Data audit for technical evaluation of **Spider (Arachnida, Araneae) fauna of the lowland part of the Balkhash-Alakol basin (SE Kazakhstan): an integrated assessment**
Resource links: https://www.gbif.org/dataset/92cc4b3a-97f2-4aa3-b260-883ba061c1a9
https://www.gbif.org/dataset/791d91fe-ce27-4044-8a3d-a2a8b203f979
the Darwin Core archives **dwca-almaty_literature-v1.6.zip** and
dwca-almaty_collection-v1.4.zip were downloaded on 2025-10-27 from:
http://gbif.ru:8080/ipt/archive.do?r=almaty_literature
http://gbif.ru:8080/ipt/archive.do?r=almaty_collection

Maxim Shashkov (m.shashkov@pensoft.net)

_____

About this evaluation

Pensoft does a technical evaluation of the dataset (or datasets) referred to in the data paper. If the dataset passes or has only minor problems, the data paper manuscript is referred to reviewers. If the dataset has major problems, a review of the paper is postponed until the dataset has been corrected.

To see what features of a dataset are checked in a technical evaluation, please go to
https://zookeys.pensoft.net/about#Data Quality Checklist Recommendations

Please note that Pensoft does not check the details of the content of a dataset, for example, whether the correct author is given for a scientific name or whether the correct latitude / longitude is given for a locality.

_____

Recommendation: the datasets were improved but still contain a row of issues, some of them boeing reported repeatedly, they should be addressed before the datapaper is accepted for review.

Part 1 - literature
1. According to the Darwin Core definition the **bibliographicCitation** is "A bibliographic reference for the resource.", so it should be unique for a particular occurrence or event within the dataset. While, for example, "Huber B.A. 2011. Revision and cladistic analysis of Pholcus and closely related taxa (Araneae, Pholcidae) // Bonner Zoologische Monographien. Nr.58. P.1–509" mentioned for 26 different occurrences.
Furthermore, if you provide a reference to some literature source in this field, you must be sure that almost all information from this source is represented by a particular occurrence or event in the dataset, and vice versa, there are no more occurrences or events described in this source besides the considered record. In other words, the literature source that you provide as a bibliographic citation and the related event or occurrence must be (near) equivalent. I am not sure that most of the cited literature sources in the dataset contain details regarding only one occurrence.
If you can prove that some bibliographic references contain description only of the particular occurrence or event, you can keep them in this field; otherwise, please move them to the **associatedReferences** field. Briefly, any bibliographic citation in the dataset must unambiguously relates to a particular record, just as you can find exactly one article you need by bibliographic citation.

2. The *informationWithheld* field in the dataset means that some details from the dataset were not provided in the published version of the dataset. It implies that withheld information had been digitazed, however. Please provide what information and for what reasons were withheld rather than reference to sources, from which data users must extract information themselves.

3. It is not a good idea to remove useful details from the dataset. Please get the *acceptedNameUsage* field back, and add the *taxonomicStatus* field that clarify relation the *scientificName* field with the first one.

4. Point regarding the *references* field was my mistake. Please get it back to the dataset.

5. The *occurrenceRemarks* field contains details regarding collection code and catalog number. Please place these details to the correspondent fields.

6. Please remove excessive quotation:
   "Almaty Area, ""Shazyn riverside"""
   "Долина Б. Алматинки, ""Приютская колония"""
   "Алма-Атинский у., долина Б. Алматинки, ""Приютская колония"""
   "Б. Алматинка, ""Приютская колония"""
        and others

7. Is there some sence in the order of the identifiers?
       79   Logunov D.V. | Rakov S.Yu.
       30   Rakov S.Yu. | Logunov D.V.

Part 2 - collection

8. The *verbatimIdentification* provides mostly other values that in previous version. What caused this inconsistency? Make sure it was not resulted from improper table sorting.

9. (partly repeated) Please unify spelling of the taxonRemarks field:

   5   sp.n.: paratypes
   2   sp.n.: paratypes

   1   syntype
   2   syntypes

10. The *eventRemarks* field contains details that could be formalized as *vitality*.

11. There are inconsistencies between the *eventRemarks* and *verbatimEventDate* fields.
        03-05.06.2024   05.09.2024 - found dead.
        05.06.2024   04.11.2024 - found dead.
        05.06.2024   Found dead 02.10.2024 and 05.02.2025.
            and others

12. There are possible inconsistencies between the coordinates and *locationRemarks*

2	46.22432,75.01124	
4	46.22432,75.01124	KATO: 193600000
1	45.9475,74.97284	
1	45.9475,74.97284	KATO: 193657100
6	43.97606,77.41633	
2	43.97606,77.41633	KATO: 191037100
1	43.2452,76.9328	
1	43.2452,76.9328	KATO: 750000000
2	43.685,76.589	
3	43.685,76.589	KATO: 196800000
1	43.2451,76.9326	
10	43.2451,76.9326	KATO: 750000000
1	44.11592,78.70265	
7	44.11592,78.70265	KATO: 334035100
1	43.542,79.436	
1	43.542,79.436	KATO: 196630100
5	43.2142,76.9226	
1	43.2142,76.9226	KATO: 750000000
7	43.2664,76.8178	
26	43.2664,76.8178	KATO: 750000000
1	43.9782,77.4147	
2	43.9782,77.4147	KATO: 191037100
4	43.95361,77.04639	
13	43.95361,77.04639	KATO: 191000000
5	43.239,76.866	
24	43.239,76.866	KATO: 750000000
5	43.224,76.921	
10	43.224,76.921	KATO: 750000000
4	43.224,76.921	N part
15	43.219,76.923	KATO: 750000000
3	43.219,76.923	S part
1	43.2349,76.9203	
2	43.2349,76.9203	KATO: 750000000
1	43.2259,76.9224	
1	43.2259,76.9224	KATO: 750000000