6351		Limonium		gmelinii	(Willd.) Kuntze							2		8		5	16
26880	26927	238a	6351		Limonium		gruhovii	Lincz.												

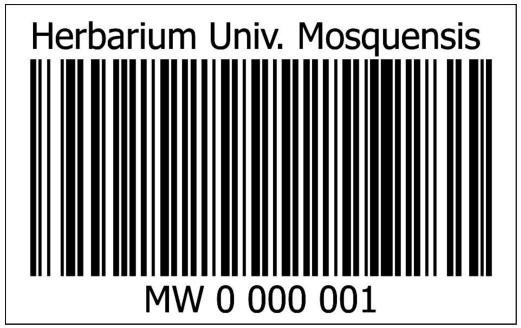
Table showing presence of 11 specimens of *Limonium dichroanthum* (Rupr.) Ikonn.-Gal. from M3 geographical area (Western Tian Shan).

Barcodes as an Accurate Counter

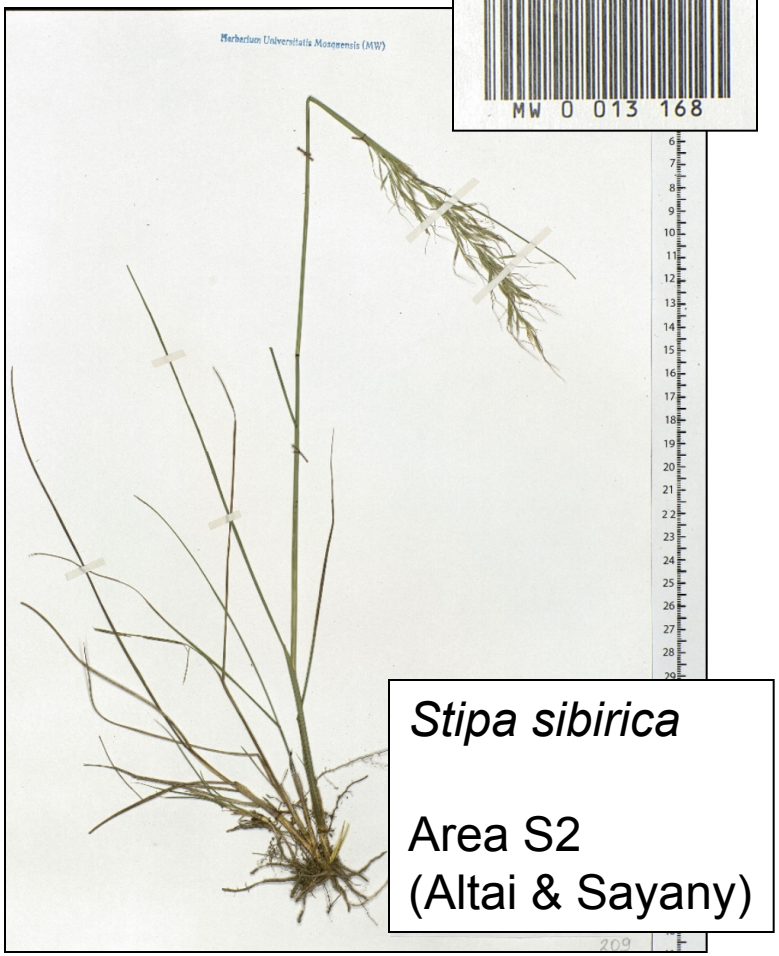
Eastern European collections:

- 338,940 specimens before barcoding
- 352,720 specimens after barcoding
- 13,780 increase (+4.07%)

Barcoding prior to scanning in 2015



Metadata Example



Digitisation: Quick Geotagging

- Country names were used as geotags
- 756K specimens were geotagged using existing country names from folders
- 200K specimens were geotagged manually in 2017-2019

Collection Growth Due to Digitisation

- Mean annual increase in 2005-2016 was 15,100 specimens

- In 2016, we added 22,013 specimens (expeditions, old collections, gifts, exchange, etc.)

Herbarium Specimen

A dry plant holds

-morphological
-anatomical
-genetic
-chemical

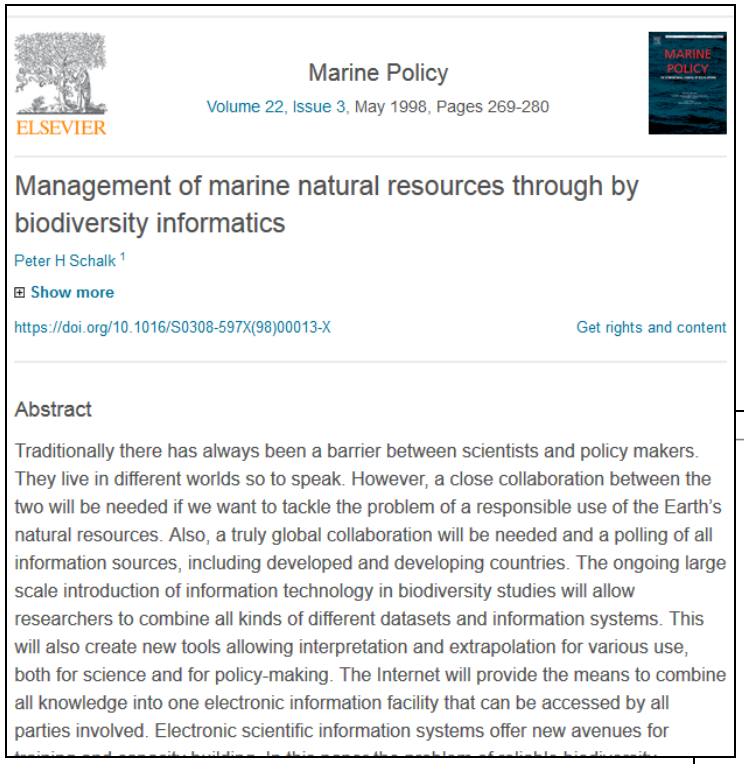
information.

A label holds

-geographical
-ecological
-historical
-material

information.

Biodiversity Informatics



Marine Policy
Volume 22, Issue 3, May 1998, Pages 269-280

Management of marine natural resources through by biodiversity informatics

Peter H Schalk [1]

Show more

https://doi.org/10.1016/S0308-597X(98)00013-X

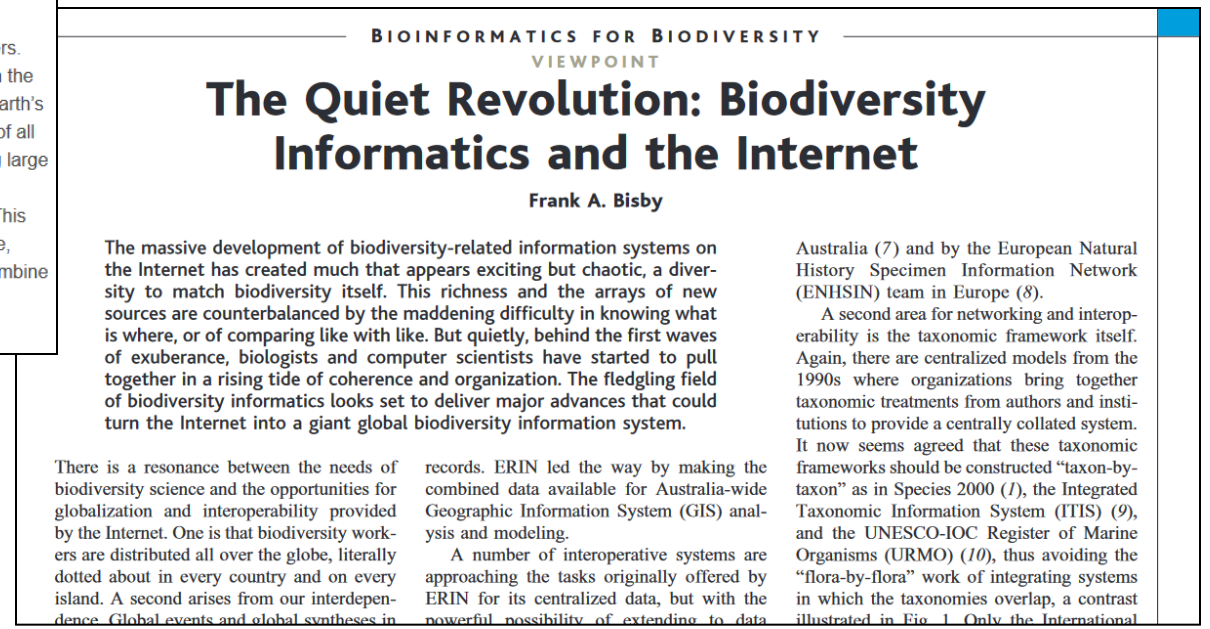
Get rights and content

Abstract

Traditionally there has always been a barrier between scientists and policy makers. They live in different worlds so to speak. However, a close collaboration between the two will be needed if we want to tackle the problem of a responsible use of the Earth's natural resources. Also, a truly global collaboration will be needed and a polling of all information sources, including developed and developing countries. The ongoing large scale introduction of information technology in biodiversity studies will allow researchers to combine all kinds of different datasets and information systems. This will also create new tools allowing interpretation and extrapolation for various use, both for science and for policy-making. The Internet will provide the means to combine all knowledge into one electronic information facility that can be accessed by all parties involved. Electronic scientific information systems offer new avenues for

Marine Policy (1998)

A discipline on the crossing of
- biology
- geography
- computer science



BIOINFORMATICS FOR BIODIVERSITY

VIEWPOINT

The Quiet Revolution: Biodiversity Informatics and the Internet

Frank A. Bisby

The massive development of biodiversity-related information systems on the Internet has created much that appears exciting but chaotic, a diversity to match biodiversity itself. This richness and the arrays of new sources are counterbalanced by the maddening difficulty in knowing what is where, or of comparing like with like. But quietly, behind the first waves of exuberance, biologists and computer scientists have started to pull together in a rising tide of coherence and organization. The fledgling field of biodiversity informatics looks set to deliver major advances that could turn the Internet into a giant global biodiversity information system.

There is a resonance between the needs of biodiversity science and the opportunities for globalization and interoperability provided by the Internet. One is that biodiversity workers are distributed all over the globe, literally dotted about in every country and on every island. A second arises from our interdependence. Global events and global syntheses in

records. ERIN led the way by making the combined data available for Australia-wide Geographic Information System (GIS) analysis and modeling.

A number of interoperative systems are approaching the tasks originally offered by ERIN for its centralized data, but with the powerful possibility of extending to data

Australia (*7*) and by the European Natural History Specimen Information Network (ENHSIN) team in Europe (*8*).

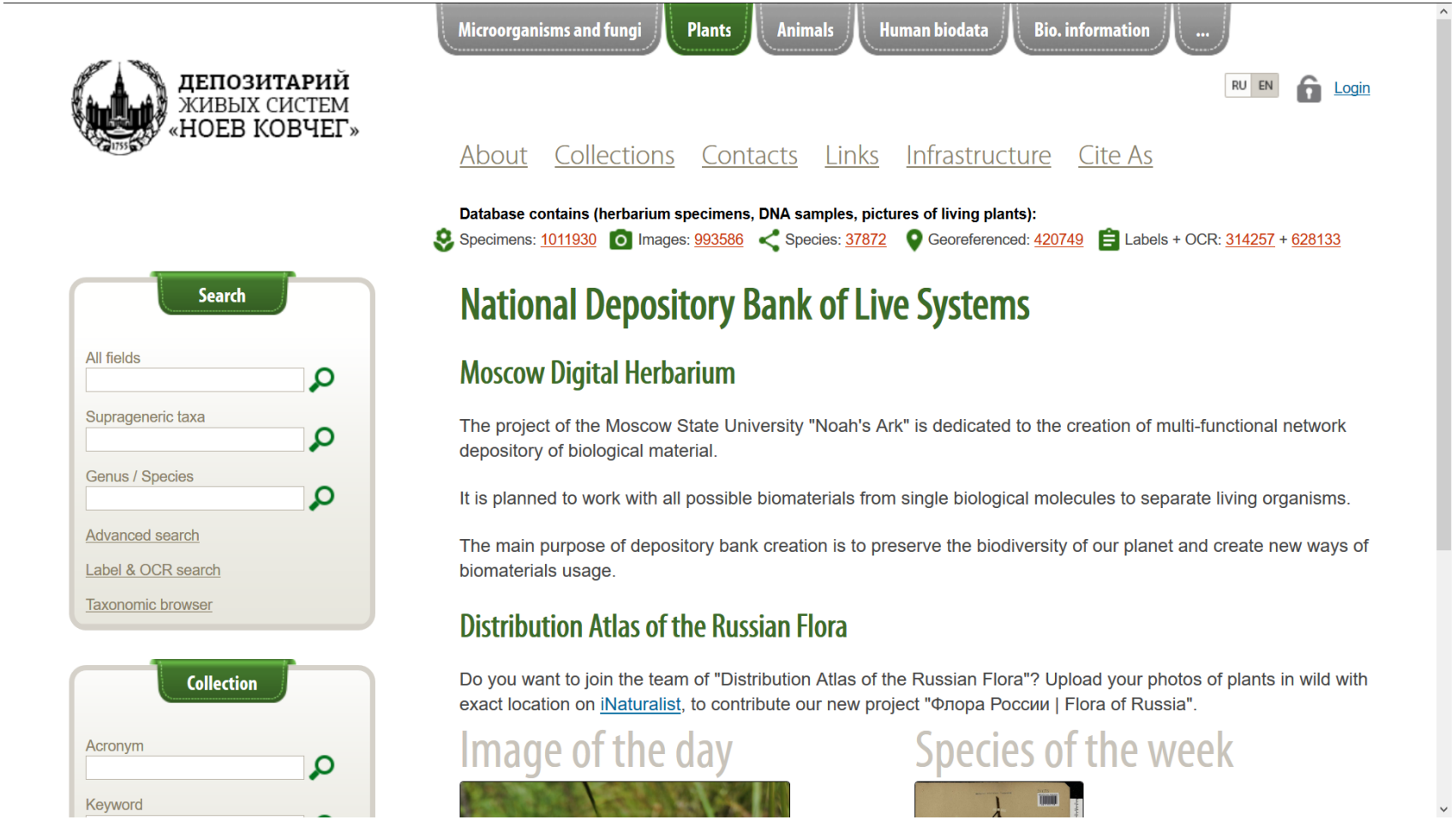
A second area for networking and interoperability is the taxonomic framework itself. Again, there are centralized models from the 1990s where organizations bring together taxonomic treatments from authors and institutions to provide a centrally collated system. It now seems agreed that these taxonomic frameworks should be constructed "taxon-by-taxon" as in Species 2000 (*1*), the Integrated Taxonomic Information System (ITIS) (*9*), and the UNESCO-IOC Register of Marine Organisms (URMO) (*10*), thus avoiding the "flora-by-flora" work of integrating systems in which the taxonomies overlap, a contrast illustrated in Fig. 1. Only the International

Science (2000)

Online Access Started in 2017

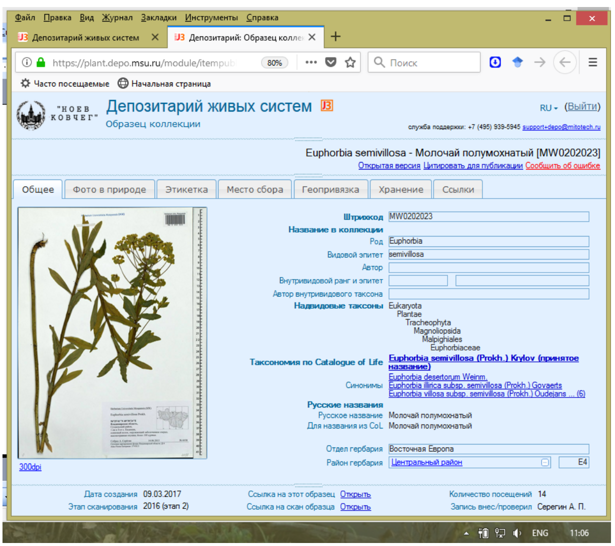
- http://plant.depo.msu.ru/ (public access)
- Russian and English versions
- 786K specimens at 300 dpi (JPG)
- Automatic linking of the collection names with *Catalogue of Life* taxonomy
- No label data at that moment
- No geographical coordinates

