

Many thanks for these valuable corrections. All possible changes in the data and the manuscript are made and accompanied by replies **painted by blue color**.

Part 1 - literature

1. (partly repeated) The **bibliographicCitation** field is intended to provide a reference for the resource itself rather than the source of initial details:

<https://dwc.tdwg.org/terms/#dcterms:bibliographicCitation>

For literature sources, it is more appropriate to use **associatedReferences**:

<https://dwc.tdwg.org/terms/#dwc:associatedReferences>

If you duplicated the **bibliographicCitation** field as **associatedReferences**, please remove the first one.

The example you provided in the response to the initial technical evaluation report is regarding the event, not a particular occurrence.

We disagree that using **bibliographicCitation** for citation the bibliography is a mistake preventing data publishing. In opposite, this is the best match for the title and the content of this column. Mention "... is regarding the event" is doubtful in the current context as far as each occurrence record has all Event columns (potentially) and can be considered as a single-occurrence event. We could split the records at events according the articles or more detailed, but this would be just needless complexity, sophisticating perception and using these data. So we disagree that existence a bibliographic reference as an 'event example' at <https://dwc.tdwg.org/terms/#dcterms:bibliographicCitation> should consider as clear and unambiguous block for using this way in occurrence-type datasets.

There is one another citation from <https://dwc.tdwg.org/terms>: "Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible". This comes from the DublinCore (<http://purl.org/dc/terms/bibliographicCitation>

<https://www.dublincore.org/specifications/dublin-core/dc Terms/#bibliographicCitation>), which is more general (parent) scheme to the DwC, and should be treated as a priority (when examples don't give enough of clarity for us). So putting bibliographic reference here is sufficient enough to show what the resource is and where to find it.

Earlier, you refenced the 'preferred way to cite the resource itself' and now it is more unclear why are not you ready to consider original open available article as 'resource itself' where records comes from. Just to avoid duplicating with **associatedReferences**? This reason is not enough. Whole the DwC standard serves the aim to describe data as detailed as possible, which makes duplicating unavoidable sometimes. **bibliographicCitation** and **associatedReferences** are separated fields, just in this case they are filled by equal values. We could enlarge the second one by a few references more to, concerning the occurrences, species and places to avoid duplication formally, but this makes dataset too noisy.

Keeping in mind that datasets itself citation is another matter, as well as individual records, is well specified here (<https://www.gbif.org/citation-guidelines>), we are insistent on:

- accepting bibliographic reference as a valid value for the **bibliographicCitation** field
- or
- more strong explanation why original article isn't a resource itself where data records comes from (at least in this case).

2. There is no "references" field in the Darwin Core standard. You can use the associatedReferences for specifying URLs as well.

This is not true. Such DwC field exist: <https://dwc.tdwg.org/terms/#dcterms:references>, but now this field is removed from the dataset in order to avoid senseless argues.

3. (partly repeated) Please explain more clearly in the **informationWithheld** field, that for more details data user could refed to the source

This field indicates the literature data (occurrences extracted and provided in the dataset) are based not specimens preserved or observed, but on another literature source, and these sources are cited as well. Additional details are added into the manuscript, section `Data resources`.

4. Please verify the field lists in the **Data resources** section of the manuscript thoroughly. All fields from the datasets should be listed in the manuscript, and vice versa - all fields mentioned in the manuscript must appear in the datasets.

Fixed

5. The **eventDate** and **verbatimEventDate** fields are inconsistent:

1993-06-30/08-12 1993-6-30/1995-8-12
1993-06-30/08-12 1993-6-30/1995-8-12
1993-06-13/06-14 1993-6-13/1996-6-14 and
others.

Changes in the year were not taken into account.

Fixed

6. If the month does not change, it is not necessary to specify it twice:

1919-06-01/06-29
1920-06-01/06-10
1993-05-28/05-15
1993-06-13/06-14

All these months were from years range and these cases were already fixed on the previous step (see above)

7. Is there any difference between the values in the **habitat** field:

building
buildings

This is exact letter by letter copy of original habitats provided by authors so doesn't require clarification remarks in the dataset or data paper. We just can assume first case (Jäger P., Otto S. 2007. New records of Olios sericeus (Kroneberg, 1875) with notes on its taxonomy and biogeography (Araneae: Sparassidae: Sparassinae) // Rev. Ibér. Aracnol. Vol.14 (for 2006). P.19-24) is about single catch when second one is a sort of authors's generalization or the specimen comes from a sample of a few buildings.

