
Presentation

Being consistent with the world latest innovations regarding biodiversity data publishing and scientific papers, the SiB Colombia and *Biota Colombiana* attended for the worldwide call “Activities to enrich metadata and promote the publication of data paper”. This activity took place in late 2012 and was offered by the Global Biodiversity information Facility Secretariat (GBIF), with the aim of support proposals to promote worldwide the publication of Data Papers, papers about biodiversity data.

In this way we are pleased to report that the SiB Colombia and *Biota Colombiana* were granted by the call in Latin America. Thus, with the support of GBIF and implementation by the SiB, *Biota Colombiana* now includes a special section of data papers.

These types of papers are academic publications that have arisen as a mechanism to encourage the publication of biodiversity data, achieving scholarly and professional recognition for people involved in generating and managing biodiversity information.

The data papers should describe as best as possible the Whom, What, Where, When, Why and How to document and record data, without becoming the instrument to make a detailed analysis of the data, as happens in other academic publications. The innovation and improvement of this publishing model relies on the manuscript being always connected to the dataset through a link providing the persistent and reliable web repository of the biodiversity information publishing platform of SiB Colombia.

With the incorporation of this model *Biota Colombiana* is placed at the vanguard in scientific publications, being the first Latin American journal that incorporates this model and one of the top fifteen worldwide, among which are the journals of Pensoft publishing house.

Brigitte L. G. Baptiste

Directora General

Carlos A. Lasso

Editor *Biota Colombiana*
Coordinador
Biología de la Conservación
y Uso de la Biodiversidad

Juan Carlos Bello

Coordinador
Gestión de Información
y Conocimiento

Danny Vélez

Editor asistente
Artículos de Datos

Instituto de Investigación de Recursos Biológicos
Alexander von Humboldt

Guidelines for authors - Data Papers

www.humboldt.org.co/biota - biotacol@humboldt.org.co | www.sibcolombia.net - sib+iac@humboldt.org.co

The purpose of this guide is to establish and explain the necessary steps to prepare a manuscript with the potential to become a publishable data paper in Biota Colombiana. This guide includes aspects related to the preparation of both data and the manuscript.

What is a Data Paper?

A data paper is a scholarly publication that has emerged as a mechanism to encourage the publication of biodiversity data as well as an approach to generate appropriate academic and professional recognition to all those involved in the management of biodiversity information.

A data paper contains the basic sections of a traditional scientific paper. However, these are structured according to an international standard for metadata (information that gives context to the data) known as the *GBIF Metadata Profile* (GMP)¹. The structuring of the manuscript based on this standard enables the community of authors publishing datasets globally, with presence in networks such as the Global Biodiversity Information Facility (GBIF) and other related networks, to publish data easily while getting proper recognition for their work and to encourage the authors of this type of data sets that have not yet published in these global information networks to have the necessary incentives to do so.

A data paper should describe in the best possible way the Whom, What, Where, When, Why and How of documenting and recording of data, without becoming the instrument to make a detailed analysis of the data, as happens in other academic publications. To deepen this publishing model, it is recommended to consult Chavan & Penev (2011)².

Which manuscripts are suitable for publication as data paper?

Manuscripts that describe datasets containing original primary biological records (data of occurrences in a particular place and time); information associated with specimens of biological collections, thematic or regional inventories of species, genomic data and all data likely to be structured with the standard *Darwin Core Darwin Core*³ (DwC). This standard is used in the community of authors publishing biodiversity datasets to structure the data and thus to consolidate and integrate from different sources

globally. It is not recommended to submit manuscripts describing secondary datasets, such as biological records compilations from secondary sources (e.g. literature or compilations of records already published in networks such as GBIF or IABIN).

