

Literature data digitization



IPAE
INSTITUTE OF PLANT
AND ANIMAL ECOLOGY

Artëm Sozontov

A.N.Sozontov@gmail.com
ipae.uran.ru/Sozontov_AN
<https://sozontov.cc/about>

Karaganda, 2025

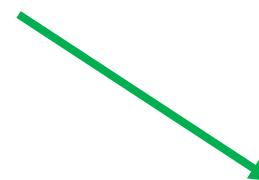
1. Breaking news! Zoologists kill to much



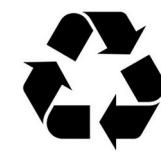
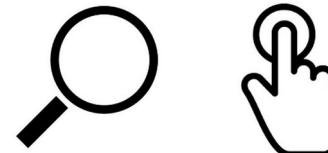
2. Primary data -> FAIR data -> Data reuse!

The primary data of biological species occurrences is a basis of numerous studies concerning the ecology & environment

Less deaths
More data
Fast access
Easy use
Do more science!



F_{indable} A_{ccessible} I_{nteroperable} R_{Reusable}



Open access to these data makes researches faster, wider, deeper



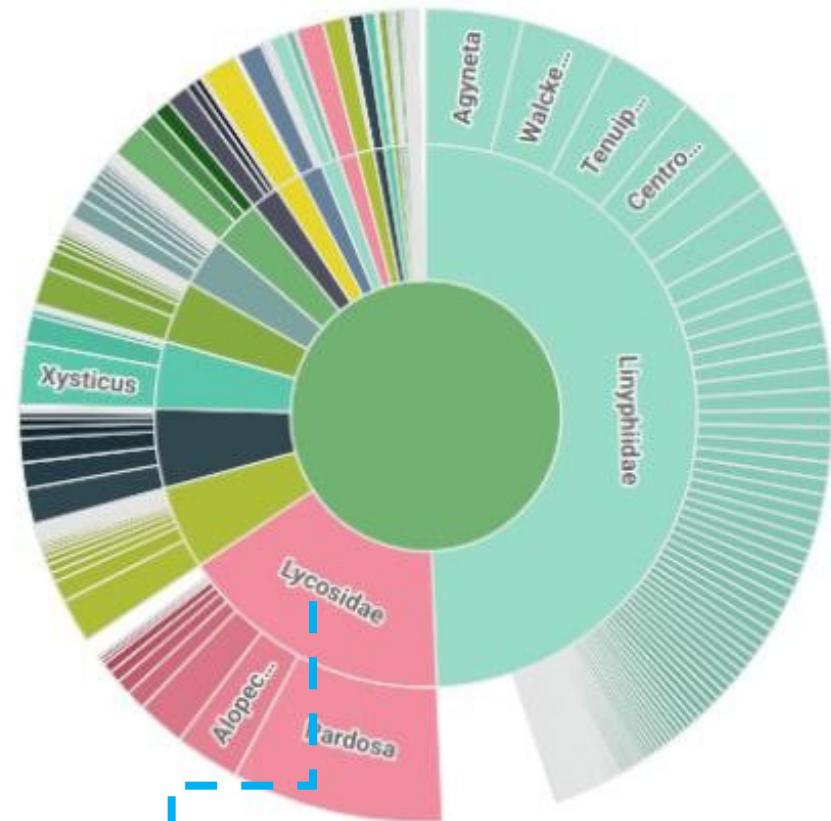
What about data reuse? Being once collected, specimens can provide data again and again

3. Data sources for SDM (and beyond)

Source	Trust score	Bias ¹	Machine-readable
Museum collections	Academician	More or less	Yes (sometimes)
Your own collectoins	Academician (hopefully)	-_(ツ)_/-	Yes (hopefully)
Academician open sources: GBIF	Academician	More or less	Yes
Volunteer open sources: iNaturalist	Amateur	Yes	Yes
Literature data	Academician	Less	No

¹ Bias spatial, taxonomical etc

4.1. Biased species composition



4.2. Biased species composition

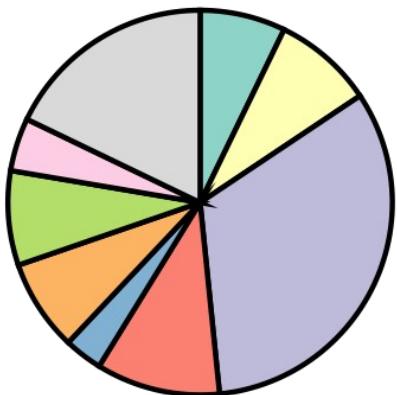
1. Literature



2. Visim Reserve



3. Udm Republic



4. iNaturalist



Family



5. Why can't we search across these data

still?

1. To know if the paper exists

AND

2. To obtain the full text

AND

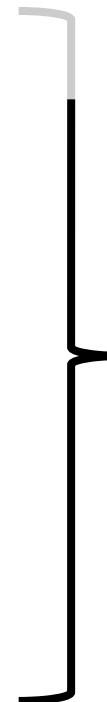
3. To get through language
barrier

AND

4. To uniform all data standards

AND

- - -
ive after this labor



Things get worse
due to exponential
growth
in number of
publications



6. Literature data access: easy!



7. Literature data access: easy... Or not?!

Species	LAT	LON	Records in Novosibirsk Area									
			CLUBIONIDAE									
<i>Clubiona phragmitis</i> * C.L. Koch, 1843	P	TP	56, 80									
<i>Clubiona stagnatilis</i> * Kulczyński, 1897	P	WP	36									
<i>Clubiona subtilis</i> L. Koch, 1867	SH	WP	36									
<i>CYBAEIDAE</i> — 1 genus, 1 species	SH	TP	27, 86									
(Clerck, 1758)												
<i>DICTYNIDAE</i> — 7 genera, 13 species	P	WP	106									
Argyrono-												
Archae-												
1871)												

Биотопическое распределение видов пауков с.н. *Gnaphosidae* с о. Биотопическая шкала относительного обилия видов, предложенная Ю.А. Песенко [в аранжировке] (пятибалльная шкала отрицательного обилия видов, предложенная Ю.А. Песенко [в аранжировке])

Виды	Группы местообитаний										
	Степи					Сосновые леса					
Номера местообитаний	1	2	3	4	5	6	7	8	9	10	11
Приуроченные к мозаичным стациям											
<i>Drossodes longispinus</i> Marusik and Logunov, 1995											
<i>Drossodes chrysodenoides</i> Eysunin and Tuneva, 2002											
<i>Gnaphosa mongolica</i> Simon, 1895											
<i>Zelotes rufi</i> Etymik, 1996											
<i>Drossodes rotundatus</i> Eysunin and Tuneva, 2002											
<i>Drossyllus</i> sp. Tuneva and Eysunin, 2003											
<i>Gnaphosa pilosa</i> Saveljeva, 1972											
<i>Gnaphosa steppica</i> Ovtsharskii, Platnick and Song, 1992											
<i>Zelotes mithailovi</i> Marusik, 1995											
<i>Zelotes arenarius</i> Tuneva and Eysunin, 2003											
<i>Zelotes segregus</i> Simon, 1878											
<i>Berlandina cincta</i> Mengen, 1872											
<i>Crozeutes pygmaeus</i> (Miller, 1945)											
<i>Graphosa lugufa</i> (Walckenaer, 1802)											
<i>Gnaphosa saurica</i> Ovtsharskii, Platnick and Song, 1992											
<i>Micaria russica</i> Thorac, 1875											
<i>Zelotes atrocaneus</i> (Simon, 1878)											