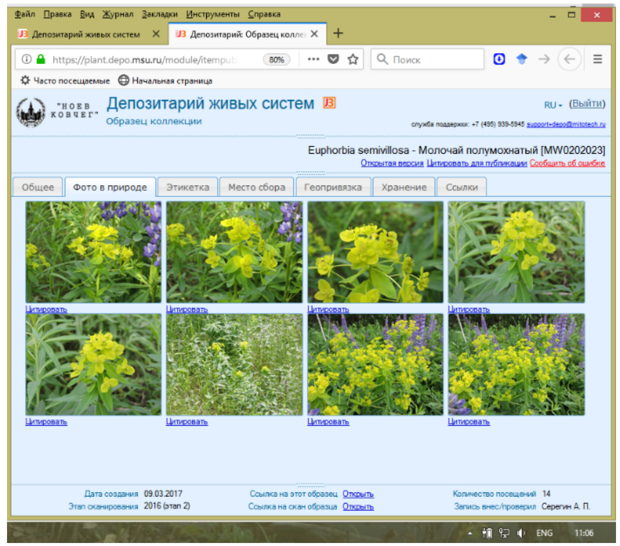
https://plant.depo.msu.ru/

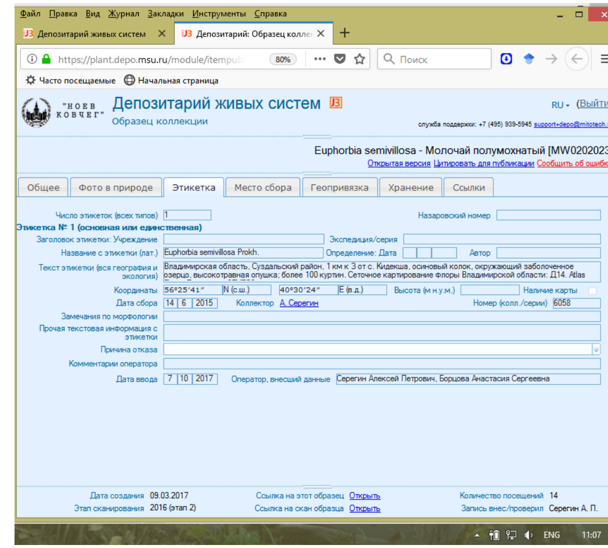


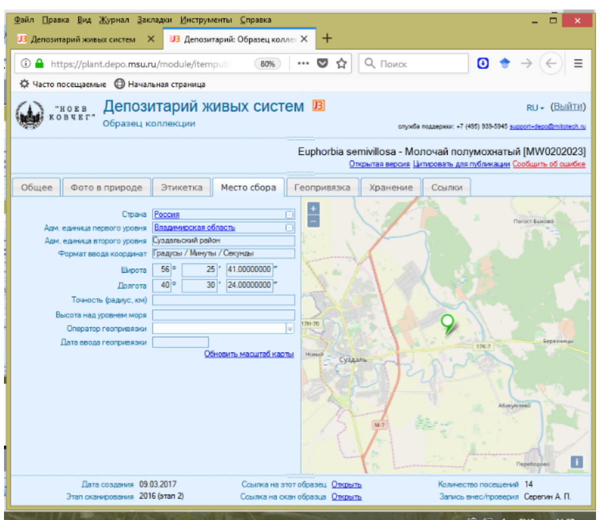
Home page

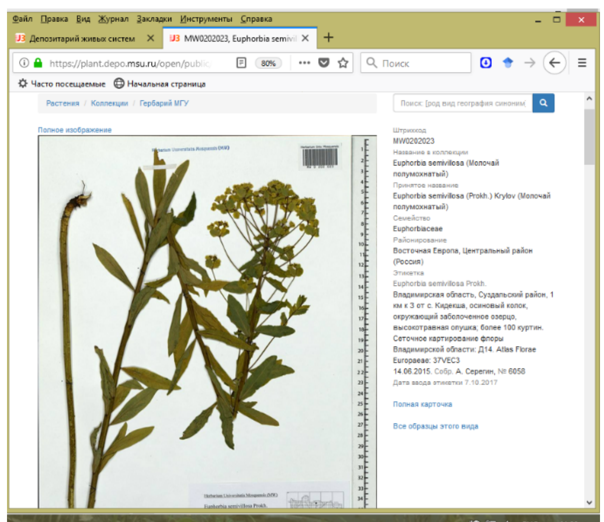
Moscow Digital Herbarium Screenshots

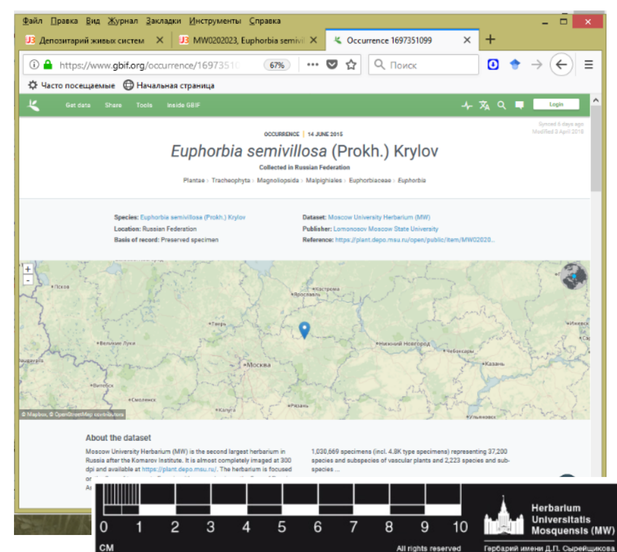




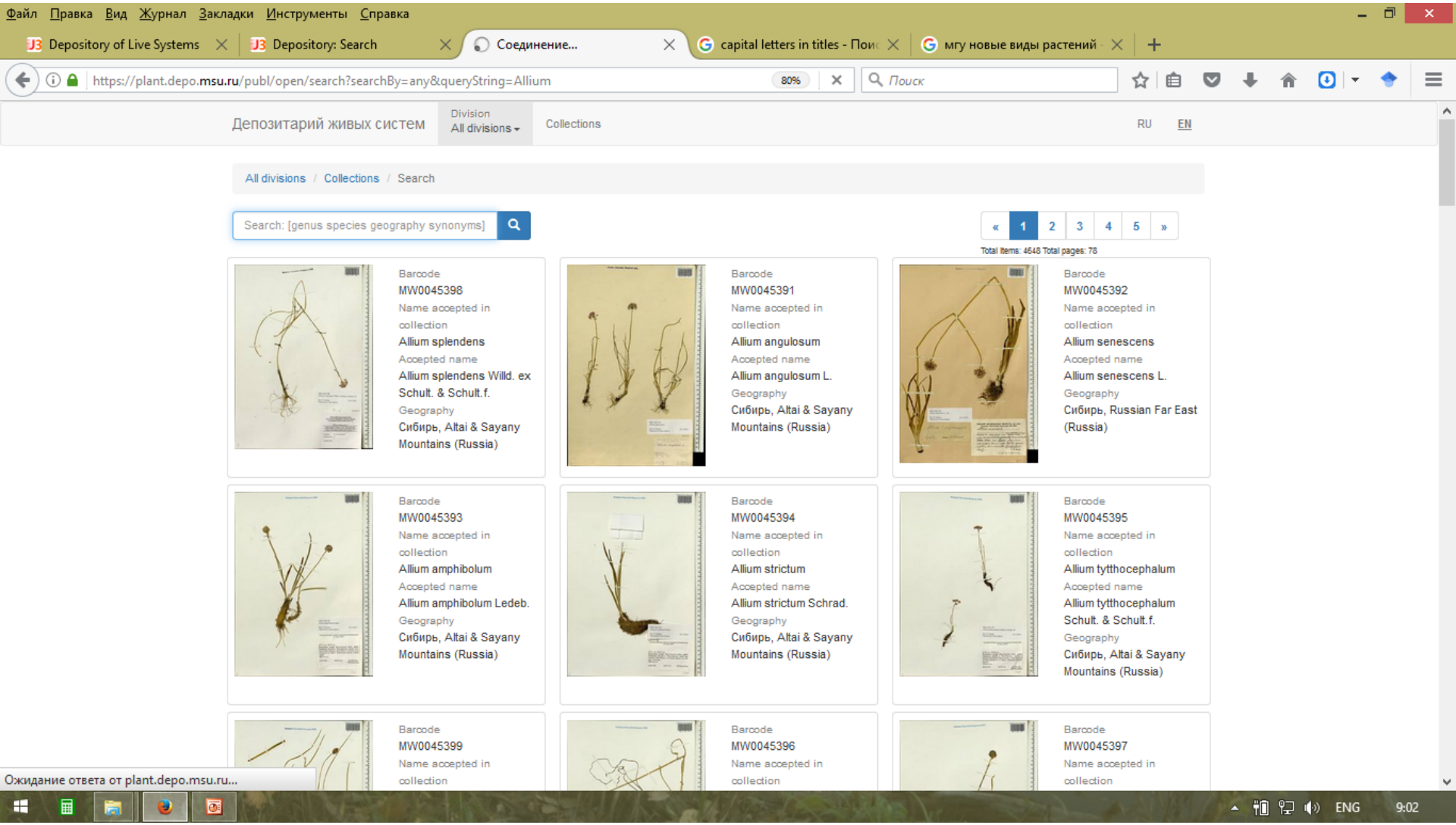






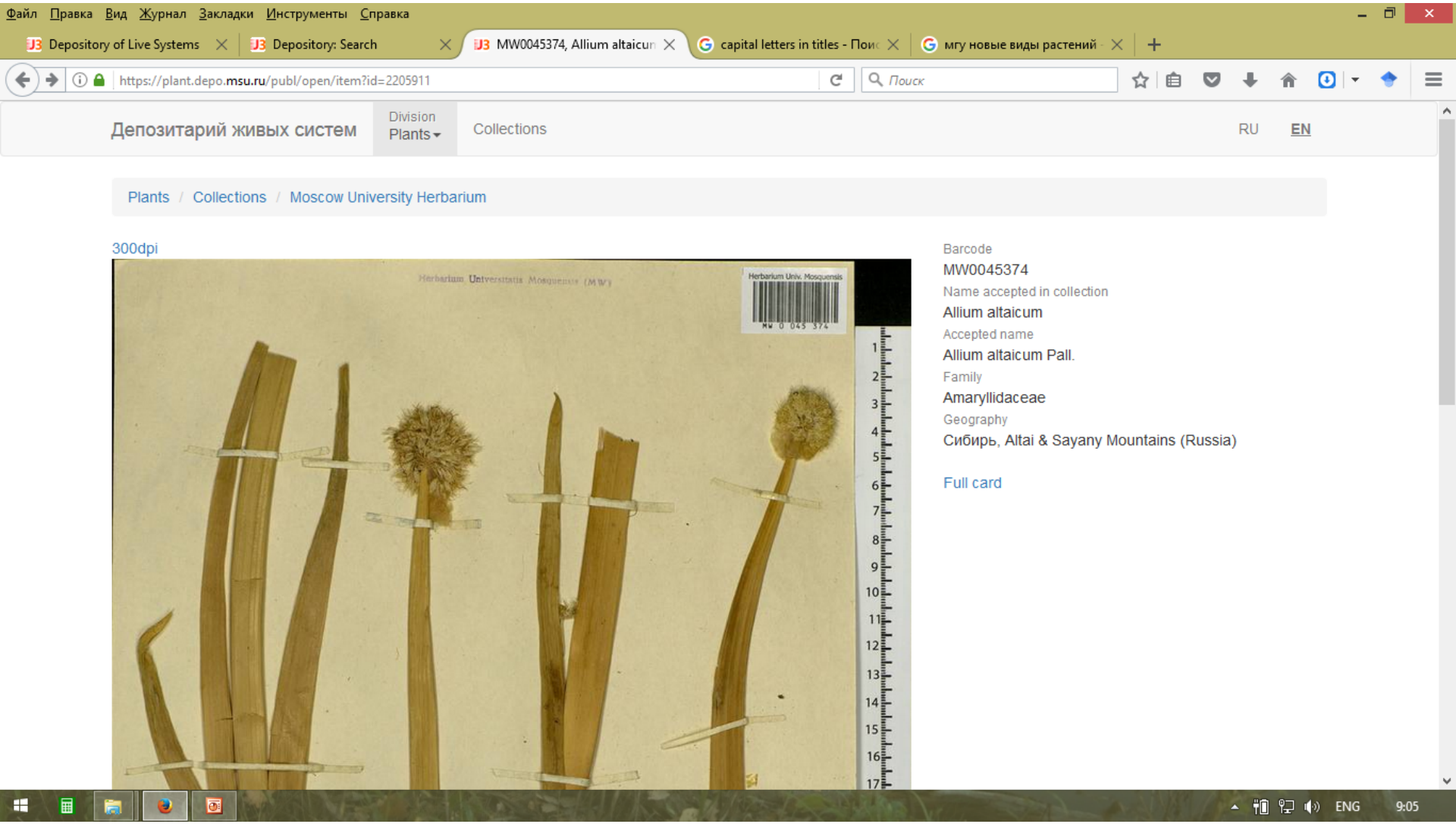


https://plant.depo.msu.ru/open/



Alternative results page used for *Google* indexing

https://plant.depo.msu.ru/open/

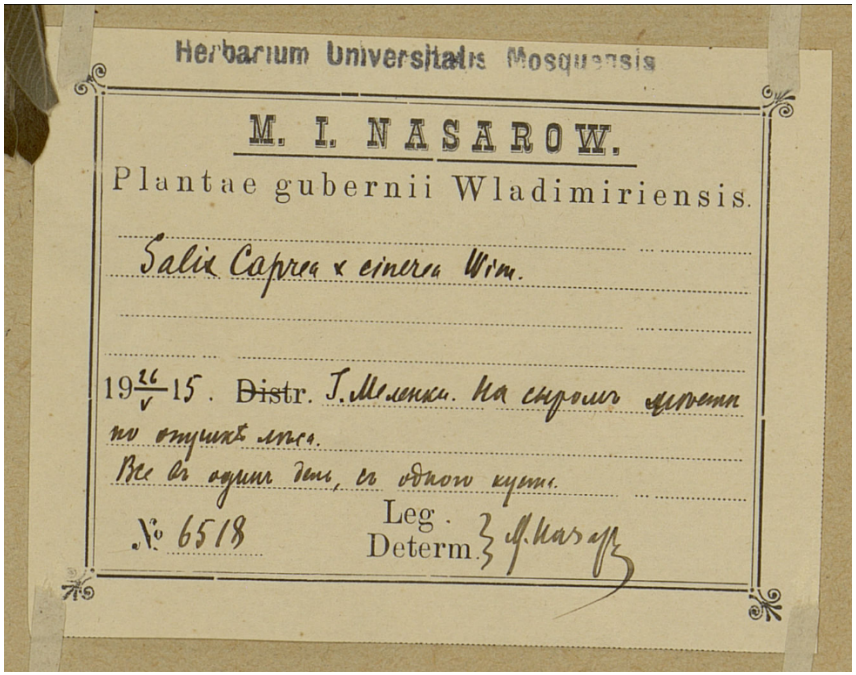


Alternative specimen record used for *Google* indexing

Digitisation: Label Capturing as the Second Step

- 5,000 labels were databased by September 2017
- 796,508 partly captured labels by 3.09.2019 (collector & date)
- 314,257 fully captured labels by 3.09.2019
- Original label language (no translation)
- 50% of labels captured by the commercial partner

Why We Really Need Label Capturing?