8. Please round geographic coordinates up to 5 decimal places: 43.881100000000004

Fixed

9. Please provide full scientific names with authorship and year:

Agalena taurica
Agroeca thorelli
Dictyna latens
Euophrys thorelli
Larinoides folium
and others

If you want to keep initial names according to the literature, you can use the verbatimIdentification field: <https://dwc.tdwg.org/terms/#dwc:verbatimIdentification>

Full scientific names (scientificName) are provided with “+author, year” now (including parentheses where necessary).

According the recommendation, initial names are kept into the field verbatimIdentification now. In opposite to the field scientificName, verbatimIdentification doesn't have taxon' authors sometimes, where it is unclear or/and unnecessary.

10. If you use both **scientificName** and **acceptedNameUsage** fields, the **taxonomicStatus** field would be useful as well: <https://dwc.tdwg.org/terms/#dwc:taxonomicStatus>
acceptedNameUsage field is removed, initial names are in the field verbatimIdentification now, according the previous recommendation (see above).

11. Please unify the spelling of the scientificName field:

Pardosa turcestanica
Pardosa turkestanica (Roewer, 1951)

unified

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

29 sp.n. holotype
4 sp.n.: holotype

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

unified

part2

13. There are multiple inconsistencies between geographic coordinates and localities:

- | | |
|---|--|
| 4 | 46.22432,75.01124 |
| 2 | 46.22432,75.01124 Balkhash Lake, S coast |
| 1 | 45.9475,74.97284 20 km from Karaoi village |
| 1 | 45.9475,74.97284 NW part of Saryesik-Atyrau desert, 20 km NE from Karaoi village |
| 2 | 43.97606,77.41633 near Shengeldy village |

- 6 43.97606,77.41633 Balkhash Lake
- 2 44.91361,76.51472
- 1 44.91361,76.51472 22 km NE from Bakanas village
- 1 43.2452,76.9328
- 1 43.2452,76.9328 Shevchenko str., 89b
- 2 43.685,76.589
- 3 43.685,76.589 Sorbulak Lake
- 5 44.11592,78.70265
- 3 44.11592,78.70265 Altyn-Emel National Park, Shygan cordon
- 1 43.62823,76.74617
- 1 43.62823,76.74617 Sorbulak Lake
- 1 43.542,79.436
- 1 43.542,79.436 Chundzha village
- 1 43.2142,76.9226
- 5 43.2142,76.9226 Al-Farabi ave.
- 1 43.846,79.419
- 1 43.846,79.419 Chundzha, wooden lake
- 4 44.03047,79.5535
- 2 44.03047,79.5535 2.5 km E from Aidarly village
- 1 43.87487,78.19593
- 1 43.87487,78.19593 Altyn-Emel National Park, Zhantogai
- 1 43.9782,77.4147
- 1 43.9782,77.4147 near Shengeldy village
- 1 43.9782,77.4147 Kapshagay reservoir N coast, near Shengeldy village
- 7 43.95361,77.04639
- 10 43.95361,77.04639 Ili river bank, abandoned station of the Mining Institute
- 24 43.239,76.866
- 4 43.239,76.866 Sairan reservoir
- 1 43.239,76.866 Sairan reservoir, W coast

1 44.11639,78.70111
1 44.11639,78.70111 Altyn-Emel National Park

10 43.224,76.921
9 43.224,76.921 KazNU campus

15 43.219,76.923
3 43.219,76.923 KazNU campus

1 43.91278,75.50111 Tamgaly
1 43.91278,75.50111 Tamgaly village

7 43.9516,77.0603 Ili river bank
11 43.9516,77.0603 Ili river bank, abandoned station of the Mining Institute

Fixed

14. There are inconsistencies between the **fieldNumber** and **eventDate** fields: No. of records | fieldNumber | eventDate

1 | WP37 | 2022-10-14
1 | WP37 | 2023-04-05

1 | WP247 | 2023-09-28
1 | WP247 | 2023-09-29

3 | WP262 | 2024-05-08
1 | WP262 | 2024-05-09

1 | WP266 | 2024-05-27
2 | WP266 | 2024-05-18

1 | 43 | 2021
1 | 43 | 2020-07-27

1 | 48 | 2020-07-07
1 | 48 | 2020-07-27

1 | 609 | 2021-08-15
1 | 609 | 2021-08-19

1 | 627 | 2021-12-15
1 | 627 | 2021-08-19

Unfortunately, a few tubes were filled by spiders collected at different dates or even at different places. This is impossible to repair now.

15. It can be advisable to add the ***collectionCode*** field, as your collection was registered in the GRSciColl registry:

<https://scientific-collections.gbif.org/institution/7e82dc97-c81e-4361-845f-4338170452b2/collections>

added with “IZRK_Ara” value