Биотопическая шкала относительного обилия видов, предложенная Ю.А. Песенко [в аранжировке]

Виды	Группы местообитаний										
	Степи					Сосновые леса					
Номера местообитаний	1	2	3	4	5	6	7	8	9	10	11
Приуроченные к мозаичным стациям											
<i>Drossodes longispinus</i> Marusik and Logunov, 1995											
<i>Drossodes chrysodenoides</i> Eysunin and Tuneva, 2002											
<i>Gnaphosa mongolica</i> Simon, 1895											
<i>Zelotes rufi</i> Etymik, 1996											
<i>Drossodes rotundatus</i> Eysunin and Tuneva, 2002											
<i>Drossyllus</i> sp. Tuneva and Eysunin, 2003											
<i>Gnaphosa pilosa</i> Saveljeva, 1972											
<i>Zelotes mithailovi</i> Marusik, 1995											
<i>Zelotes arenarius</i> Tuneva and Eysunin, 2003											
<i>Zelotes segregus</i> Simon, 1878											
<i>Berlandina cincta</i> Mengen, 1872											
<i>Crozeutes pygmaeus</i> (Miller, 1945)											
<i>Graphosa lugufa</i> (Walckenaer, 1802)											
<i>Gnaphosa saurica</i> Ovtsharskii, Platnick and Song, 1992											
<i>Micaria russica</i> Thorac, 1875											
<i>Zelotes atrocaneus</i> (Simon, 1878)											

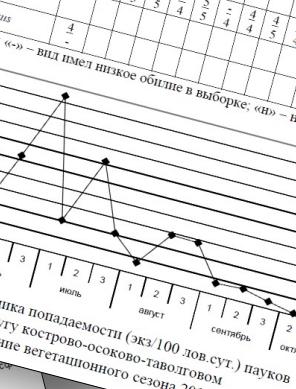
Биотопическая шкала относительного обилия видов, предложенная Ю.А. Песенко [в аранжировке]

Виды	Группы местообитаний										
	Степи					Сосновые леса					
Номера местообитаний	1	2	3	4	5	6	7	8	9	10	11
Приуроченные к мозаичным стациям											
<i>Drossodes longispinus</i> Marusik and Logunov, 1995											
<i>Drossodes chrysodenoides</i> Eysunin and Tuneva, 2002											
<i>Gnaphosa mongolica</i> Simon, 1895											
<i>Zelotes rufi</i> Etymik, 1996											
<i>Drossodes rotundatus</i> Eysunin and Tuneva, 2002											
<i>Drossyllus</i> sp. Tuneva and Eysunin, 2003											
<i>Gnaphosa pilosa</i> Saveljeva, 1972											
<i>Zelotes mithailovi</i> Marusik, 1995											
<i>Zelotes arenarius</i> Tuneva and Eysunin, 2003											
<i>Zelotes segregus</i> Simon, 1878											
<i>Berlandina cincta</i> Mengen, 1872											
<i>Crozeutes pygmaeus</i> (Miller, 1945)											
<i>Graphosa lugufa</i> (Walckenaer, 1802)											
<i>Gnaphosa saurica</i> Ovtsharskii, Platnick and Song, 1992											
<i>Micaria russica</i> Thorac, 1875											
<i>Zelotes atrocaneus</i> (Simon, 1878)											

Биотопическая шкала относительного обилия видов, предложенная Ю.А. Песенко [в аранжировке]

Таблица 2
на лугу кострово-осоково-таволговом в 2010 (под чертой) и 2011 гг.

Виды	(под чертой)									
	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Май	Июнь	Июль	Август
<i>Allomengea scopulera</i>	3	1	2	3	1	2	3	1	2	3
<i>Allomengea vidua</i>										
<i>Pardosa fulvipes</i>										
<i>Pardosa plumipes</i>										
<i>Pardosa prativaga</i>										
<i>Pelecopsis menetriesii</i>										
<i>Alytus viduus</i>										



	Номер	Виды	Материал
1		<i>Linyphantes</i> (trid.).	Из этого представителя у него более половины особей, обитающих в Тарасовской роще.
2		<i>Clubionidae</i>	Коллективные гнездовые, обитающие в Тарасовской роще.
3		<i>Clubionidae</i> (Tarsius)	Из этого представителя Шишков имеет значения от 2,3 до 2,7, что соответствует с высоким разнообразием птиц, группирующихся в нем.
4		<i>Clubionidae</i> (Tarsius)	Самые высокие значения от 2,3 до 2,7, что соответствует с высоким разнообразием птиц, группирующихся в нем.

303

	Clad.	Материал
1	<i>Clubionidae</i>	2 ♀♀, [2], bank of former river-bed, 12-20.VI.2012.
2	<i>Clubionidae</i>	2 ♀♀, [2], bank of former river-bed, flood-plain 10.VII.2012; 1 ♂, 3 ♀♀, [4], banks of former river-bed, 11.VI.2012.

	Clad.	Материал
1	<i>Clubionidae</i>	1 ♀, [3], flood-plain of Kama River, 4-10.VII.2011.
2	<i>Clubionidae</i>	1 ♀, [3], flood-plain of Kama River, 4-10.VII.2011.

	CORINIIDAE	Phrurolithus festivus
1		MATERIAL 1 ♂, 1 ♀, [1], immoderately herbaceous, pitfall-traps, 1.VI-2012; 1 ♂, 1 ♀, 1 subadult ♀, [2], Populus tremula-Tilia forest with oak and Acer, 14.VI.2012; 1 ♂, [3], Populus tremula-Tilia forest with oak and Acer, 11-20.VI.2012; 7 ♀♀, [3], dry and sunny bank of former river-bed, 2-5.VII.2012.
2		MATERIAL 2 ♀♀, [2], bank of former river-bed, 12-20.VI.2012.

DICTYNIDAE

7.1. Oh not... is it worth it?

CORINIIDAE

Phrurolithus festivus (C.L. Koch, 1835)

MATERIAL. 1 ♂, 1 ♀, [1], inundated *Quercus*-forest, pitfall-traps, 12–20.VI.2012; 1 ♀, 1 subadult ♀, [2], *Populus tremula-Tilia* forest with oak and *Acer*, 14.VI.2012; 1 ♀, [3], *Populus tremula-Tilia* with oak and *Acer*, 11–20.VI.2012; 7 ♀♀, [3], dry and sunny bank of former river-bed, 2–5.VIII.2012.