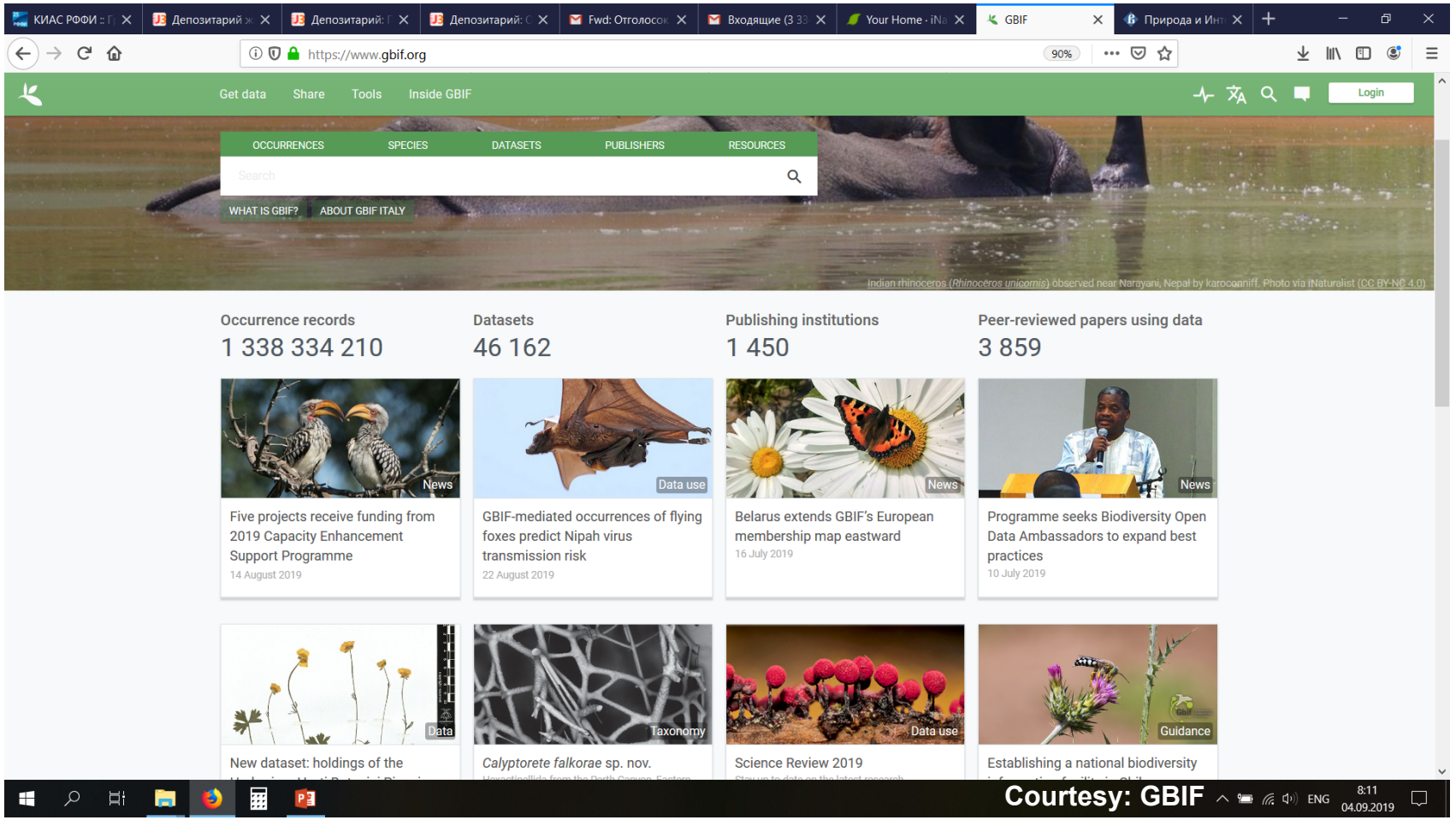
- M.I. NASAROW. Plantae gubernii Wladimiriensis.
  Salix caprea × cinerea Wim.
  Г. Меленки. На сыром месте по опушке леса.
  26.05.1915. Собр. М. Назаров, № 6518
  Все в один день, с одного куста.

- AUTOMATIC TRANSLATION: M.I. NASAROW. Plantae gubernii Wladimiriensis.
  Salix caprea × cinerea Wim. Melenki. In a damp place along the edge of the forest. 26.05.1915. Coll. M. Nazarov, No. 6518 All in one day, from one bush.

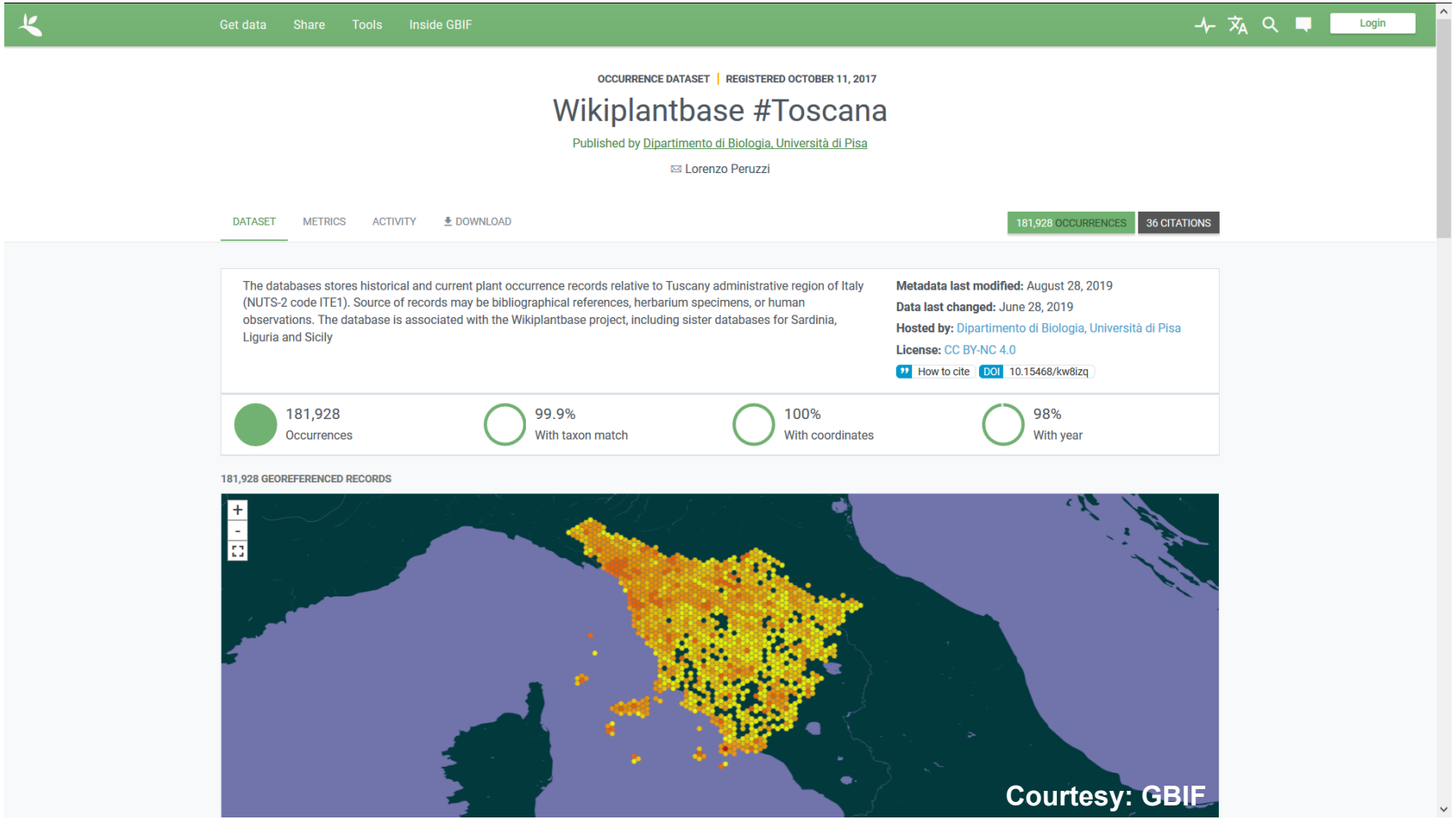
Digitisation: Georeferencing as the Third Step

- 5,000 labels were georeferenced by September 2017
- 420,749 georeferences by 3.09.2019:
  - 216,952 precise manual georefs
  - 203,797 automatic georefs based on collector/date matches
- N, E coordinates (+ uncertainty)
- 100% of georefs made by the herbarium team

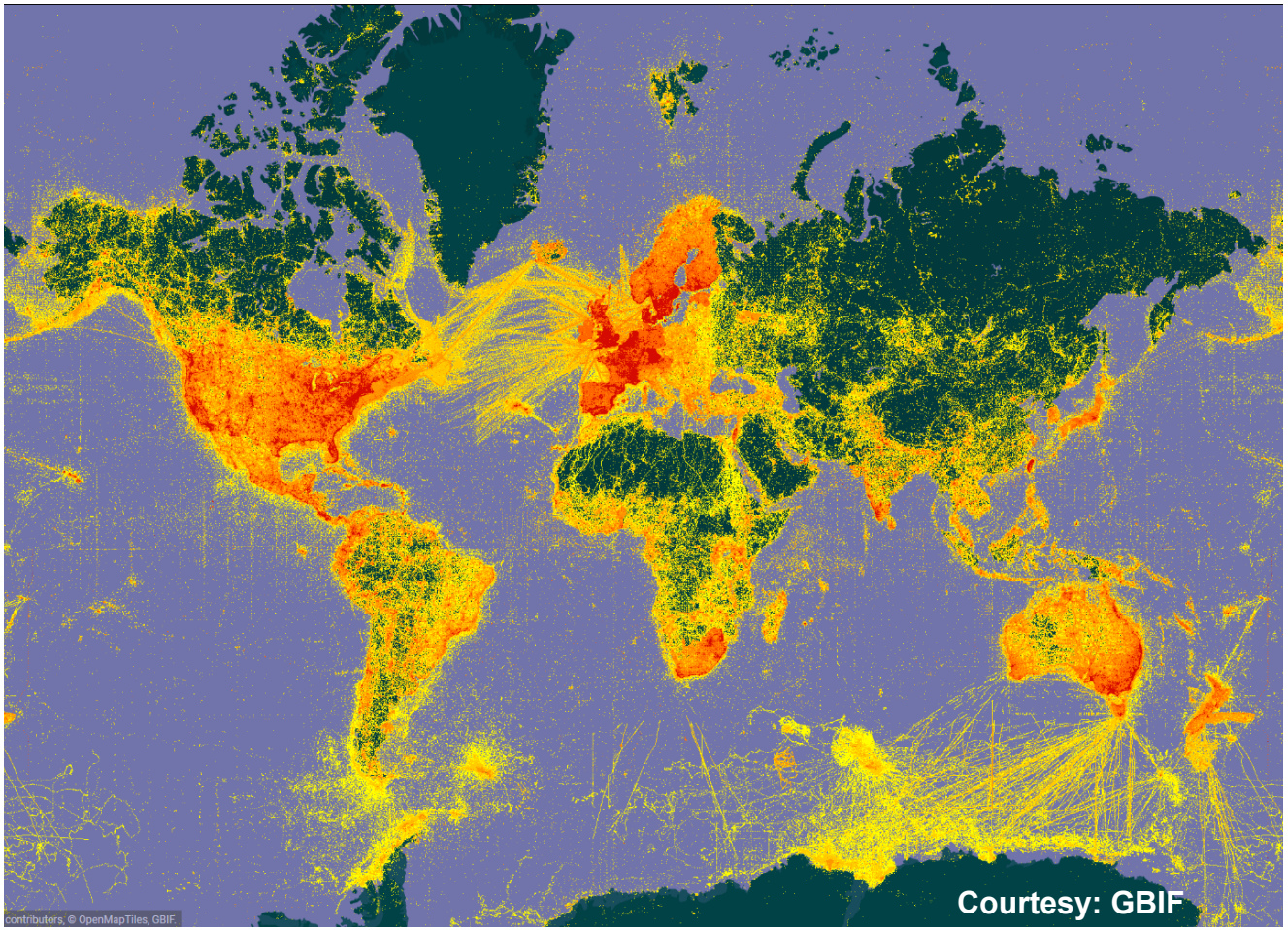
Global Biodiversity Information Facility (GBIF)



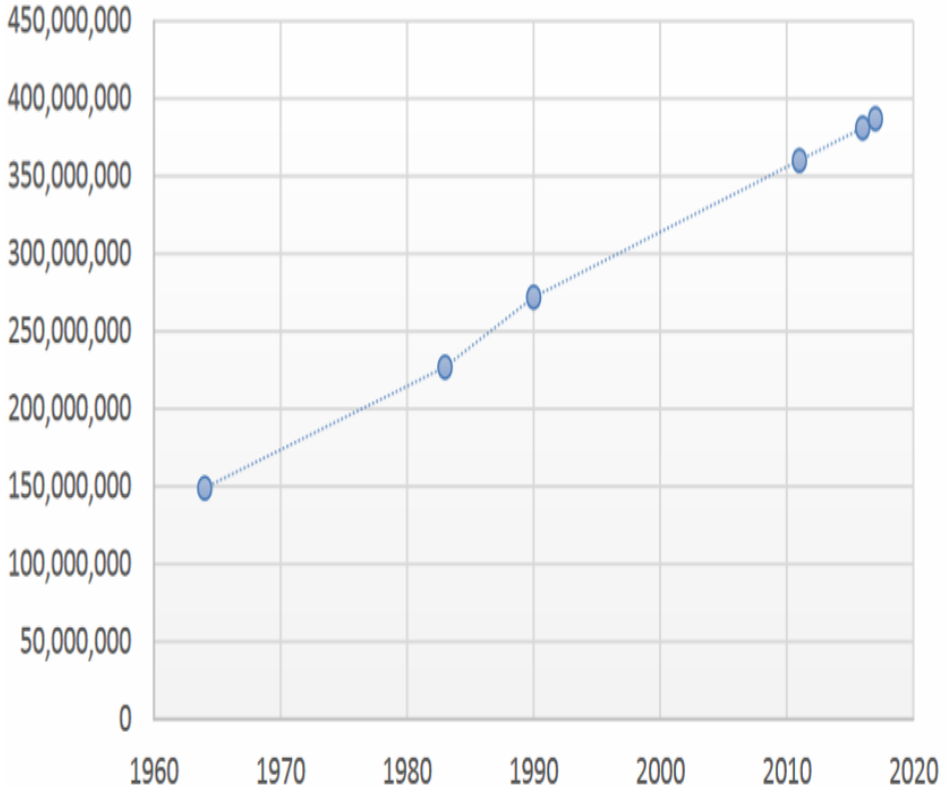
The Largest Italian Dataset in GBIF



Russia on GBIF Data Map



World Herbaria Holdings

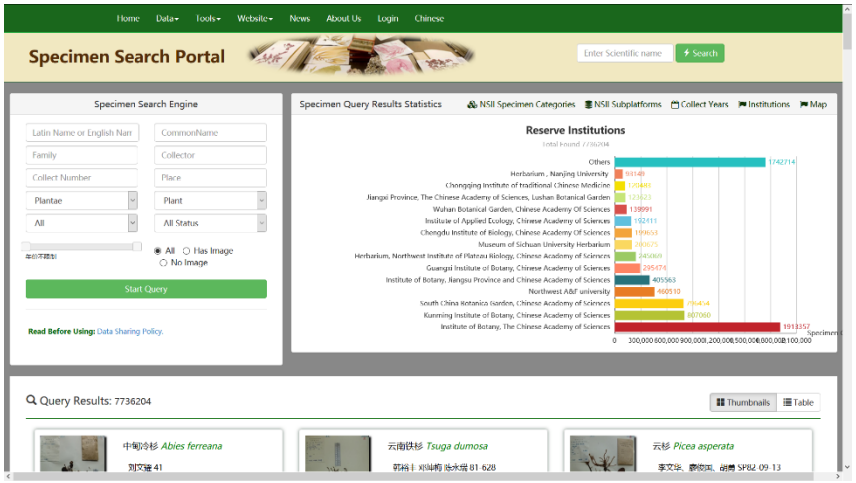


- 387 500 000 specimens are preserved in world herbaria (Thiers, 2019)
- 75 098 731 specimen records are indexed in GBIF, or **19,4%**
- 34 826 192 specimen records have coordinates in GBIF, or **9,0%**