Dataset preparation

As mentioned above data submitted in this process should be structured based on DwC standard. For ease of structuring, the Biodiversity Information System of Colombia (SiB Colombia), created two templates in Excel; one for occurrences and other for species checklist. Carefully read and follow the template instructions for structuring and publishing data. For any questions about the structure process of data please contact the Coordinator Team of SiB Colombia (EC-SiB) at sib+iac@humboldt.org.co

Manuscript preparation

To assist the creation and structuring of the manuscript in the GMP standard, an electronic writing tool is available (<http://ipt.sibcolombia.net/biota>) to guide the author in the process and ultimately generate a first version of the manuscript. The use of GMP manual as an information guide to include in each section of the manuscript, as well as the annex 1 is recommended.

Steps required for the manuscript preparation:

- 1 Request access to the electronic writing tool at sib+iac@humboldt.org.co. The EC-SiB will assign a username and password.
2. Login to the electronic writing tool, then go to the tab Manage Resources and create a new resource by assigning a short name for your manuscript and clicking on the Create button. Use the format: "InstitutionAcronym_Year_DatasetFeature", e.g. NMNH_2010_rainforestbirds.
3. In the overview of the writing tool click on edit in Metadata section (please, do not use any other section), once there you will find different sections (right panel) that will guide you creating your manuscript. Save the changes at the end of each section, otherwise you will lose the information. Remember to use the GMP manual. Here are some recommendations for editing the metadata, sections are indicated in CAPS and the elements of these sections in **bold**.

¹ GBIF (2011). GBIF Metadata Profile, Reference Guide, Feb 2011, (contributed by O Tuama, E., Braak, K., Copenhagen: Global Biodiversity Information Facility, 19 pp. Accessible at http://links.gbif.org/gbif_metadata_profile_how-to_en_v1.

² Chavan, V. y L. Penev. 2011. The data paper: The mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12 (Suppl 15): S2.

³ Biodiversity Information Standards – TDWG. Accessible at <http://rs.tdwg.org/dwc/terms/>

- In ASSOCIATED PARTIES include only those who are not listed in BASIC INFORMATION.
 - PROJECT DATA and COLLECTION DATA are optional depending on the data type. When using these sections extend or complement information already provided, i.e. do not repeat the same information describing the **description** (GEOGRAPHIC COVERAGE) in the **study area description** (PROJECT DATA).
 - Likewise, in SAMPLING METHODS, you must expand or complete the information, not repeat it. The information in **study extent** should give a specific context of the sampling methodology.
 - It is essential to document the **quality control** in SAMPLING METHODS. Here you should describe what tools or protocols were used to ensure the quality and consistency of data structured with DwC standard.
 - To create the **resource citation** in the CITATIONS section, follow one of the two formats proposed (Annex 2). Do not fill out the **citation identifier**, this will be provided later by the EC-SiB.
 - To include the manuscript bibliography in **citations**, enter each of the citations individually, adding a new citation each time by clicking in the bottom left.
4. Check that the format of the information provided meets the guidelines of the journal (e.g. abbreviations, units, number

formatting, etc.) in the *Biota Colombiana* Guidelines for Authors.

5. Once included and verified all information in the writing tool, notify to EC-SiB at sib+iac@humboldt.org.co, indicating that you have finished editing the manuscript. Additionally attach the Excel template with structured data (remove all columns that were not used). The EC-SiB will perform corrections and final recommendations about the structure of the data and give you the final instructions to submit the paper.

Submit the manuscript

Once you have finished editing your manuscript and getting the instructions from EC-SiB, send a letter submitting your article to email biotacol@humboldt.org.co, following the instructions of *Biota Colombiana* Guidelines for Authors.

Remember to attach:

- Excel template with the latest version of the data reviewed by the EC-SiB.
- Word document with figures and tables followed by a list of them.

At the end of the process, your information will be public and freely accessible in the data portal of SiB Colombia and GBIF. This will allow your data to be available for national and international audience, while maintaining credit to the authors and partner institutions.

Annex 1. Basic structure of a data paper and its mapping to the writing tool elements based on GM.