7.2. Oh not... is it worth it?

Таблица 2

Доминантный комплекс (значения индекса Песенко) пауков
на лугу кострово-осоково-таволговом в 2010 (над чертой) и 2011 гг.
(под чертой)

Виды	Месяцы и декады													
	Май		Июнь			Июль			Август			Сентябрь		
	3	1	2	3	1	2	3	1	2	3	1	2	3	2
<i>Allomengea scopigera</i>											<u>4</u> 4	<u>4</u> -	<u>4</u> -	<u>4</u> -
<i>Allomengea vidua</i>												<u>4</u> -	<u>5</u> -	<u>5</u> -
<i>Pardosa fulvipes</i>					<u>4</u> 4	<u>4</u> 5	<u>4</u> 4	<u>-</u> 4						
<i>Pardosa plumipes</i>							<u>-</u> 4							
<i>Pardosa prativaga</i>	<u>5</u> 4	<u>5</u> 5	<u>5</u> 5	<u>5</u> 5	<u>5</u> 4	<u>5</u> 5	<u>-</u> 4	<u>4</u> 4	<u>4</u> 5	<u>-</u> 4				
<i>Pelecopsis mengei</i>														<u>4</u> -
<i>Xysticus viduus</i>	<u>4</u> -													

Примечание: «-» – вид имел низкое обилие в выборке; «н» – нет данных.

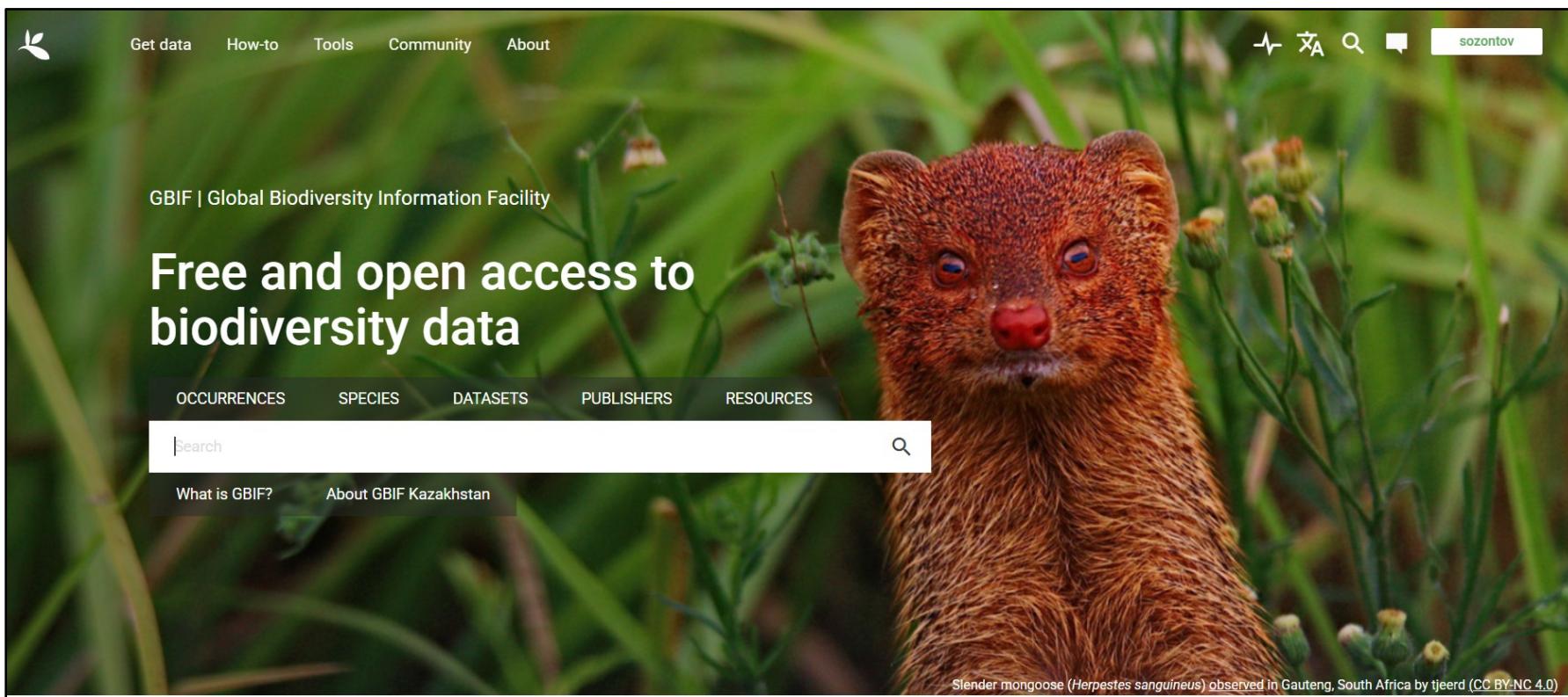
7.3. Oh not... is it worth it?

Таблица 1

Видовой состав, плотность (экз./м²) и показатели разнообразия населения стратифицированных пауков лесных биотопов национального парка «Припышминские боры», летний аспект

Таксон	Номер биотопа*						
	1	2	3	4	5	6	7
Araneidae							
<i>Araneus</i> sp. (неполовозрелые)	0.6	0.5					
Clubionidae							
<i>Clubiona</i> sp. (неполовозрелые)			1.1				
Dictynidae							
<i>Emblyna burjatica</i> (Danilov, 1994)							4.0
Gnaphosidae							
<i>Haplodrassus silvestris</i> (Blackwall, 1833)	0.6						
<i>H. soerrenseni</i> (Strand, 1900)	0.6				3.2		
<i>H. umbratilis</i> (L. Koch, 1866)	1.7						
<i>Haplodrassus</i> sp. (неполовозрелые)	12.6		2.3				
<i>Zelotes clivicola</i> (L. Koch, 1870)	0.6						
<i>Zelotes</i> sp. (неполовозрелые)	2.9						

8. GBIF current data (amount)



The screenshot shows the GBIF homepage with a large banner image of a slender mongoose. The page includes a navigation bar with links for 'Get data', 'How-to', 'Tools', 'Community', and 'About'. A search bar is present, along with a user profile for 'sozontov'. Below the banner, the text 'Free and open access to biodiversity data' is displayed. A red callout points to the first data statistic: '3 160 204 849 Occurrence records'.

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search

What is GBIF? About GBIF Kazakhstan

Slender mongoose (*Herpestes sanguineus*) observed in Gauteng, South Africa by tjeerd (CC BY-NC 4.0)

7 271 610 of them (2%) are spiders!

3 160 204 849 Occurrence records	116 293 Datasets	2 545 Publishing institutions	13 324 Peer-reviewed papers using data
-------------------------------------	---------------------	----------------------------------	---

9. A little part of literature data (estimation)

References	Occurrences	Individuals
Танасевич А.В. 1985. К изучению пауков (Aranei) Полярного Урала // Фауна и экол. пауков СССР. Тр. Зоол. ин-та АН СССР 139: 52–62.	100	156
Есюнин С.Л., Новокшенов В.Г. 1992. Интересные находки пауков (Aranei) из Юганского заповедника // Тр. Зоол. ин-та АН СССР 226: 115–117.	14	26
Esyunin S.L., Kazantsev D.K. 2008. Spider fauna of the Pechoro-Ilychskiy Reserve, with the description of a new Agroeca species (Liocranidae) // Arthropoda Selecta 16: 245–250.	75	313
Тунёва Т.К., Есюнин С.Л. 2012. К фауне пауков (Aranei) Леоновских гор (Челябинская область) // Евразиатский энтомологический журнал 11: 373–377.	118	422
Mikhailov K.G., Trushina E.E. 2013. On the spider fauna (Arachnida: Aranei) of the Mordovian State Reserve, Russia: preliminary results // Arthropoda Selecta 22: 189–196.	441	1701
Sozontov A.N., Esyunin S.L. 2014. On the spider fauna (Arachnida: Aranei) of the 'Ust'-Belsk' Natural Park and its vicinities // Arthropoda Selecta 23: 301–310.	304	1220
Azarkina G.N., Lyubechanskii I.I., Trilikauskas L.A. et al. 2018. A check-list and zoogeographic analysis of the spider fauna of Novosibirsk Area // Arthropoda Selecta 27: 73–93.	364	916
Fomichev A.A., Ballarin F., Marusik Yu.M. 2022. A new genus of the family Nesticidae (Arachnida: Aranei) from the Caucasus // Arthropoda Selecta 31: 99–110.	5	45
Total	1 421	4 779
Per one reference	178	600
Estimation (5 300 references)	943 400	3 180 000