World Herbaria in GBIF: 75,098,731 Specimens

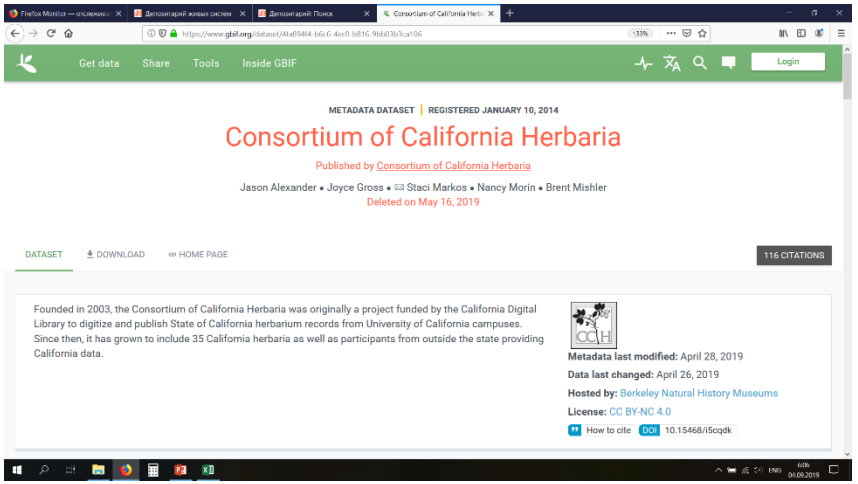
• The vascular plants collection (P) at the Herbariu…	5,426,095
• MEL AVH data	5,071,904
• Naturalis Biodiversity Center (NL) - Botany	4,824,298
• Tropicos Specimen Data	4,668,829
• The New York Botanical Garden Herbarium (NY)	3,266,213
• NMNH Extant Specimen Records	2,952,134
• Meise Botanic Garden Herbarium (BR)	1,424,986
• PRECIS	1,117,942
• Harvard University Herbaria: All Records	1,060,214
• Phanerogamic Botanical Collections (S)	1,018,284
• **Moscow University Herbarium (MW)**	**1,001,832**
• EURISCO, The European Genetic Resources Search Cat…	956,422
• Royal Botanic Gardens, Kew - Herbarium Specimens	921,988
• Lund Botanical Museum (LD)	911,304
• Royal Botanic Garden Edinburgh Herbarium (E)	866,672

Some Essential Additions



7,7M records (China)

At least 10M specimen records are not currently available in GBIF making the world's total ca. 85,000,000 databased specimens, or **22%**



2,3M records (California)

Records from Russia in GBIF: a Big Bang in November, 2017

Get data Share Tools Inside GBIF Login

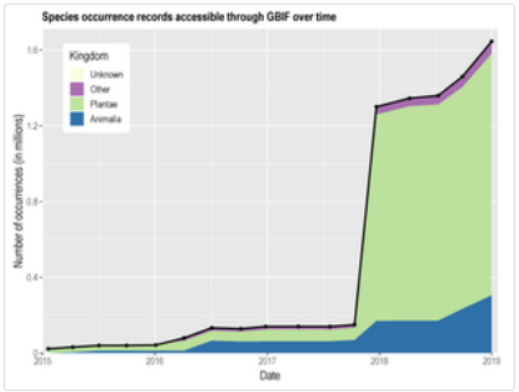
PUBLISHING TRENDS

Number of occurrence records

These charts illustrate the change in availability of the species occurrence records over time.

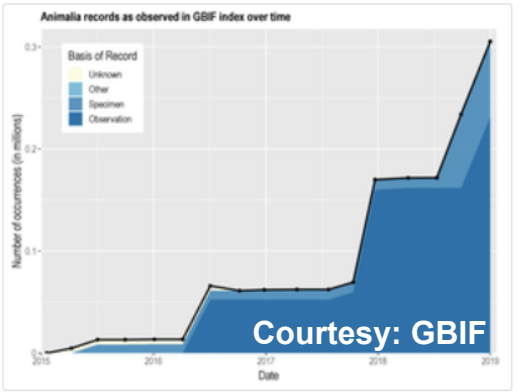
Records by kingdom

The number of available records categorized by kingdom. "Unknown" includes records with taxonomic information that cannot be linked to available taxonomic checklists.



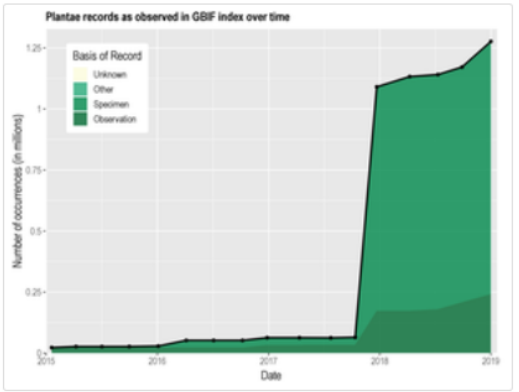
Records for Animalia

The number of animal records categorized by the basis of record. "Unknown" includes records without defined basis of record or with an unrecognised value for basis of record.

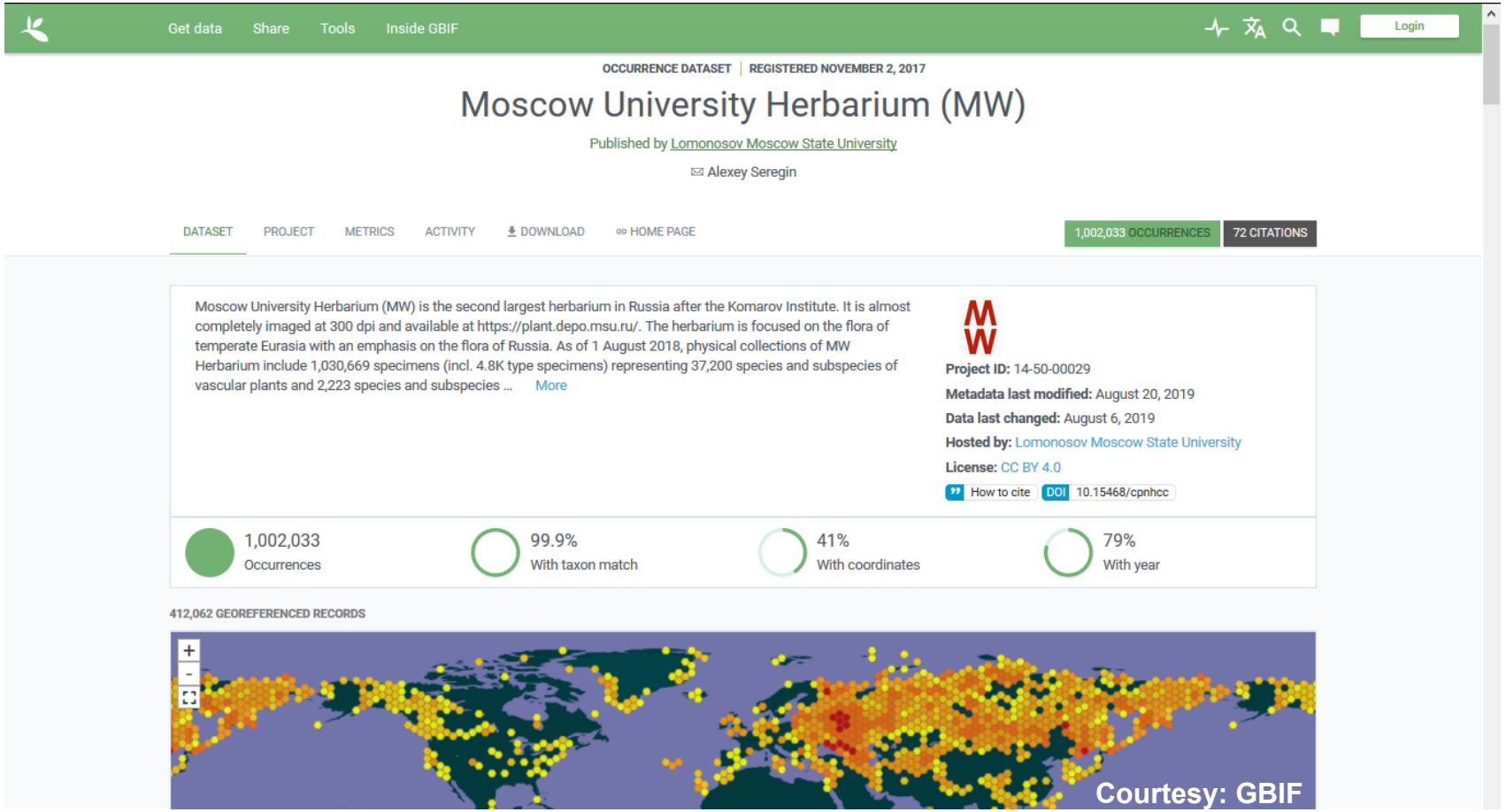


Records for Plantae

The number of plant records categorized by the basis of record. "Unknown" includes records without defined basis of record or with an unrecognised value for basis of record.



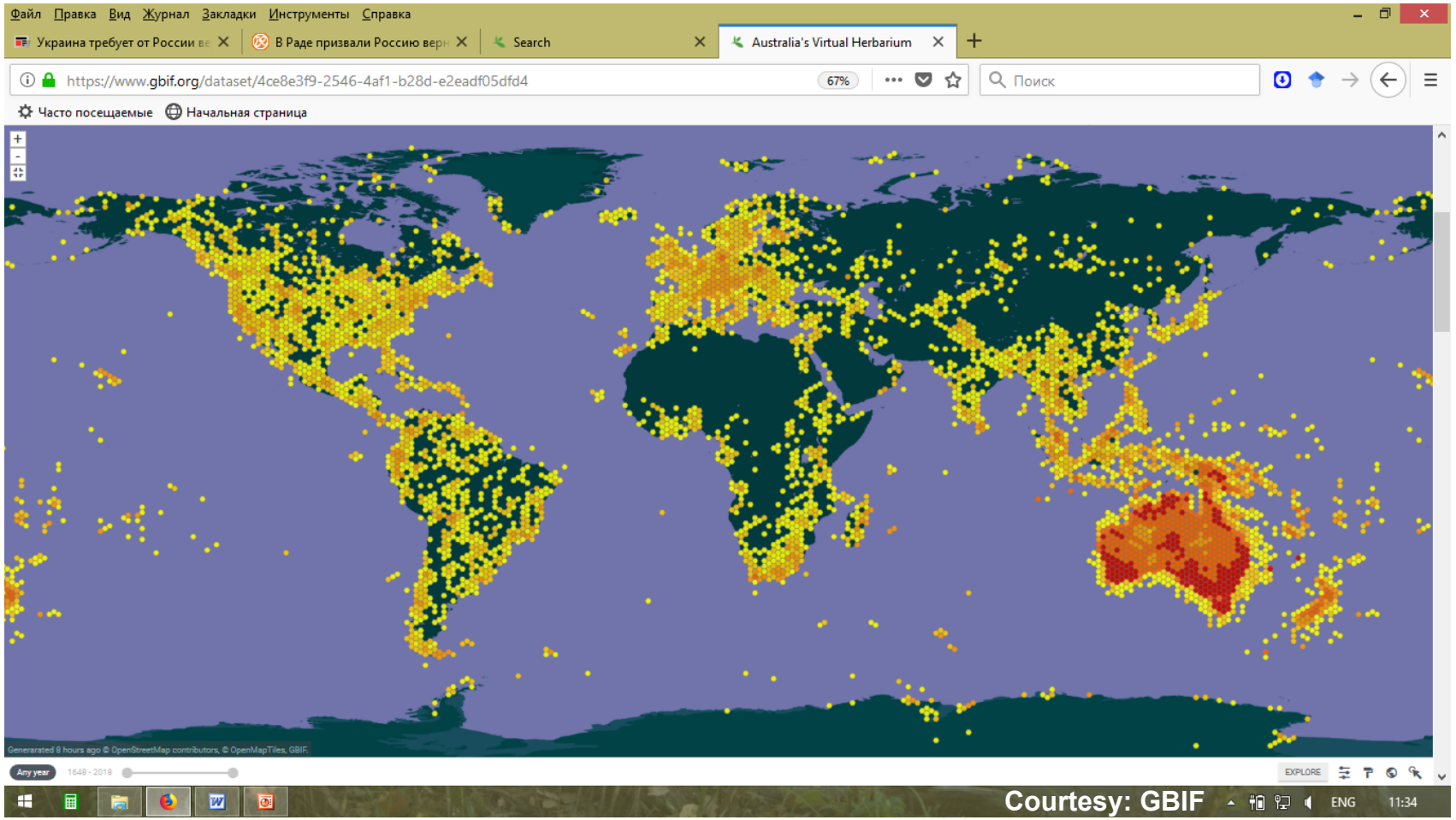
MW Dataset in GBIF: 1,002,033 Records



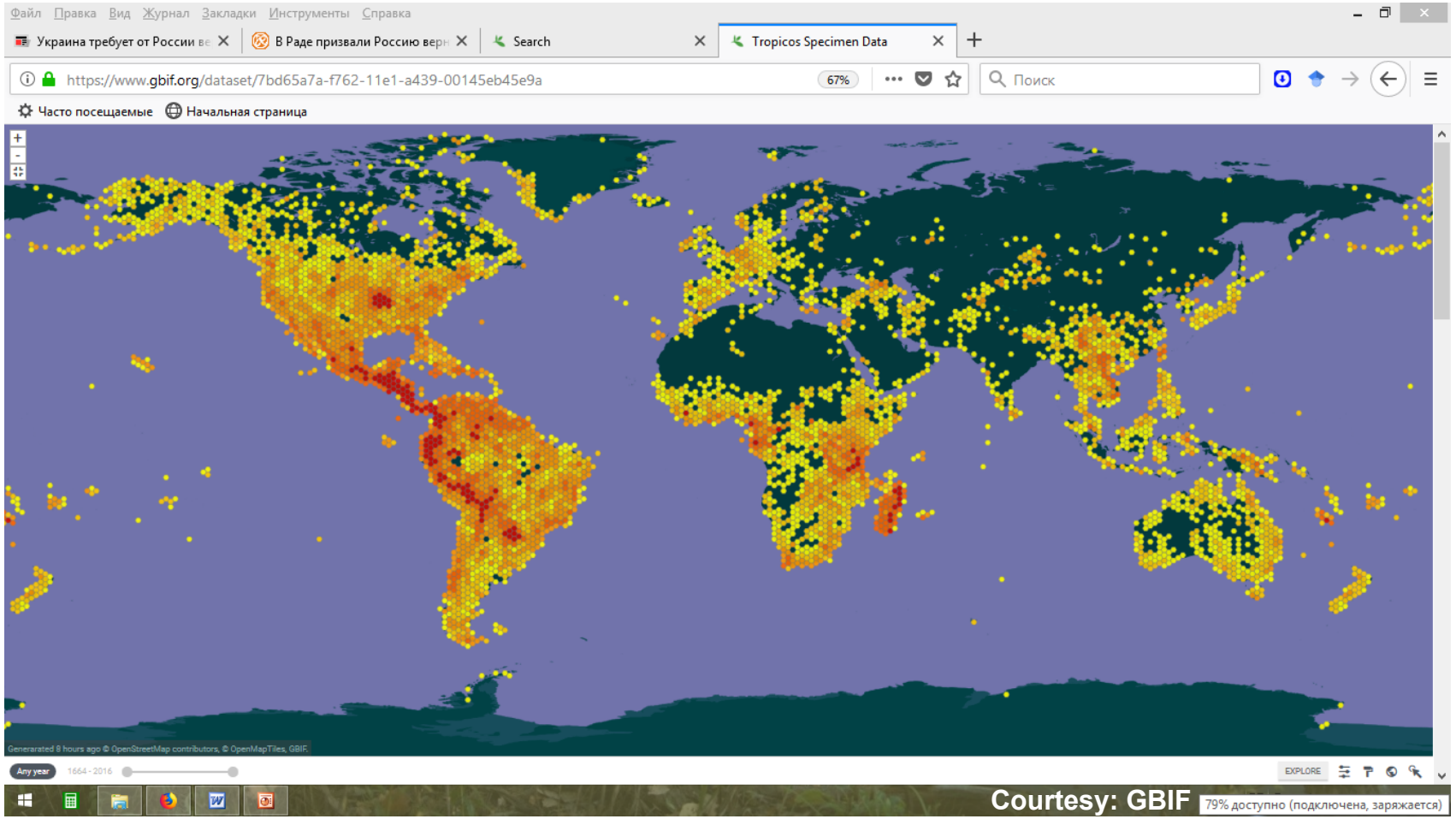
World Herbaria in GBIF: 34,826,192 Georeferences