SECTION/SUB-SECTION HEADING	MAPPING WITH WRITING TOOL ELEMENTS
TITLE	Derived from the title element.
AUTHORS	Derived from the resource creator , metadata provider , and associated parties elements.
AFFILIATIONS	Derived from the resource creator , metadata provider and associated parties elements. From these elements combinations of organization , address , postal code , city , country and email constitute the affiliation .
CORRESPONDING AUTHOR	Derived from the resource contact , metadata provider elements.
CITATION	For editors use.
RESOURCE CITATION	Derived from the resource citation element.
RESUMEN	Derived from the resumen element. 200 words max.
PALABRAS CLAVE	Derived from the palabras clave element. 6 words max.
ABSTRACT	Derived from the abstract element. 200 words max.
KEY WORDS	Derived from the key words element. 6 words max.
INTRODUCTION	Derived from the purpose (Introduction and Background section). A short text to introduce the following sections is suggested. For example, history or context of the biological collection or project related with the data described, only if that information is not present in subsequent sections.
Project data	Derived from elements title , personnel first name , personnel last name , role , funding , study area description , and design description .
Taxonomic Coverage	Derived from the taxonomic coverage elements: description , scientific name , common name and rank .
Geographic Coverage	Derived from the geographic coverage elements: description , west , east , south , north .

cont. **Annex 1.** Basic structure of a data paper and its mapping to the writing tool elements based on GM.

SECTION/SUB-SECTION HEADING	MAPPING WITH WRITING TOOL ELEMENTS
Temporal Coverage	Derived from the temporal coverage elements: temporal coverage type .
Collection data	Derived from the collection data elements: collection name, collection identifier, parent collection identifier, specimen preservation method and curatorial units .
MATERIALS AND METHODS	Derived from the sampling methods elements: study extent, sampling description, quality control and step description .
RESULTADOS	
Descripción del conjunto de datos	Derived from the discussion and acknowledgments, contains information about the format of the data and metadata: hierarchy level, date published and ip rights .
DISCUSSION	Derived from the discussion element. A short text (max 500 words), which can refer to the importance, relevance, usefulness or use that has been given or will give the data in the published literature or in subsequent projects.
ACKNOWLEDGMENTS	Derived from the acknowledgments element.
BIBLIOGRAPHY	Derived from the citations element.

Annex 2. Citation style quick guide for “resource reference” section.

The Resource Reference is the one that refer to the dataset described by the paper, publicly available through SiB Colombia and GBIF networks. Note that this reference may differ from the one of the paper. For more information about this element contact EC-SiB.

Here two formats are suggested; however you can consult other formats established by GBIF⁴.

TYPE OF RESOURCE	TEMPLATE	EXAMPLE
The paper is the result of a collective or institutional project with multiple participants.	<Institution/Research Group>. <Year>, <Title of the Resource/Paper>. <Number of total records>, <provided by :> <associated party 1 (role), associated party 2 (role), (...)>. <Online,> <resource URL>, <published on>. <Published on DD/MM/AAAA>.	National Biodiversity (2013). Vertebrates in Orinoco, 1500 records, provided by: Perez, S. (Principal investigator, content provider), M. Sanchez (Processor), D. Valencia (Custodian Steward, metadata provider), R. Rodriguez (Processor), S. Sarmiento (Publisher), VB Martinez (Publisher, Editor). Online, http://ipt.sibcolombia.net/biota/resource.do?r=verte_orin , published on 01/09/2013.
The paper is the result of a personal initiative or a defined research group.	<associated party 1, associated party 2, (...)>. <Year>, <Title of the Resource/Paper>, <Number of total records>, <Online,> <resource URL>. <Published on DD/MM/AAAA>.	Valencia, D., R. Rodríguez and V. B. Martínez. (2013). Vertebrate Orinoco Basin, 1500 records, Online, http://ipt.sibcolombia.net/biota/resource.do?r=verte_orin , published on 01/09/2001

⁴ GBIF (2012). Recommended practices for citation of the data published through the GBIF Network. Version 1.0 (Authored by Vishwas Chavan), Copenhagen: Global Biodiversity Information Facility. Pp.12, ISBN: 87-92020-36-4. Accessible at http://links.gbif.org/gbif_best_practice_data_citation_en_v1