10. Literature data liberation: requirements (part 1)



Research Ideas and Outcomes 10: e126586
doi: [10.3897/rio.10.e126586](https://doi.org/10.3897/rio.10.e126586)



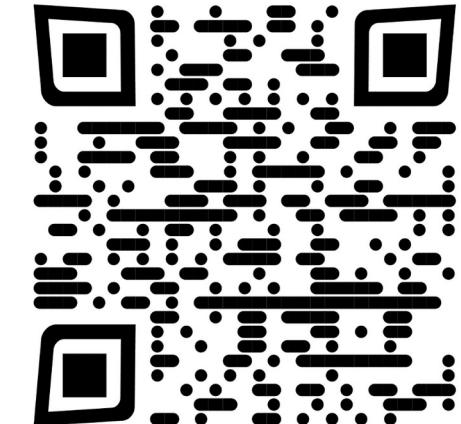
Policy Brief

Liberate the power of biodiversity literature as FAIR digital objects

Donat Agosti[‡], Laurence Bénichou[§], Ana Casino[¶], Lars Holm Nielsen[¶], Patrick Ruch^{#,¤},
Puneet Kishor[‡], Lyubomir Penev[“], Patricia Mergen[”], Christos Arvanitidis^χ

1. For biodiversity publishers and literature aggregators

- All publications should be **open access** and no restrictions for data mining should apply;
- All publications that are not available online **should be uploaded** to BLR;
- All publications should be made available in a **machine format (e.g. JATS XML)**;



Agosti D. et al., 2024. Liberate the power of biodiversity literature as FAIR digital objects

11. Literature data liberation: requirements (part 2)

2. For prospective publications

- Open Access publishing must become a norm to support open science [...] under a CC BY licence or waiving copyright (CC0);
- Alternative workflows to produce XML-first based structured publications should be developed to cover differing needs of stakeholders;
- PIDs should be assigned to most important sub-article structural metadata and research objects and embedded in the article XMLs, to facilitate machine to machine interaction and save authors time to retrieve information;

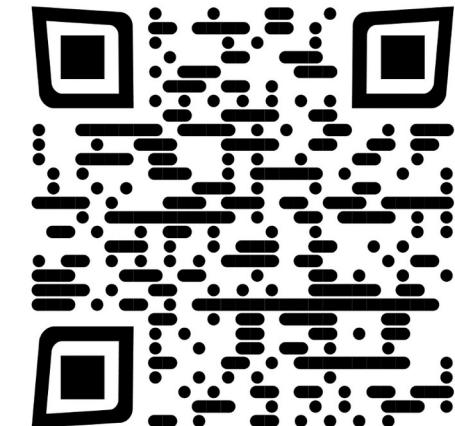
...

3. For legacy publications

- Digitisation of articles [...] Optical Character Recognition (OCR);
- Publications digitised retroactively must comply with FAIR standards [...];
 - [...] To convert publications into open digitally accessible knowledge has to be defined and added to describe the domain-specific semantic content of biodiversity works.

...

4. For integration into sustainable Infrastructures and services



Agosti D. et al., 2024. Liberate the power of biodiversity literature as FAIR digital objects

12. Literature data: just basic requirements (occurrences)

Google
sheets?
Excel?

No way!



1. Strict accordance between records and references
2. Data structure & standards
 - 2.1. Darwin Core
 - 2.2. Check for internal data conflicts
 - 2.3. Check for mistypes
 - 2.4. Show the user what to fix
3. Platform independence
4. Optimization of the repetitive data entry
5. Multi-language

13. Pensoft: data papers dissolve boundaries



**Journals published
by Pensoft have
their own hosted
data portals on GBIF**

for FAIR biodiversity data

14.1. Plazi. Focus on taxonomy

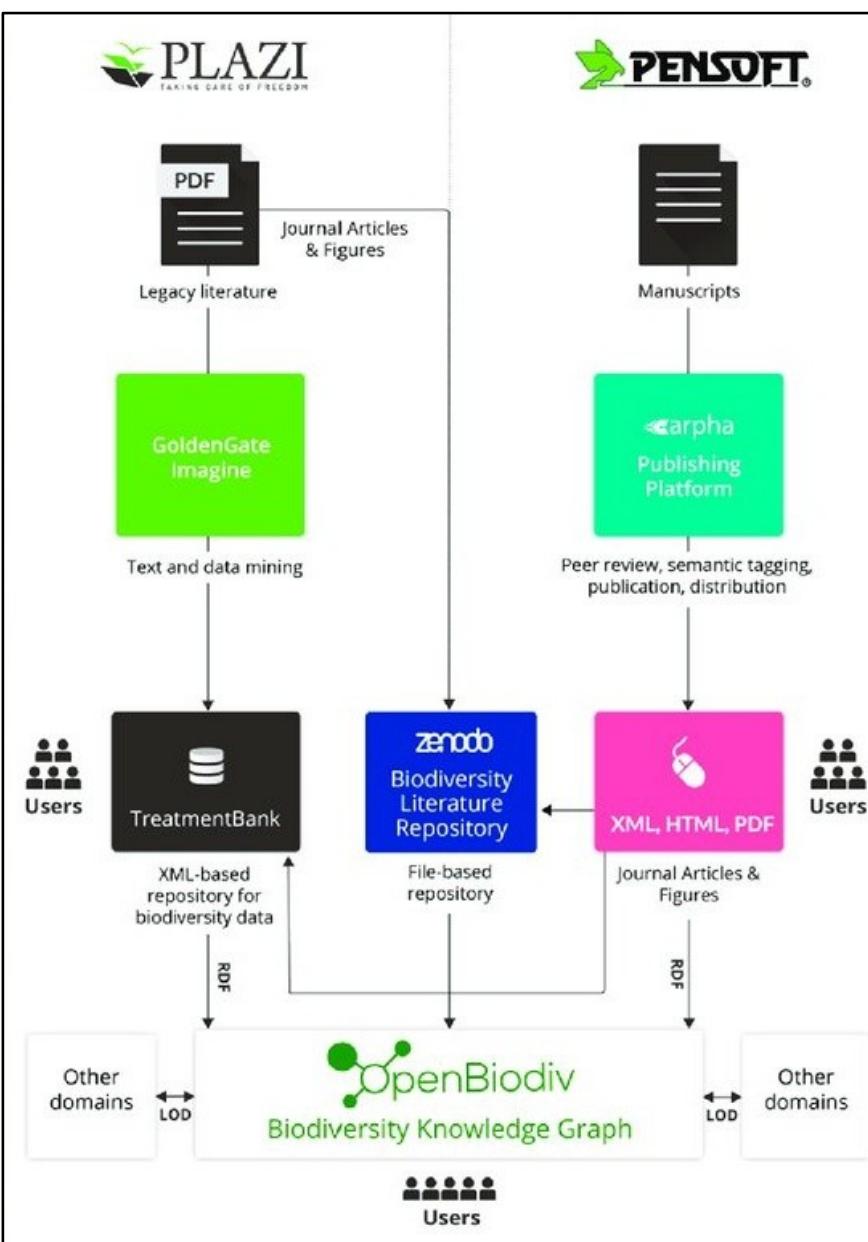


Table 10. Distribution of characters scored for ten *Onomastus* taxa and ontogroups