• MEL AVH data	4,667,925
• Tropicos Specimen Data	3,125,919
• Naturalis Biodiversity Center (NL) - Botany	1,488,255
• The New York Botanical Garden Herbarium (NY)	1,021,598
• PRECIS	903,698
• Lund Botanical Museum (LD)	871,034
• Vascular Plant Herbarium, Oslo (O)	584,762
• Field Museum of Natural History (Botany) Seed Pla…	554,080
• NMNH Extant Specimen Records	512,474
• BRI AVH data	500,208
• **Moscow University Herbarium (MW)**	**412,188**
• Finnish Floristic Database (Finnish Museum of Natu…	397,381
• Plantae of Costa Rica (INBio)	351,417
• Natural History Museum (London) Collection Specime…	332,557
• OEH Atlas of NSW Wildlife	328,787

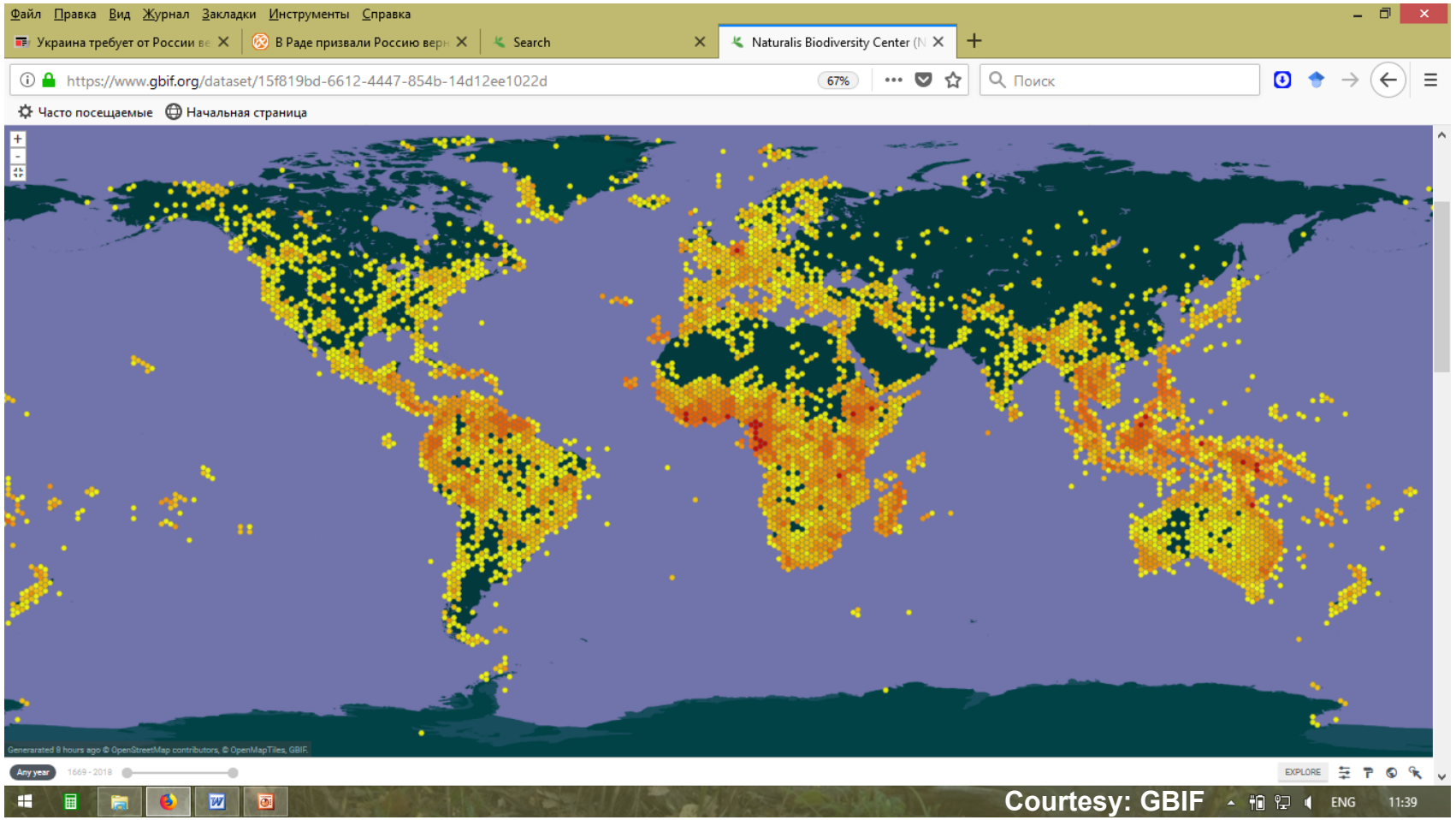
1. Australia's Virtual Herbarium



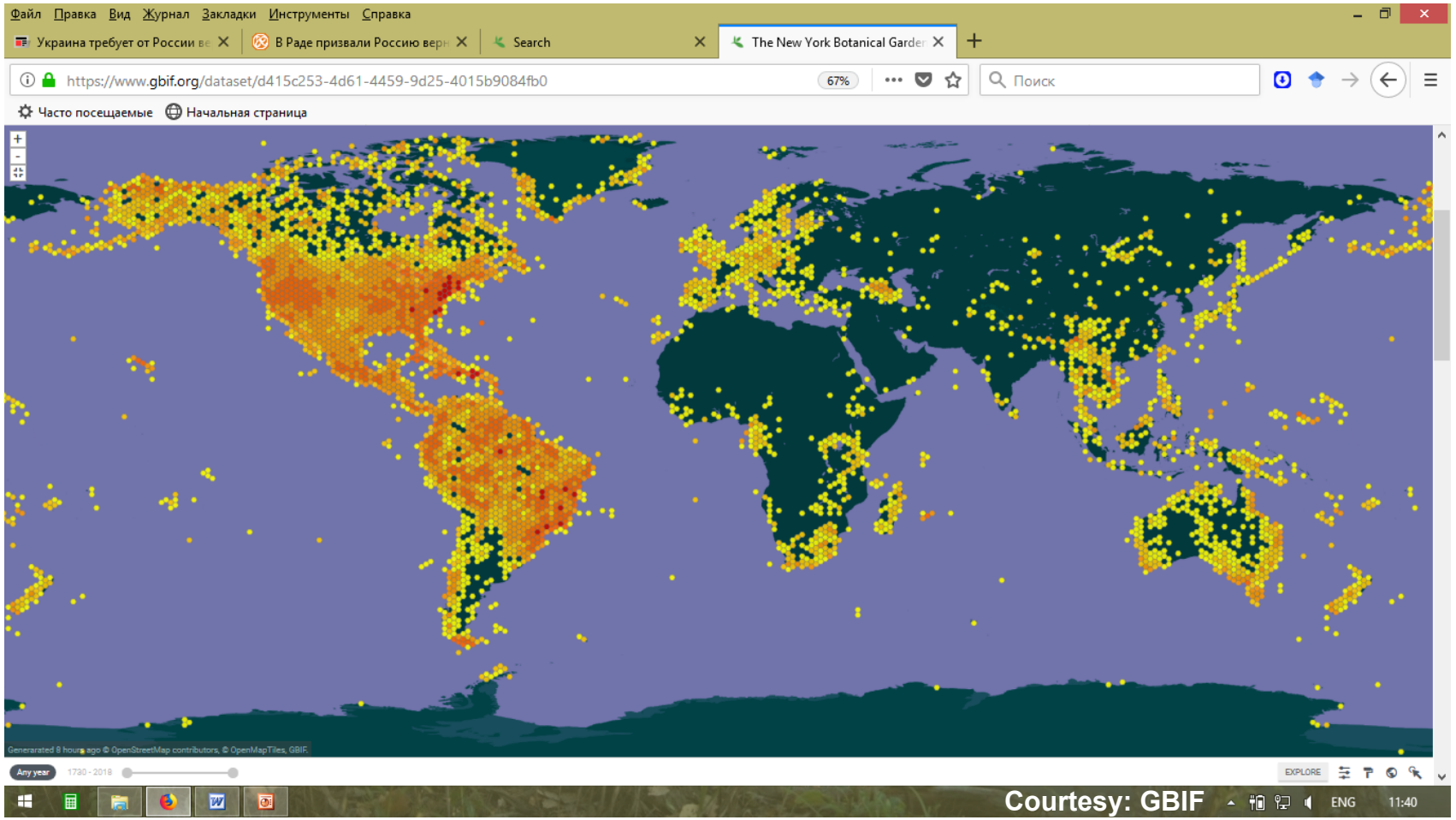
2. Tropicos Specimen Data



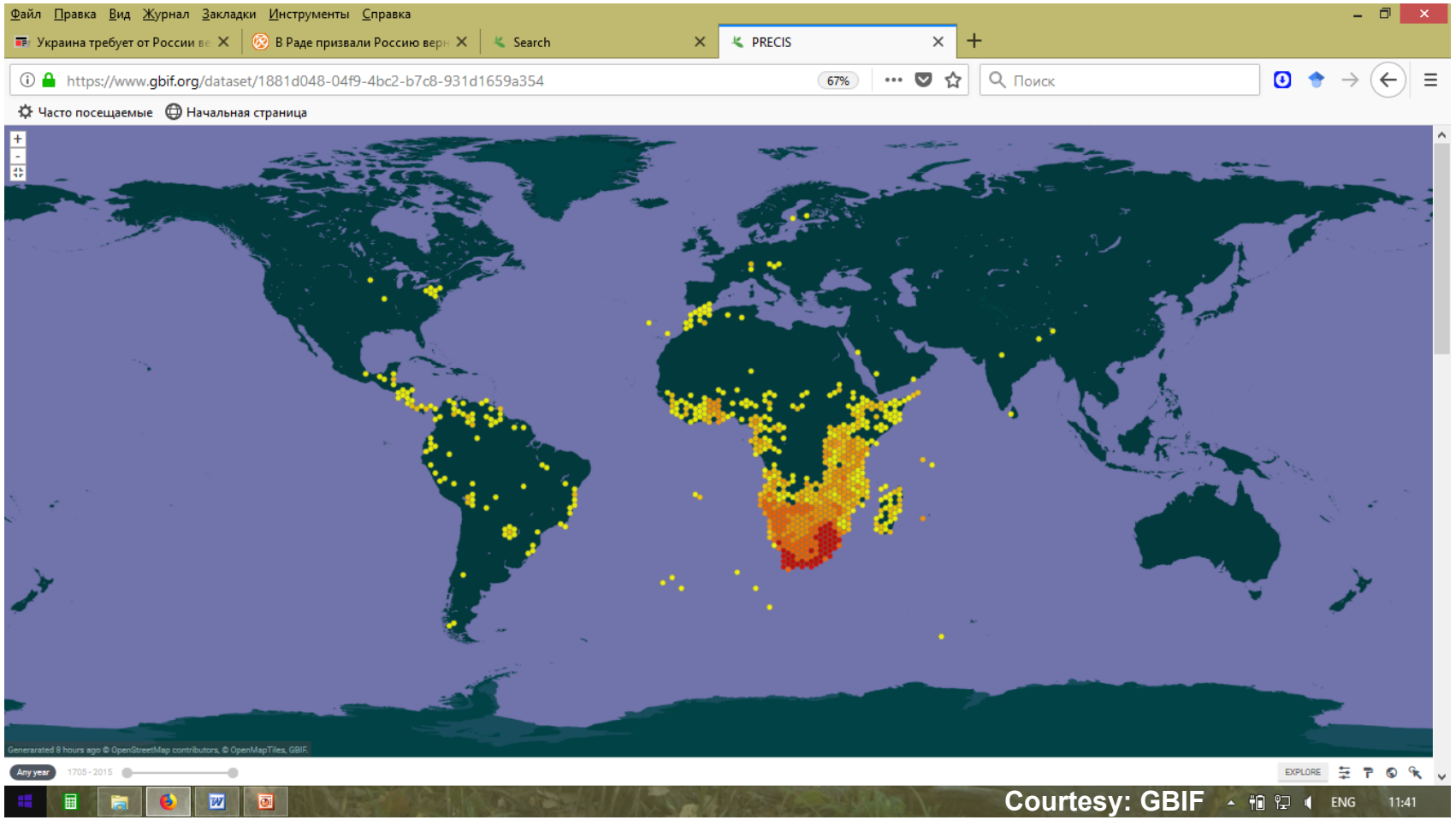
3. Naturalis Biodiversity Center (NL) - Botany



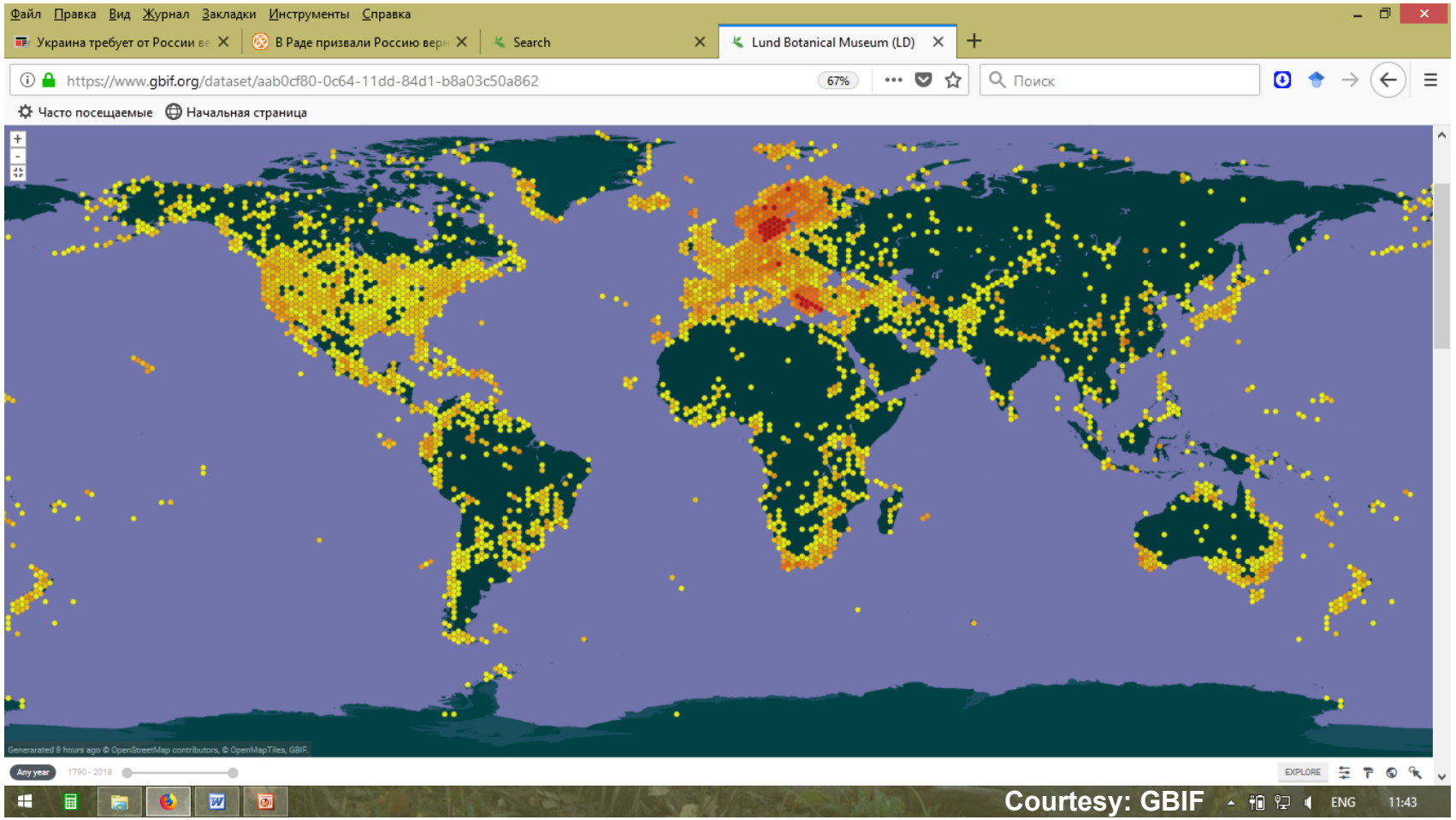
4. The New York Botanical Garden Herbarium (NY)