	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Onomastus quinquepunctatus</i>	✓	—	0	1	✓	✓	1	1	1	✓	1	1	✓
<i>Onomastus nigromaculatus</i>	✓	—	0	1	✓	✓	1	1	1	2	1	1	✓
<i>Onomastus nigricauda</i>	—	—	0	1	✓	✓	1	1	1	—	1	1	1
<i>Onomastus pallidulus</i>	✓	—	0	1	✓	1	1	1	1	1	1	0	1
<i>Onomastus indra</i> sp. nov.	✓	—	0	1	✓	1	1	1	1	1	1	0	1
<i>Onomastus puthiyagodai</i> sp. nov.	—	—	0	1	✓	0	1	1	1	1	1	0	1
<i>Onomastus rufolensis</i> sp. nov.	✓	—	0	1	✓	0	1	1	1	1	0	1	1
<i>Onomastus complexipalpis</i>	—	—	0	1	✓	1	1	1	1	1	2	1	1
<i>Onomastus kawai</i>	—	—	0	1	✓	1	1	1	1	1	2	1	1
<i>Onomastus katherinae</i> sp. nov.	✓	—	0	1	✓	0	1	1	1	1	2	1	1
<i>Pardisus saurii</i>	1	0	0	1	0	0	0	0	0	—	—	—	—
<i>Asemonea tenuipes</i> ?	1	0	1	0	—	✓	0	0	0	—	—	0	—
<i>Goleba puella</i> ?	1	1	0	0	✓	0	0	0	0	—	—	0	—

Introduction: The spider genus *Nesticus* Thorell, 1869 (family Nesticidae) is taxonomically diverse in the southern Appalachian mountains of eastern North America, with at least 30 species distributed over a geographic area extending from southern West Virginia to central Alabama (Gertsch 1984, Coyle & McGarity 1992, Hedin 1997a). Appalachian *Nesticus* are habitat specialists, reflecting apparently strict physiological constraints that limit these spiders to cool, moist microhabitats. These constraints, in combination with habitat discontinuity in both space and over time, have promoted tremendous species diversification and endemism (see speciation models of Wiens 2004a, 2004b). This fine-scale diversification

Accepted by P. Jäger, 23 Feb. 2005; published 18 Mar. 2005

Mark Page Header

- Hide Page
- Annotate >
- Annotate All >
- Mark Region
- Mark Table
- Mark Caption
- Mark Footnote
- Mark Page Header**
- Mark Parenthesis
- Mark Artifact
- Mark Image
- Merge Words
- Make Stream

14.2. Plazi. Focus on taxonomy

plazi.org

PLAZI taking care of freedom

TREATMENTBANK BLR SERVICES PARTICIPATE IMAGE SEARCH LIT SEARCH ABOUT

LATEST TREATMENTS

Strengthening FAIR Biodiversity Data Engagement in South America

Apr 13, 2025

Plazi delivered two courses for the scientific community in Bogotá [more](#)

STATS

Articles: 85773

Treatments: 1122544

Treatments with Mat. Citations: 410423

Material Citations: 1964503

Georeferenced Mat. Citations: 597015

ABOUT (8) BATLIT (1)

BHL SIBILS (1) BICIKL (3)

15.1. Examples: fungi of West Siberia

The fungal literature-based occurrence database for West Siberia (Russia)

N. Filippova, D. Ageev, S. Arefyev, S. Bolshakov, T. Bulyonkova, O. Vayshlya, A. Vlasenko, V. Vlasenko, S. Gashkov, I. Gorbunova, E. Davydov, A. Dobrynina, E. Zvyagina, V. Kapitonov, A. Mingalimova, N. Agafonova, T. Makarova, V. Mukhin, A. Paukov, N. Ryabitseva, N. Sedelnikova, I. Stavishenko, T. Tolpysheva, M. Tomoshevich, A. Filippova, Yu. Rebriev, E. Rudykina, N. Shabanova, A. Shiryaev, L. Yakovchenko

filippova.courlee.nina@gmail.com 

30 mycologists
1300 references:
200 are digitized
30 000 records



Biodiversity Data Journal 8: e52963
doi: [10.3897/BDJ.8.e52963](https://doi.org/10.3897/BDJ.8.e52963)



Data Paper

Fungal literature records database of the Northern West Siberia (Russia)

Nina Filippova[†], Stanislav Arefyev[§], Elena Zvyagina^{¶,¶}, Vladimir Kapitonov[#], Tatiana Makarova[,], Victor Mukhin[¤], Nellya Sedelnikova[«], Iraida Stavishenko[»], Anton Shiryaev[»], Tatiana Tolpysheva[^], Natalia Ryabitseva[»], Alexander Paukov[¤]

15.2. Examples: earthworms



Biodiversity Data Journal 12: e130897
doi: [10.3897/BDJ.12.e130897](https://doi.org/10.3897/BDJ.12.e130897)



Data Paper

Earthworm occurrence dataset extracted from Russian-language literature

Maxim Shashkov[‡], Natalya Ivanova[‡], Sergey Ermolov^{‡§}

[‡] Institute of Mathematical Problems of Biology RAS – the Branch
Academy of Sciences, Pushchino, Russia

[§] Center for Forest Ecology and Productivity RAS, Moscow, Russia

3 experts
159 references
5 304 occurrences

Corresponding author: Maxim Shashkov (max.carabus@gmail.com)

Academic editor: Samuel James

Received: 30 Jun 2024 | Accepted: 09 Dec 2024 | Published: 16 Dec 2024

Citation: Shashkov M, Ivanova N, Ermolov S (2024) Earthworm occurrence dataset extracted from Russian-language literature. Biodiversity Data Journal 12: e130897. <https://doi.org/10.3897/BDJ.12.e130897>

16. Examples: birds in Kazakhstan

OCCURRENCE DATASET | REGISTERED JANUARY 19, 2025

Occurrences of Phoenicopterus roseus in Northern Eurasia based on literature sources

Published by [Karaganda Buketov University](#).

Abikenova A

[DATASET](#) [METRICS](#) [ACTIVITY](#) [DOWNLOAD](#) 171 OCCURRENCES

The collected data reflect occurrences of the greater flamingo *Phoenicopterus roseus* across the countries of Northern Eurasia. The information was obtained through digitisation and analysis of publicly available literary sources. Currently, 171 records of species observations have been gathered, and the data continue to be updated. The presented information can be useful for ecological research, biodiversity conservation efforts, and monitoring the population status of the greater flamingo at the northern edge of its range.