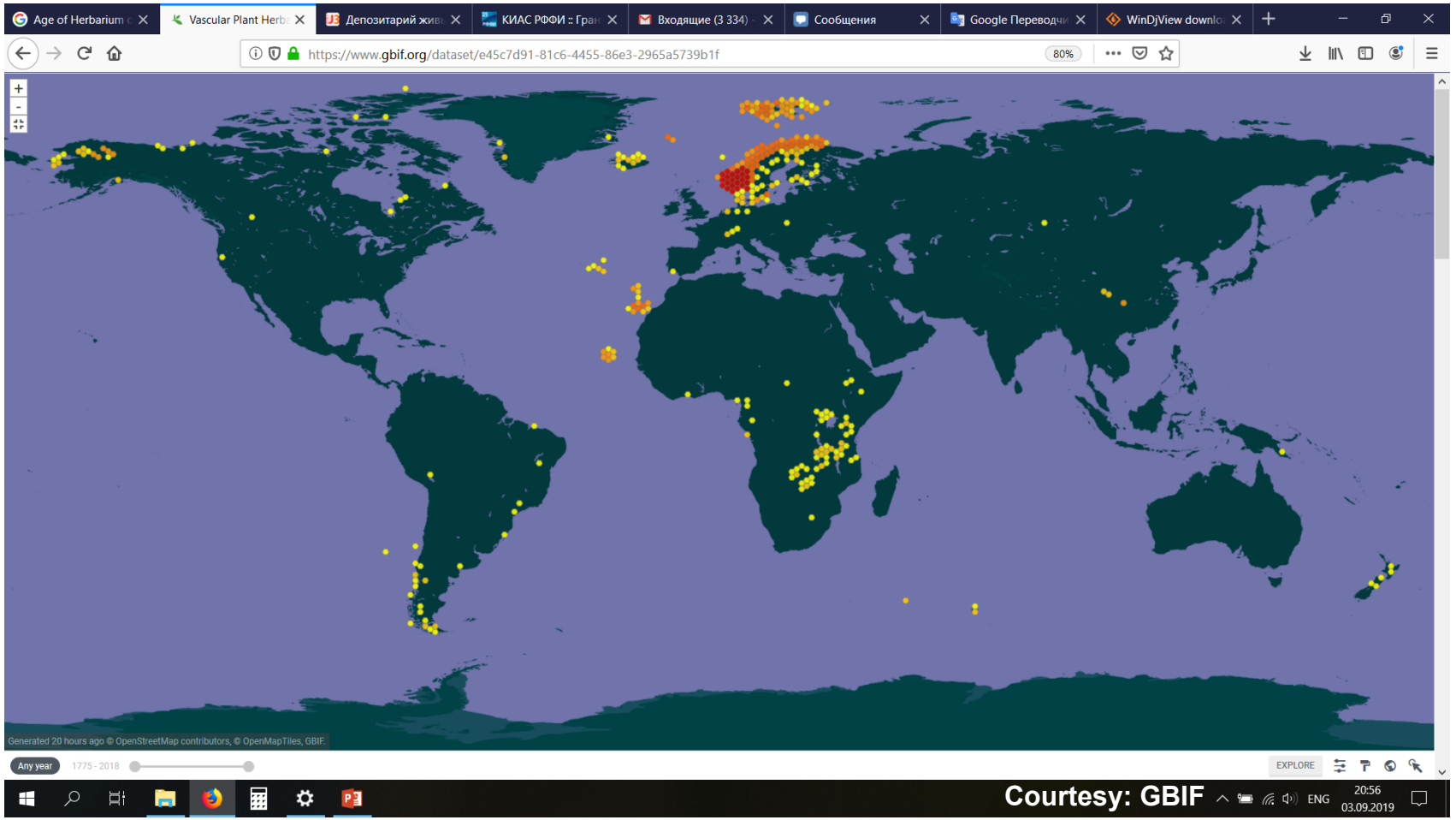
5. PRECIS



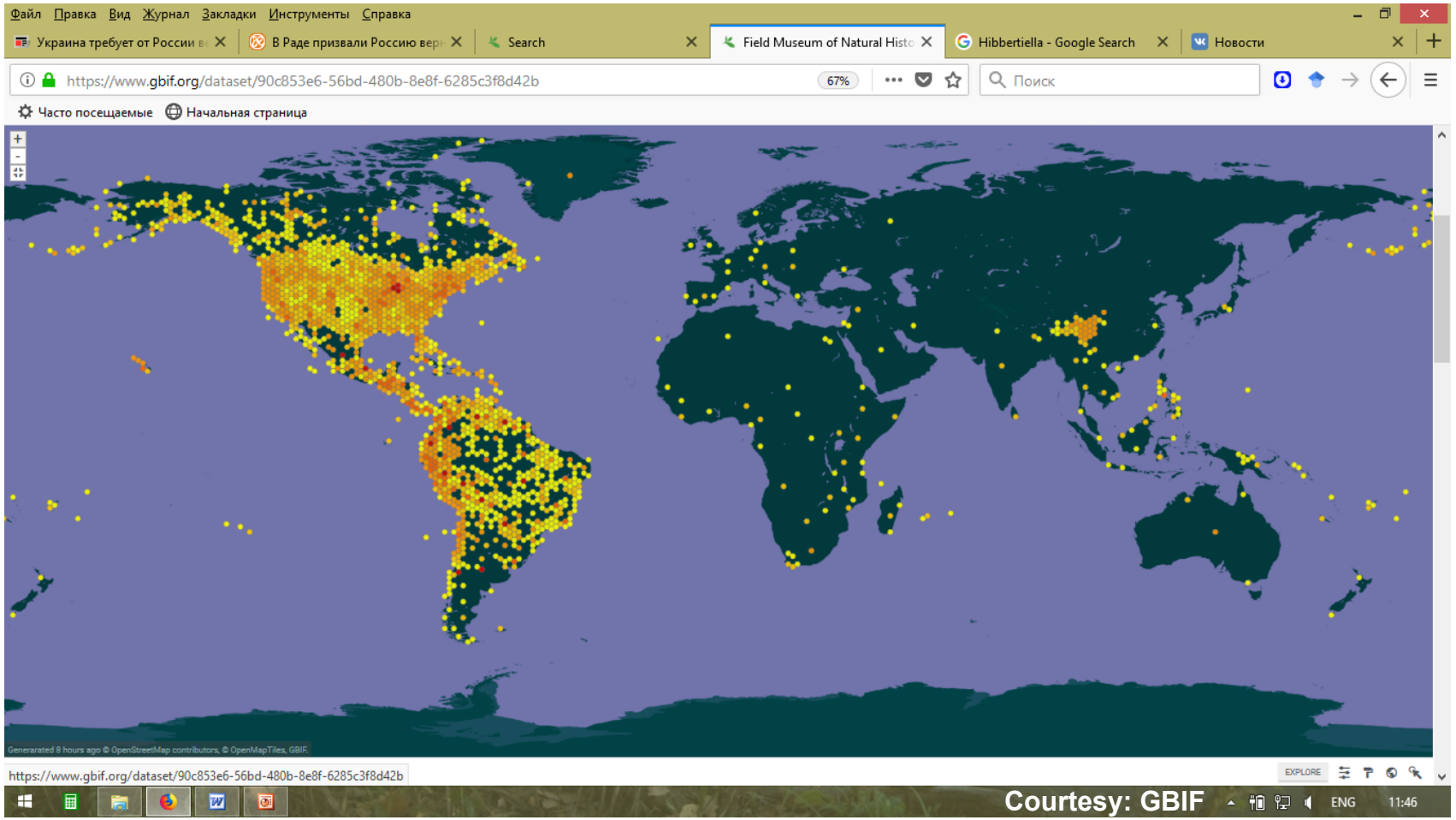
6. Lund Botanical Museum (LD)



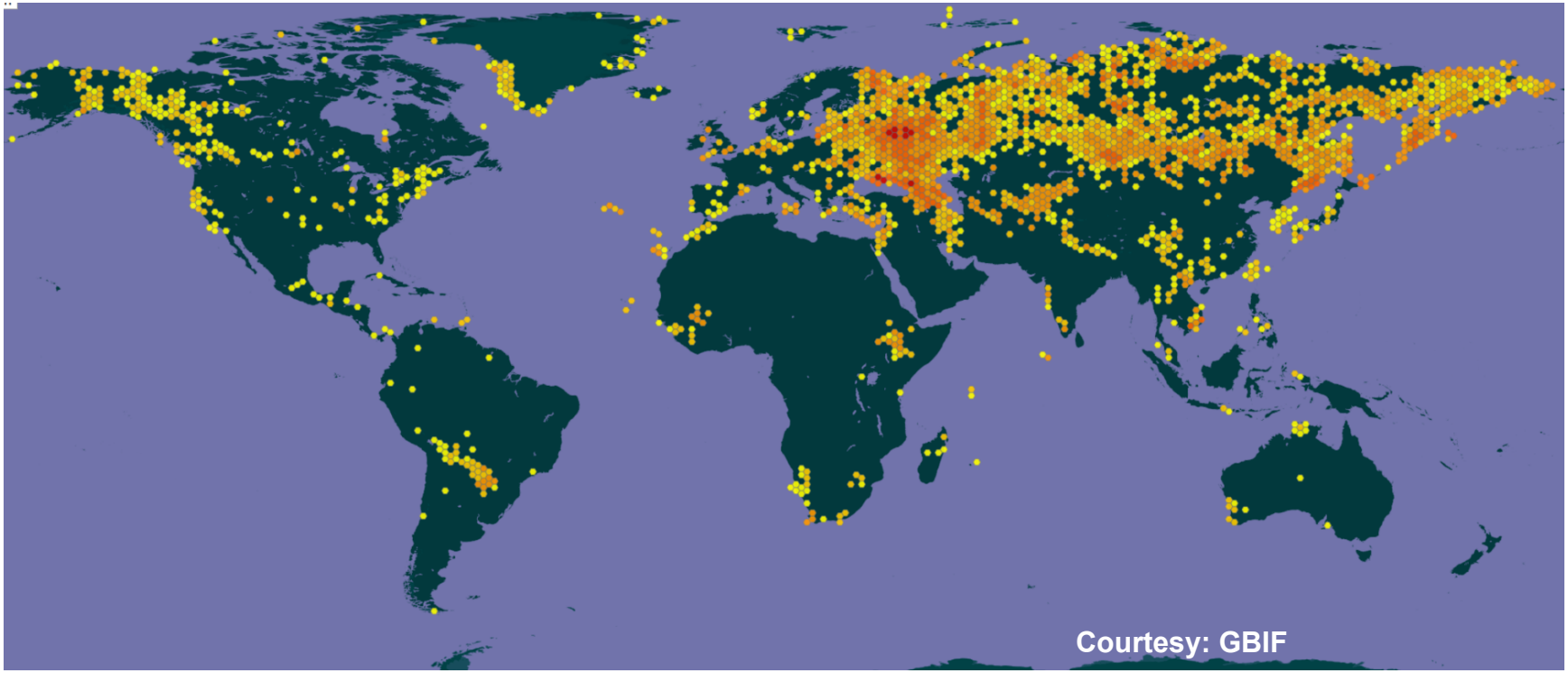
7. Vascular Plant Herbarium, Oslo (O)



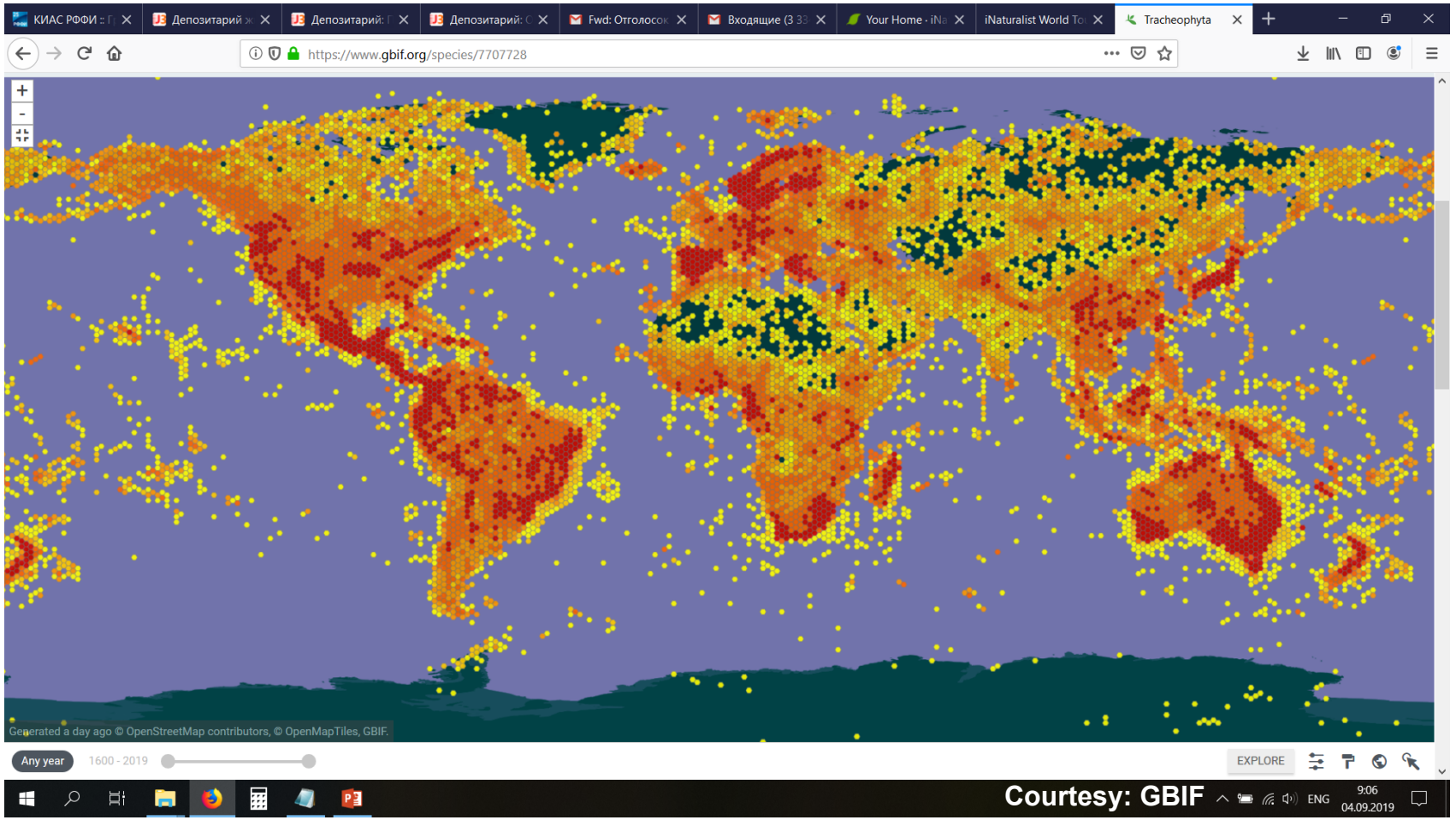
8. Field Museum of Natural History (Botany)