171 Occurrences 100% With taxon match 6% With coordinates

11 GEOFERENCED RECORDS

171 occurrences
38 references
1 species

171 OCCURRENCES

[https://www.gbif.org/dataset/
e5c9d4db-5f38-4b1c-bc1f-3e7f685d5415](https://www.gbif.org/dataset/e5c9d4db-5f38-4b1c-bc1f-3e7f685d5415)

17. Examples: spiders in Kazakhstan

Home / Resource

Spider (Arachnida, Araneae) fauna of the lowland part of the Balkhash-Alakol basin (SE Kazakhstan): Literature data

Occurrence

Latest version published on 19 August 2025

1508 occurrences,
3127 individuals
297 species
74 references

Publication date: 18 August 2025
Published by: Institute of Zoology ... Republic of Kazakhstan
License: CC-BY 4.0

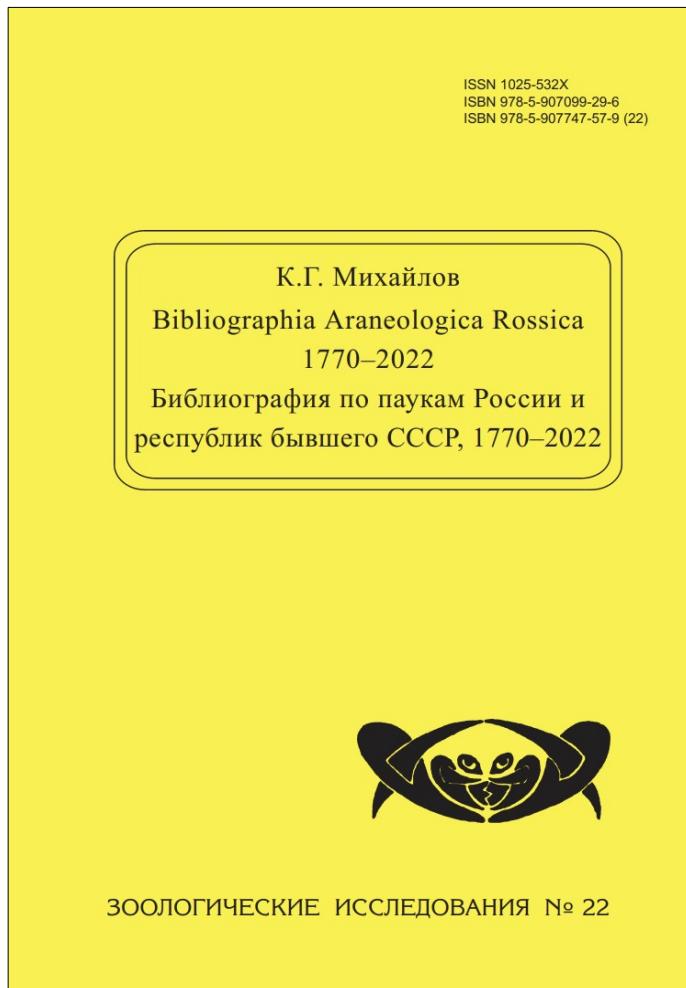
How to cite

Download the latest version of the metadata-only resource metadata as EML or RTF:

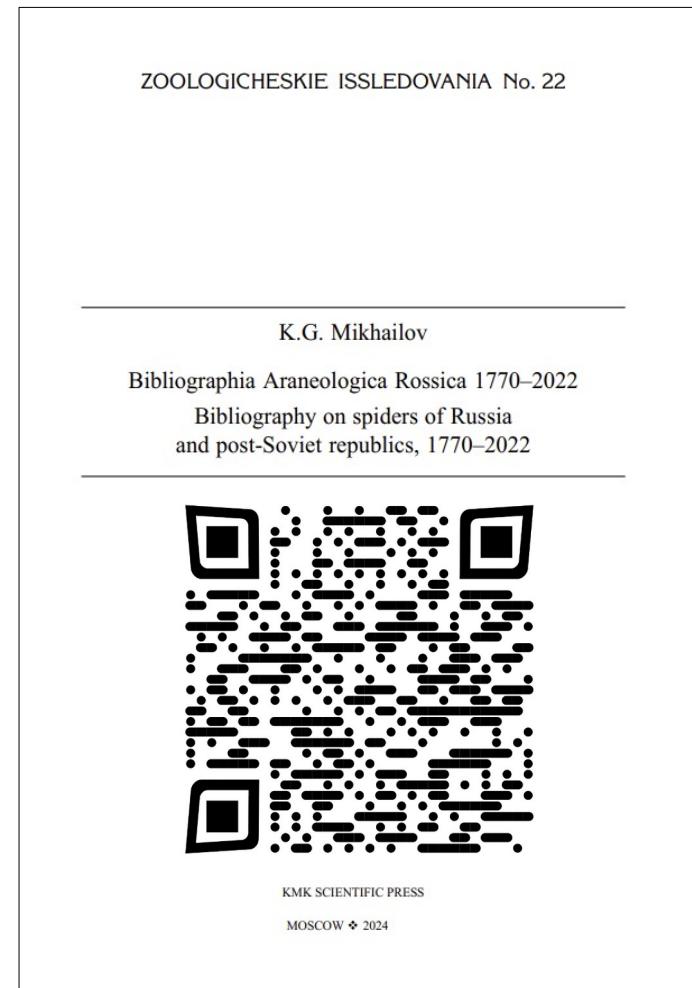
Metadata as an EML file [download in English \(0 bytes\)](#)
Metadata as an RTF file [download in English \(0 bytes\)](#)



18. All references are accounted



Mikhailov, 2012
≈3 950 references



Mikhailov, 2024
≈5 300 references
URL: sozontov.cc/arachnolibrary/files/mikhailov2024.pdf

19. Arachnological literature search & access

Press to search

Watch search results here

Arachnolibrary: digital library for arachnological literature

EN

Authors
Better to use Family Name

Year
2015 2015

Search **Clear**

Скачать найденное:
 [Excel](#) [Zotero](#)

Publication type
 Book Journal article
 Section in monograph/proceedings

Language & files
 Russian English
 others

About the library **Search results** Edit records Add new record

Web application **Arachnolibrary** accumulates all the scientific literature on the post-Soviet space, grants free and simple access to this. The database includes about 5 300 references currently. **Full texts** are available after login on the site.

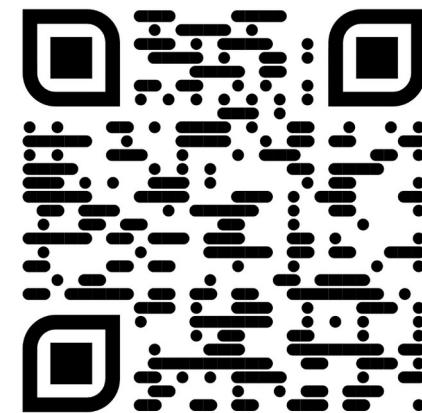
Enter password

Статистика Send new pdf file Send new reference

Arachnological literature data digitization is progressing on the separated resource | with the support of RSF № 24-24-00460.

Resource citation:
Sozontov A.N., Mikhailov K.G. 2025. [Arachnolibrary: digital library for arachnological literature](#) // Invertebrate zoology. Vol. 21. Supplement No 1. Version 2025.07. Access date: 2025-08-18.

Article citation:
Sozontov A.N., Mikhailov K.G. 2024. [Arachnolibrary: digital library for arachnological literature](#) // Invertebrate zoology. Vol. 21. № 4. P. 526-533.



20. Web app as extraction tool



Вся суть кратко:

1. О находках живых организмов существуют тысячи научных публикаций и это число прирастает лавинообразно.
2. Традиционный поиск информации из них (вручную, с перебором всех публикаций) стал слишком трудоёмкий, это тормозит научный прогресс.
3. Перевод этих данных в цифровую форму повысит скорость и эффективность поиска, а следовательно и исследований окружающей среды.
4. Волонтеры берут информацию из научных статей, структурируют и оцифровывают её на онлайн-платформе.
5. В результате волонтеры смогут не только внести свой вклад в науку, но и получить доступ к эксклюзивным материалам и мероприятиям, а также (возможно) побороть свою арахнофобию.

А теперь подробности:

Цель научного исследования

Поиск сведений о находках живых организмов – обязательный этап каждого исследования окружающей среды. Традиционное решение этой задачи – тотальный просмотр всех научных публикаций – простое, но отнимает чрезвычайно много времени, сил и других ресурсов. Необходим переход к использованию средств быстрого и эффективного поиска этих данных, без необходимости смотреть каждую статью вручную. Такие средства есть в готовом виде, есть шаблонные решения с возможностью доработки под конкретные задачи, но... Нет самих данных.

21. Web app as extraction tool: interface (top)

Administrative geography

Country

Россия

Region

District

Место сбора

Местоположение относится к Уралу

Вводить административные названия на русском языке



Locality & coordinates

ГГ.ffff° (56.83777°)

ГГ°ММ.мм' (56° 50.266')

ГГ°ММ'СС" (56° 50' 15.99")

◦ N

Происхождение координат

- Из публикации как есть
- Моя собственная привязка
- Координат нет и не будет

◦ E

Georeference remarks



Sampling event

Год

Месяц

День

Habitats

Коллектор

Фамилия И.О.

Event remarks

Интервал
дат

Месяц
определен

День
определен

Sampling effort

e.g. 20 pitfall-traps*days, 50 net sv



22. Web app as extraction tool: interface (down)

Taxonomy

Family	Genus	Specific epithet	Таксономические примечания
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/> <small>Hold the block</small>

Species defined Отсутствует в списке

sp.n. (described as new species)

ⓘ 🔒 🔓

Amount

Самцов	Самок	Взрослых (пол не определен)	Individual remarks
0	0	0	<input type="text"/>
Субвзрослых самцов	Субвзрослых самок	Ювенильных (пол не определен)	
0	0	0	

23. Requirements for occurrences are achieved

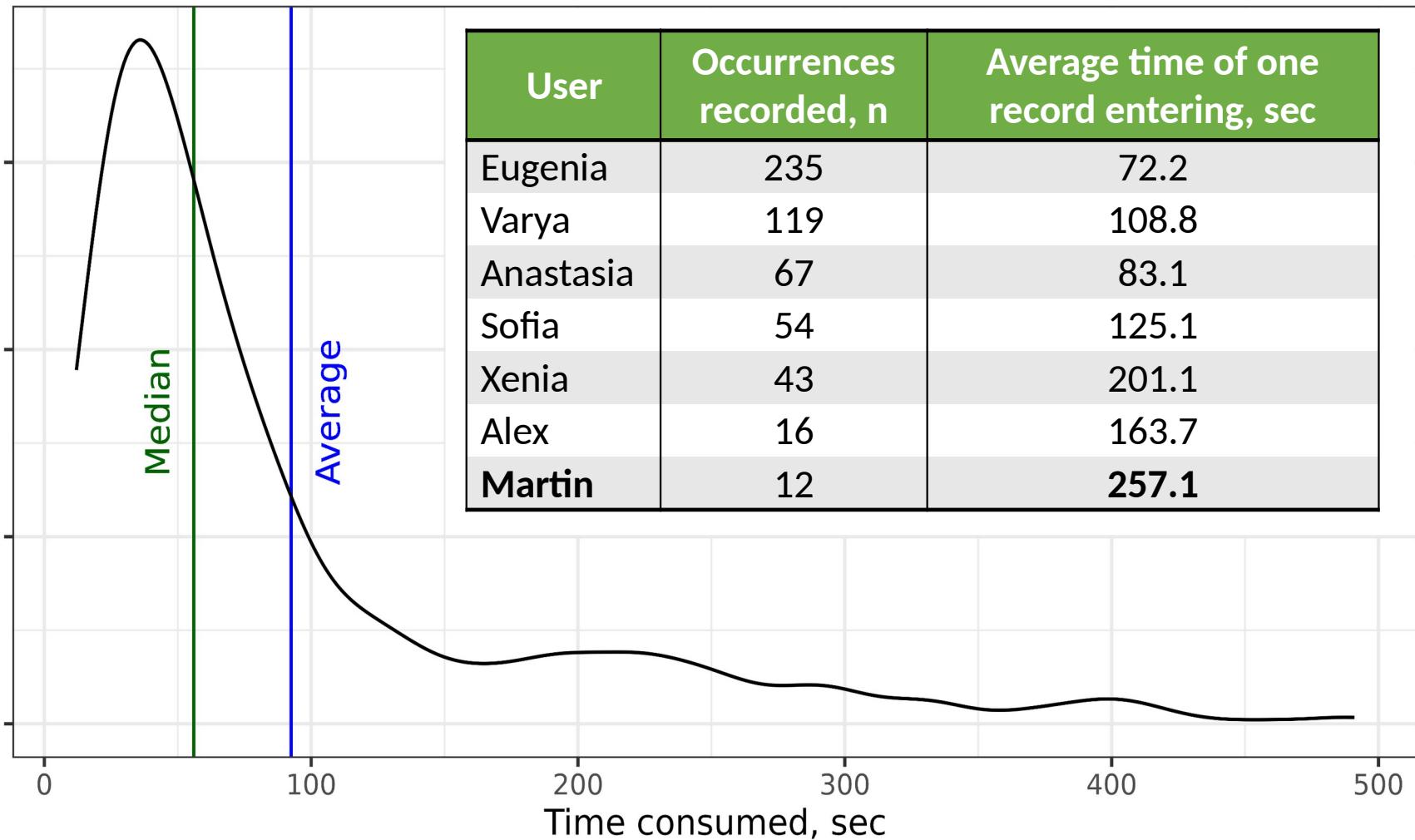
- ✓ 1. Strict accordance between records and references
- ✓ 2. Data structure & standards
 - ✓ 2.1. Darwin Core ✓
 - ✓ 2.2. Check for internal data conflicts ✓
 - ✓ 2.3. Check for mistypes ✓
 - ✓ 2.4. Show the user what to fix ✓
- ✓ 3. Platform independence
- ✓ 4. Optimization of the repetitive data entry
- ✓ 5. Multi-language

Now I know
what to use!