11. MW Dataset in GBIF: 412,062 georeferences



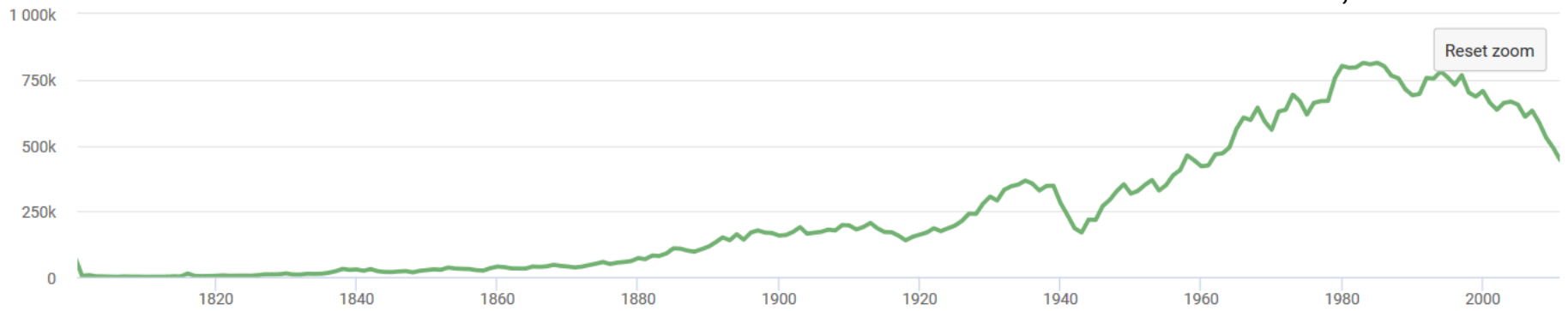
GBIF Map for Vascular Plant Specimens



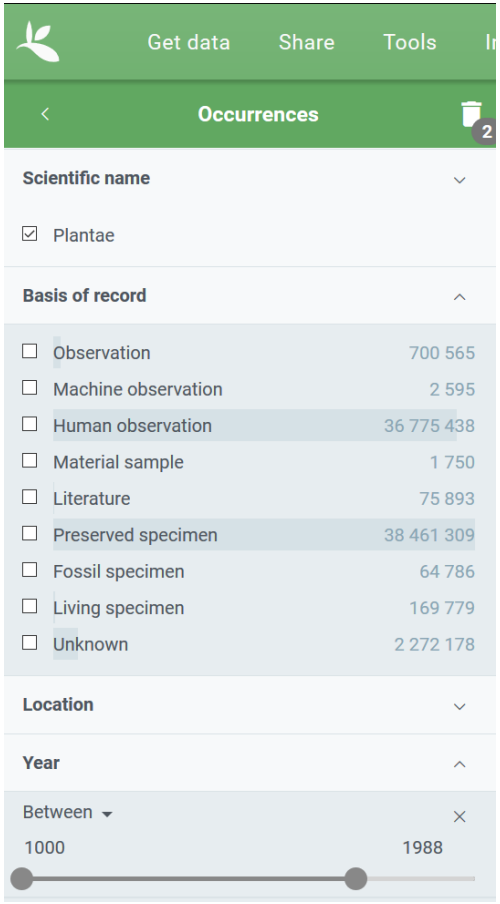
The Age of e-Records in GBIF

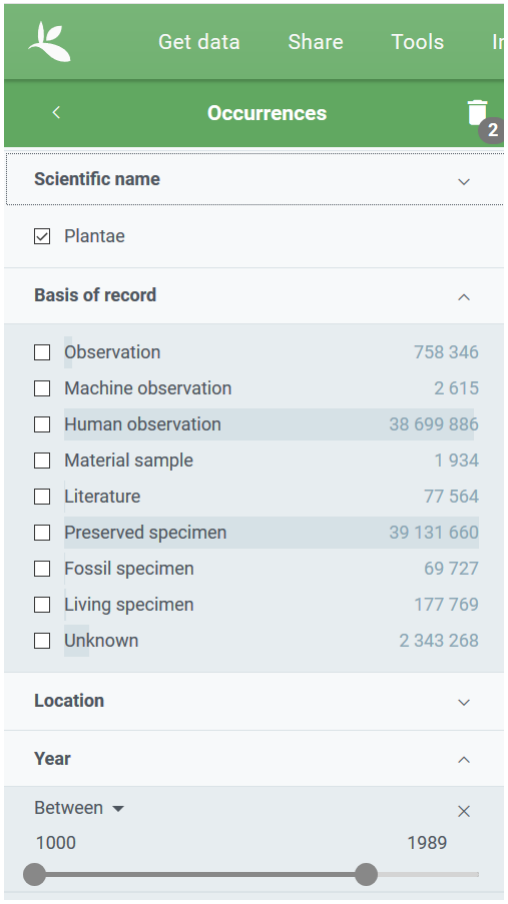
All records: max. 125,640,610 in 2018

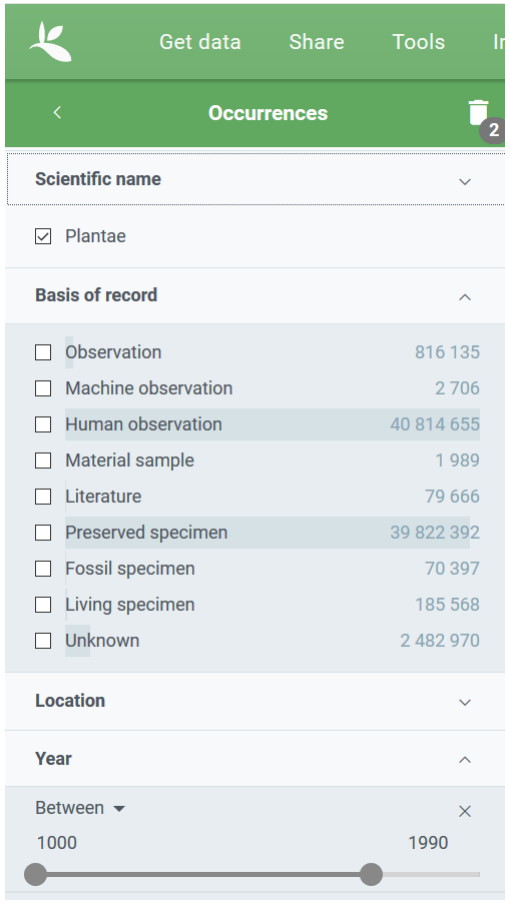
Herbarium records: max. 814,272 in 1985



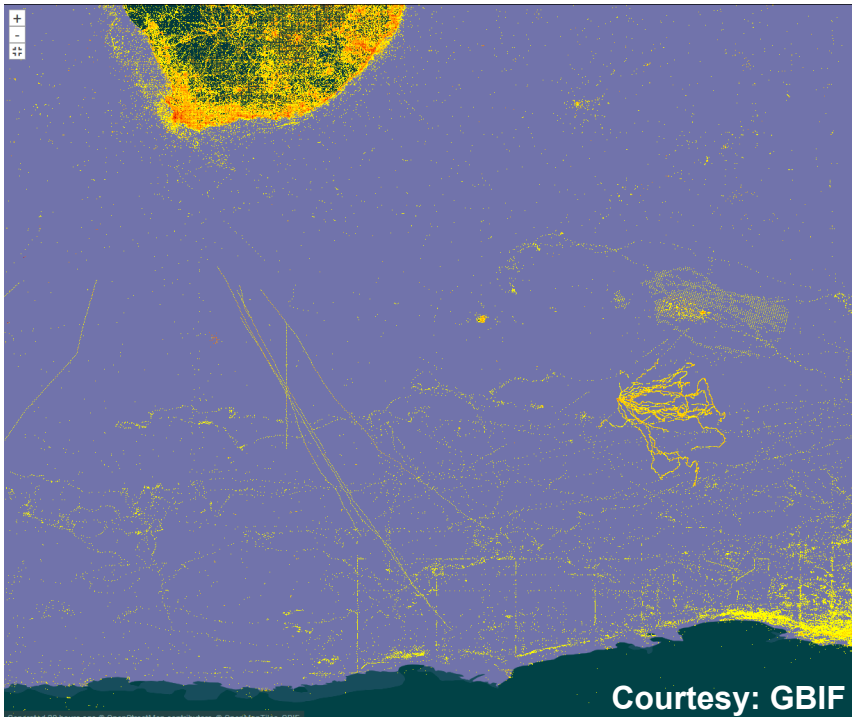
Plant e-Records in GBIF: 1988, 1989 & 1990



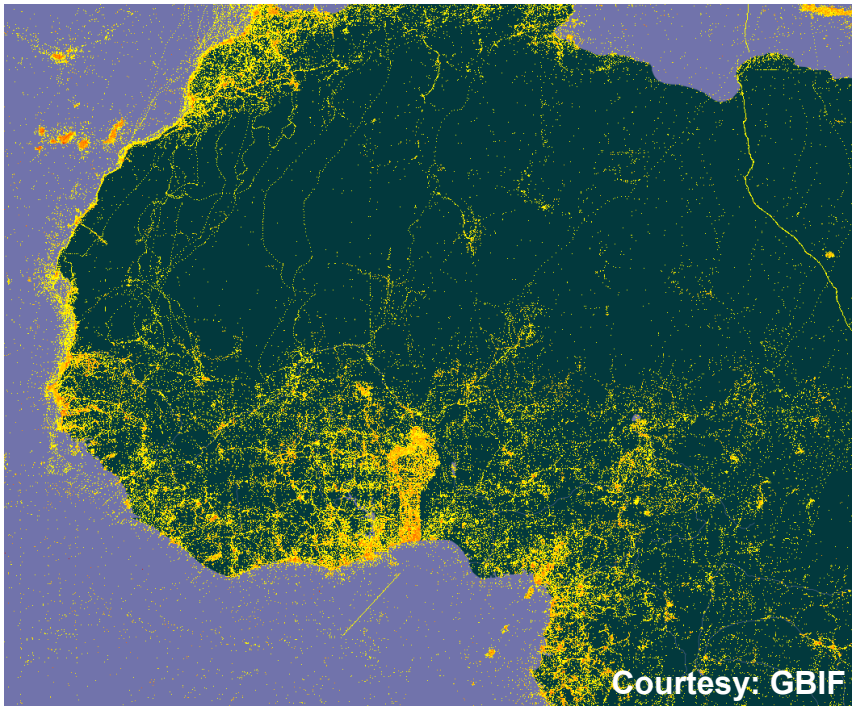




Why There Are So Many New Records? New technologies!

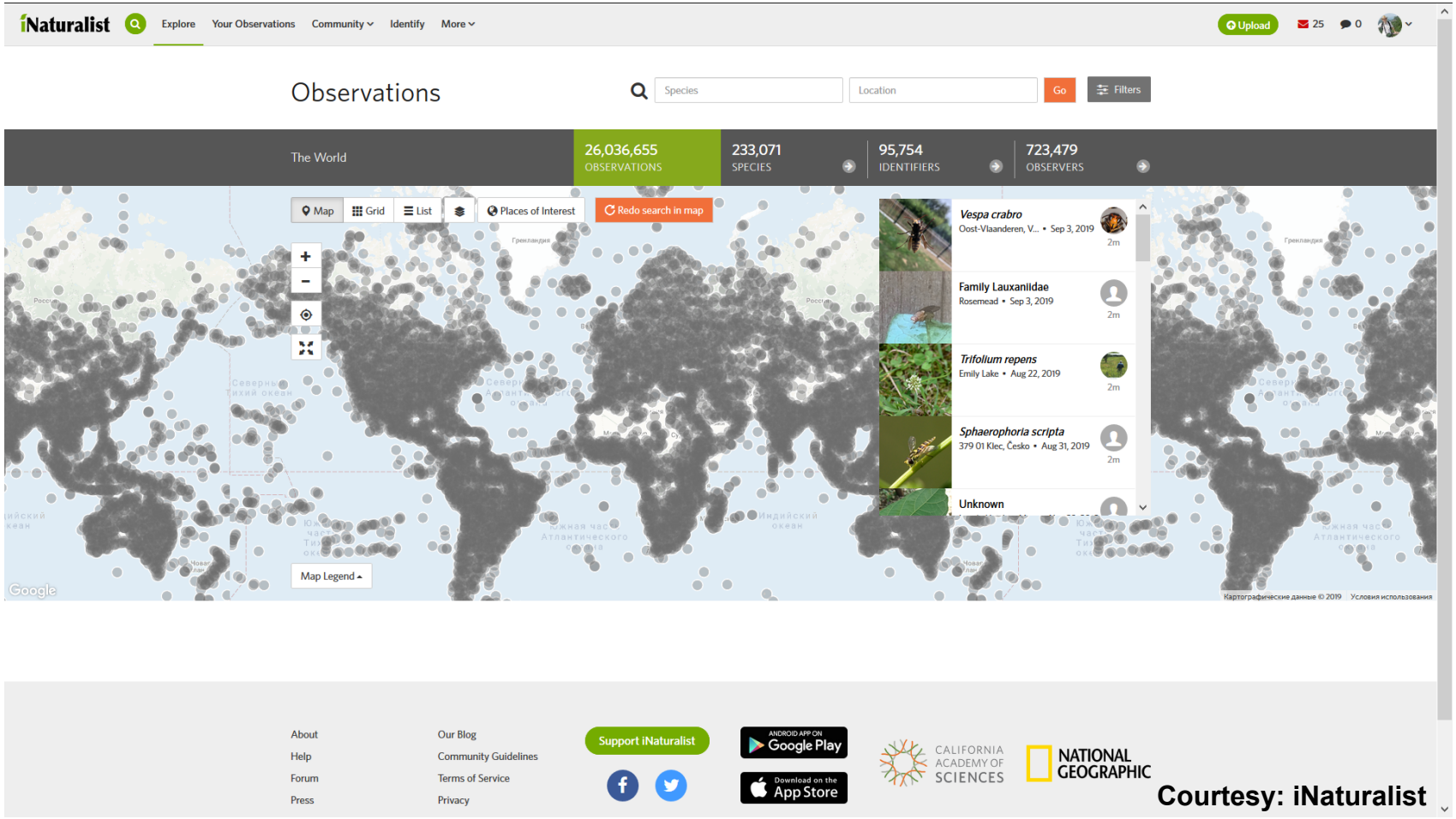


GPS trackers on king penguins and sea lions



GPS trackers on migrating birds

Why There Are So Many New Records? Citizen science!



Courtesy: iNaturalist

iNaturalist Is Popular Both in Russia and Italy!

Courtesy: GitHub

iNaturalist Country Rankings for Plants

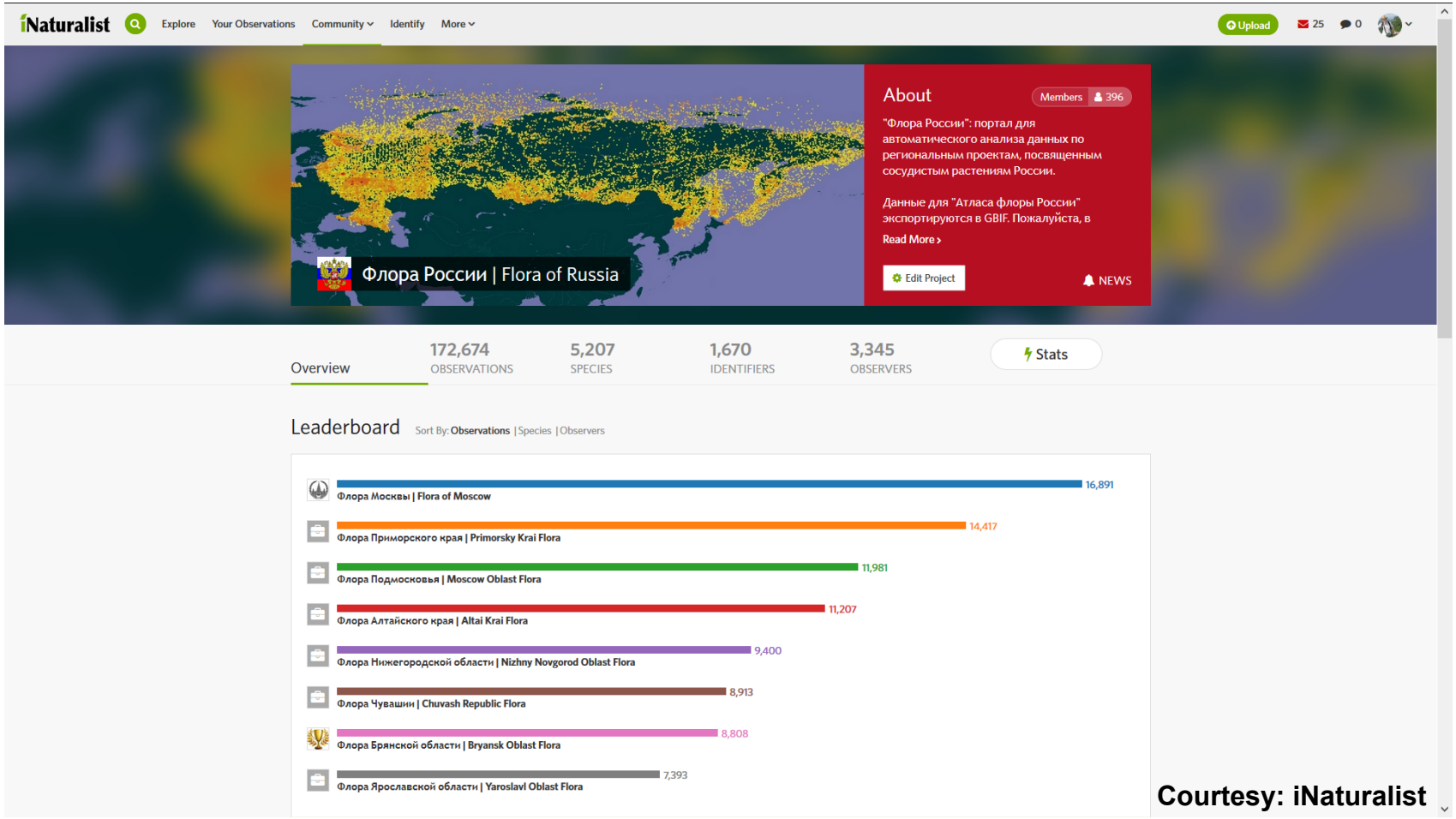
All records (including unidentified and unlicensed)

United States of America	6,293,954
Canada	716,059
Mexico	486,909
South Africa	341,706
New Zealand	274,899
United Kingdom	206,630
Russian Federation	**195,882**
Italy	**130,438**
Taiwan	113,452
Australia	104,031

Records in GBIF (identified, verified and freely licensed)

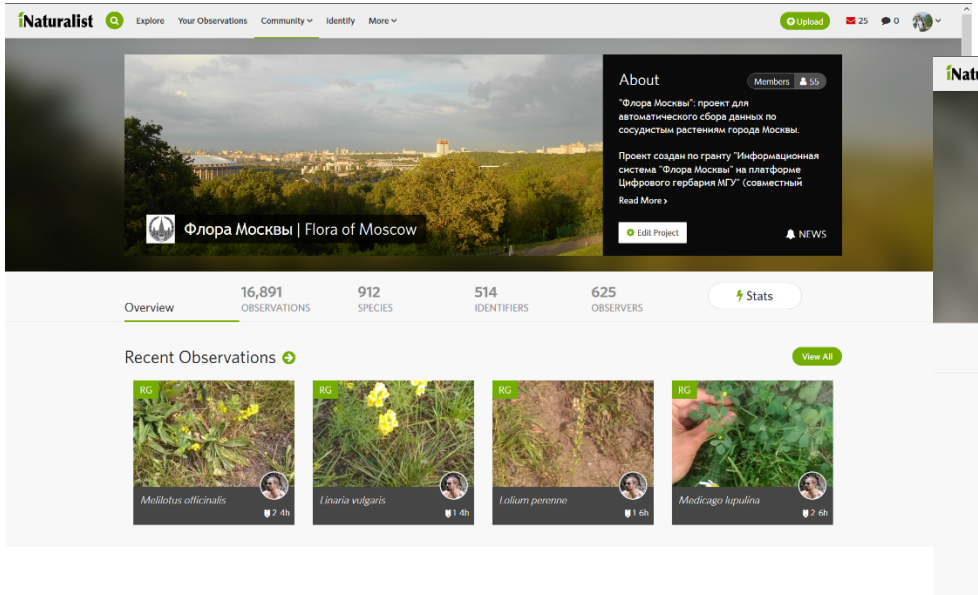
United States of America	2,262,798
Canada	293,885
New Zealand	184,462
Mexico	148,207
South Africa	138,320
Russian Federation	**126,136**
United Kingdom	67,320
Germany	47,009
Taiwan	46,896
Italy	**45,732**

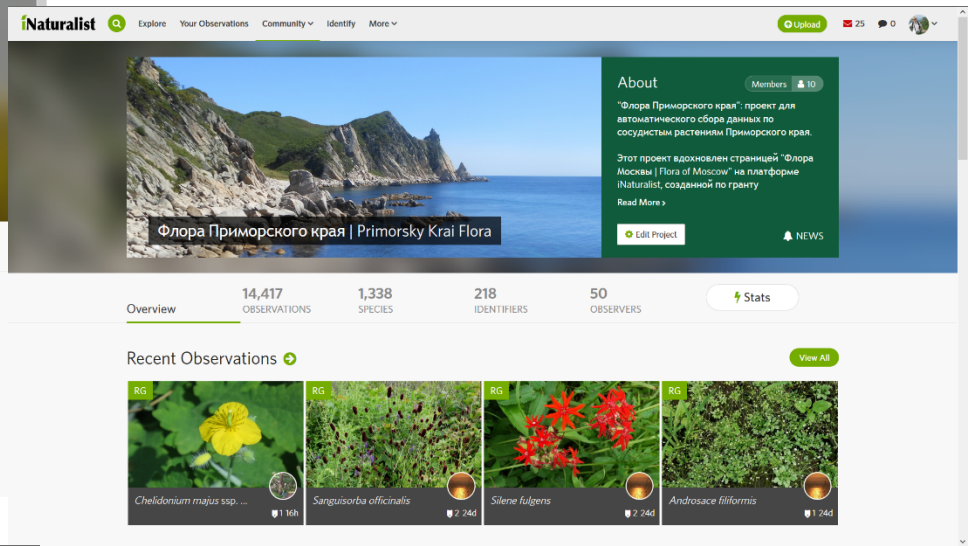
“Flora of Russia” on iNaturalist: 160,596 Records Added in 2019

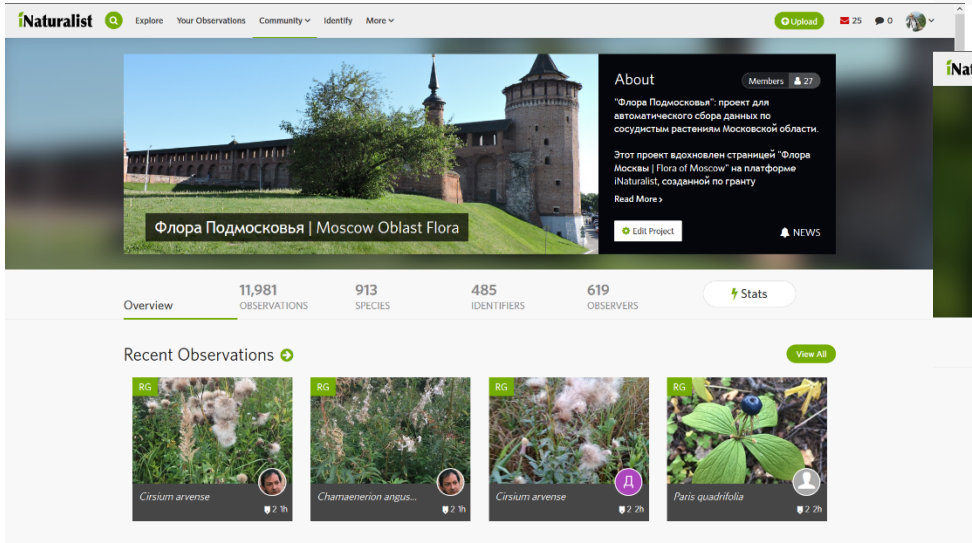


Courtesy: iNaturalist

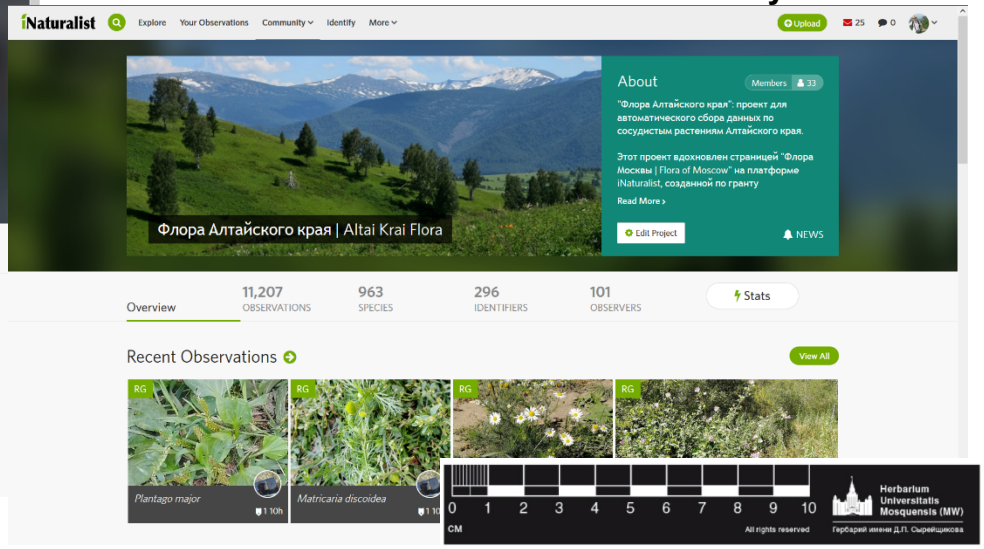
85 Regional Projects



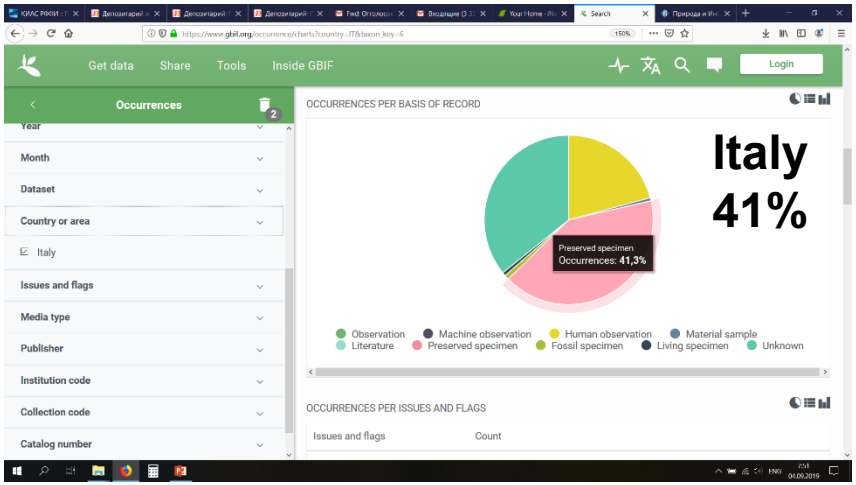


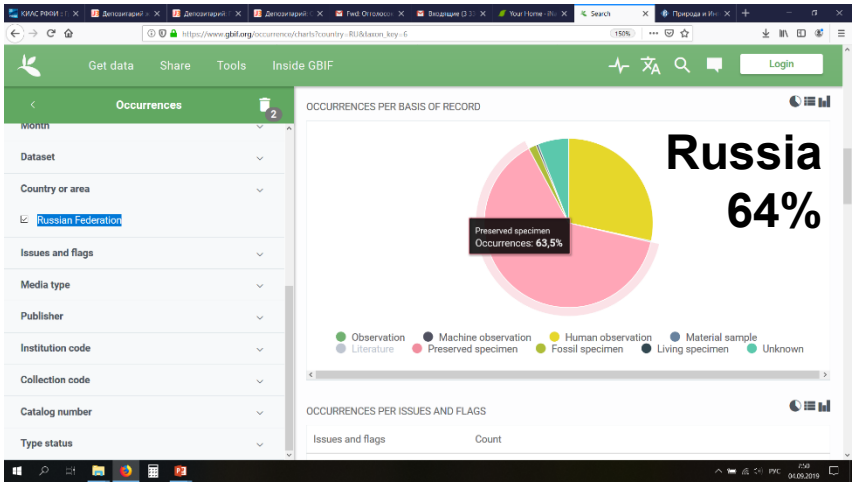


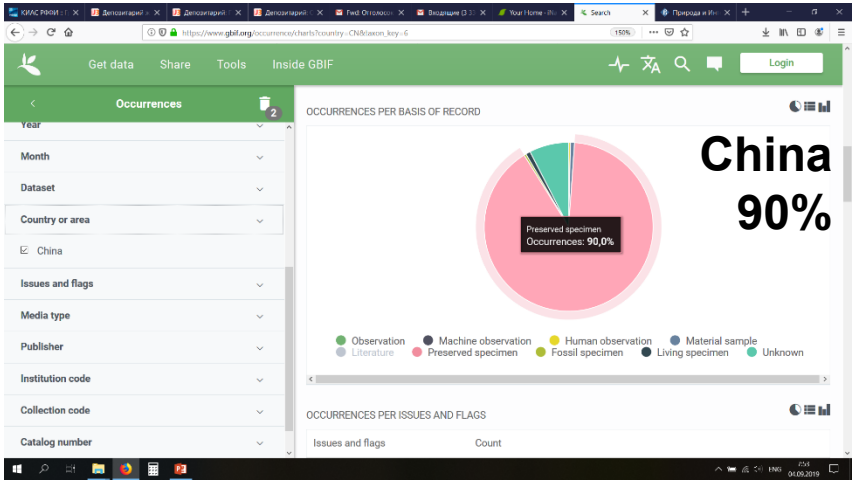
Courtesy: iNaturalist

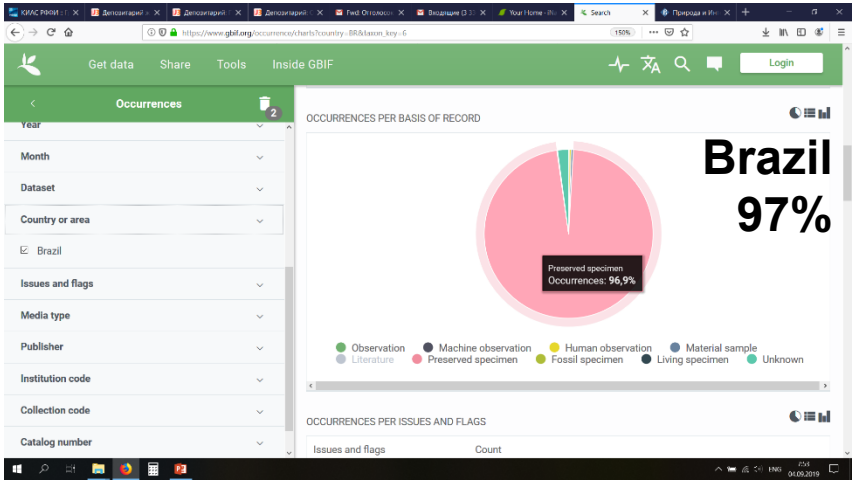


Herbarium Collections Are Still Vital









Courtesy: GBIF

Key Conclusion Notes

- World herbaria are growing constantly. Still!
- World herbaria are being rapidly digitized with current rate of digitization exceeding 22%.
- Herbaria are the major data source for plant e-records until 1989 in temperate zone and still in tropics.
- Herbarium collections are the only fully verifiable and universal source of plant records.
- Citizen science gives us tons of new data and we should start working with them.
- New recording techniques are complementing herbarium specimens, but not replacing them.

Main Collaborators

Moscow Digital Herbarium team

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
- В.И. Трошкина

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

"Flora of Russia" on iNaturalist team

Rank	User	Observations	Species
1	apseregin	11,568	1,339
2	snv2	9,772	752
3	velibortravoved	8,419	656
4	panasenkonn	5,464	866
5	eduard_garin	4,069	721
6	convallaria1128	4,042	1,240
7	dryomys	3,336	650
8	pavel_golyakov	2,957	716
9	vvolkotrub	2,760	1,245
10	phlomis_2019	2,738	1,044
11	svetlanakutueva	2,436	434
12	melodi_96	2,297	697
13	julia_shner	2,289	481
14	max_carabus	2,230	400
15	npz	2,197	391
16	vadim_prokhorov	1,972	760
17	katerina_kashirina	1,839	616
18	ramazan_murtazaliev	1,775	1,164
19	borisbolshakov	1,741	473
20	alzov	1,697	499
21	evgenyboginsky	1,634	326
22	sokolkov2002	1,569	508
23	aleksandrebel	1,535	591
24	beerolha	1,499	456
25	alexeiebel	1,438	487
26	vladimir_teplouhov	1,398	376
27	mallaliev	1,330	682
28	olga2019kuryakova	1,204	421
29	tomegatherion	1,171	466
30	denis_tishin	1,095	359

Thank you for being with me today!

Dr Sci Alexey P. Seregin

botanik.seregin@gmail.com