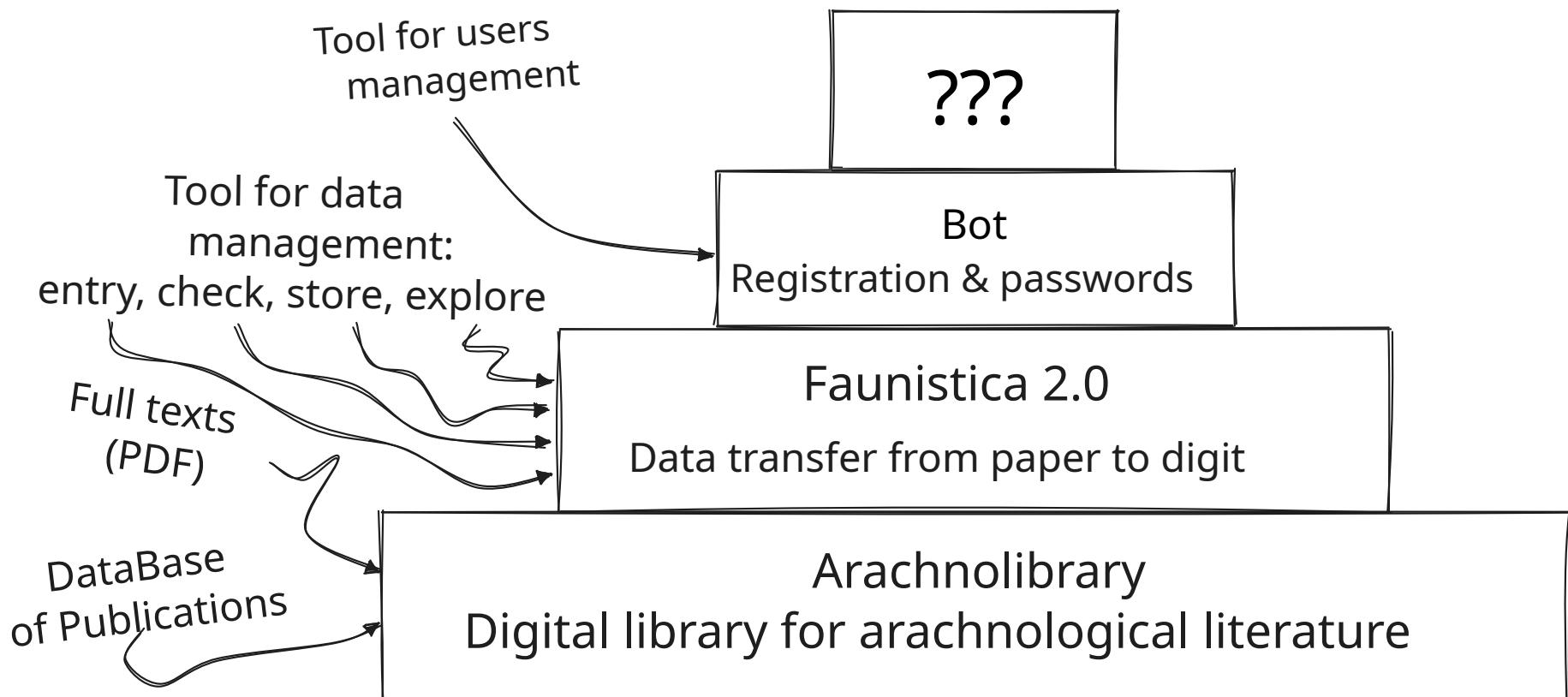


24. Speed of data entry by web-app

Time consumption for one record entering, density plot



25. Project infrastructure

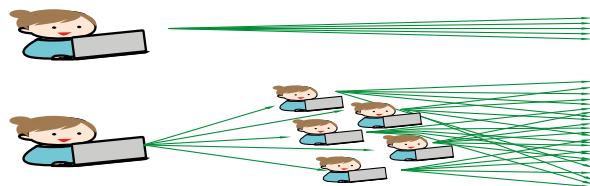


26. Conclusions

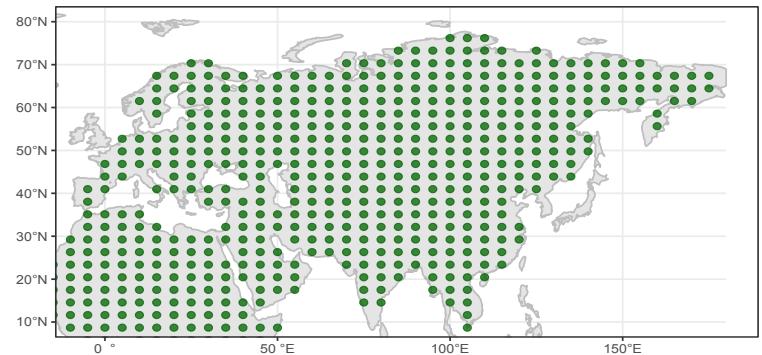
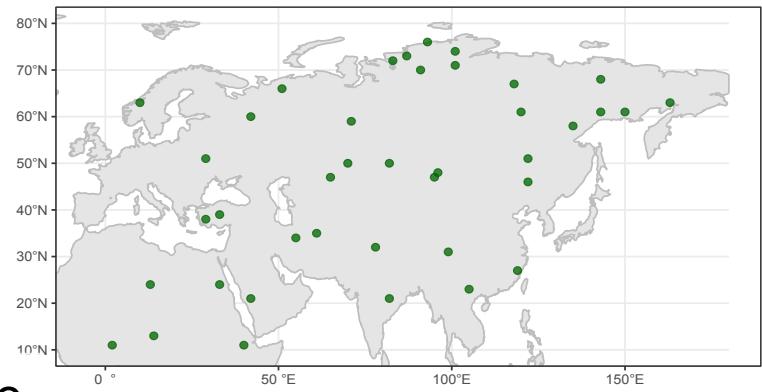
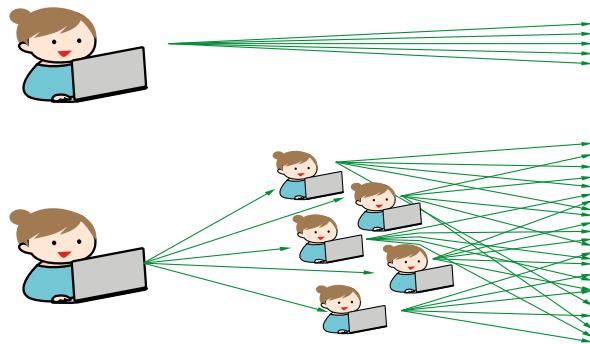
1. Easy and fast access to data enhances data reuse and intensifies researches
2. Precise occurrences data are essential
3. Literature data can also be reused quickly and efficiently
4. Now we have a tool for efficient digitization and using biodiversity data
5. You are welcome to use it

26. Epilogue: Acceleration of the data transfer process

Researcher(s)



Researcher(s) + volunteers = Citizen Science



Acknowledgements



Mikhailov K.G.

for his efforts in arachnological literature accounting and providing these data



Ivanova N.V.,
Plakhina E.V.,
Sokolova S.S.,
Ustinova A.L.

for contributing to the workflow concept, testing software developed, managing volunteer project and many other things

RSF # 24-24-00460

for foundation the project

