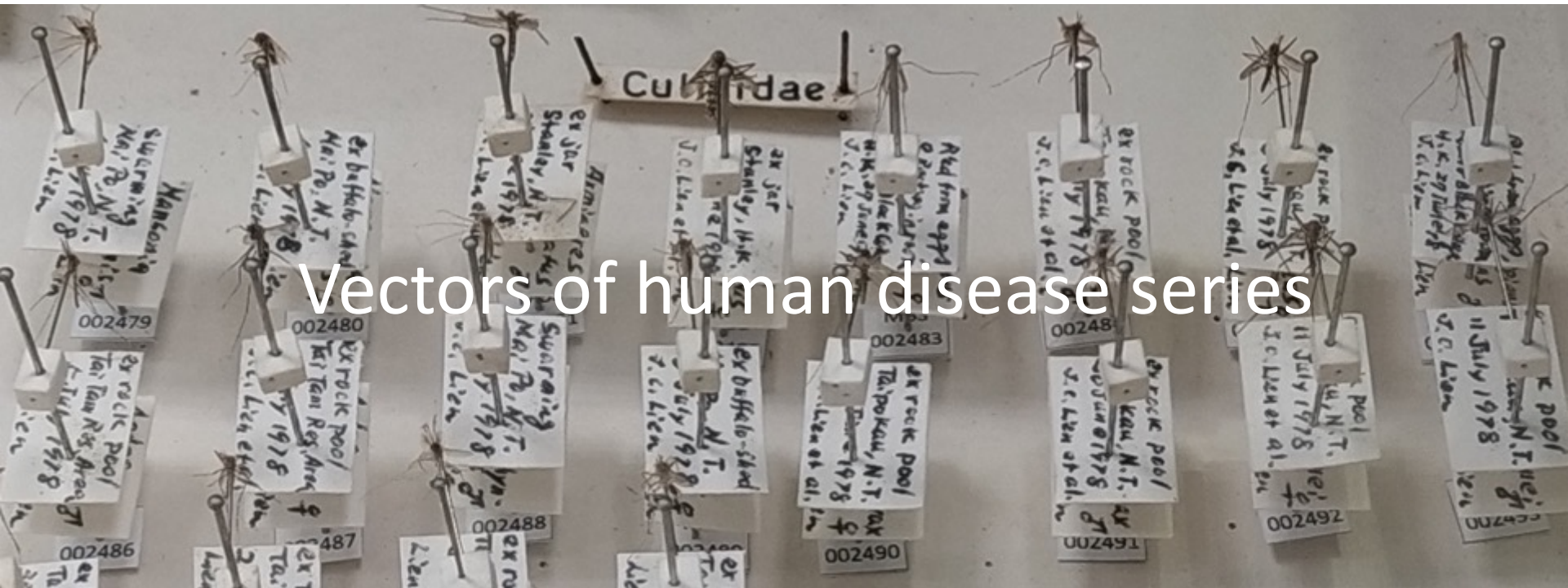


Preparing a data paper for (GIGA)byte



Vectors of human disease series

Scott Edmunds

 0000-0001-6444-1436

scott@gigasciencejournal.com

8th November 2022





Rewarding open data: *GigaScience*

Launched July 2012, now partnering with OUP. Publishes “Data Notes” for CC0 data.

OXFORD ACADEMIC Sign In Register

(GIGA)ⁿ SCIENCE (GIGA)ⁿ SCIENCE PRESS

Articles Submit Alerts About All GigaScience Advanced Search

COVID-19 – Support for our author and subscriber community



Latest Issue
Volume 11, Issue 1
2022

Publishing Director
Laurie Goodman, PhD

Editor in Chief
Scott Edmunds, PhD

Executive Editor
Nicole Nogoy, PhD

Editor
Hans Zauner, PhD

Assistant Editor
Hongling Zhou

About the journal

GigaScience is an open access, open data, open peer-review journal focusing on 'big data' research from the life and biomedical sciences.

[Learn more](#)

GigaScience has been selected as the 2018 Prose Awards Winner for "Innovation in Journal Publishing."

Published by:



2011-2016



2016-date



<http://gigasciencejournal.com/>



Rewarding open data: *GigaScience*

APC covers curation and 1TB of storage in our GigaDB repository

Since 2011,
and working
with



The screenshot shows the GigaDB website interface. At the top, there is a navigation bar with 'Login / Signup' and social media icons. Below the header, the GigaDB logo and tagline 'Revolutionizing data dissemination, organization, and use' are displayed. A main banner features the text 'GIGADB DATASETS' and a description: 'GigaDB contains 2130 discoverable, trackable, and citable datasets that have been assigned DOIs and are available for public download and use.' Below the banner, there is a 'Dataset types' section with a grid of icons and counts for various categories: Genomic (1305), Software (342), Transcriptomic (231), Imaging (1196), Neuroscience (42), Epigenomic (37), Metagenomic (67), Genome mapping (17), Workflow (79), Proteomic (28), and Metabarcoding (8), Metadata (21). To the right, an 'RSS' section lists recent dataset additions with their DOIs and titles. At the bottom, four large cards display key statistics: 2130 Datasets, 127517 Samples, 344277 Files, and 47 T Data volume(TB). The footer includes the GigaScience logo, BGI 华大, and contact information: database@gasciencejournal.com.

<http://gigadb.org/>



UNESCO Recommendation on Open Science

<https://unesdoc.unesco.org/ark:/48223/pf0000379949>

Signed on the 23rd November 2021 at the 41st session of UNESCO by 193 Members States

The Recommendation affirms the importance of open science as a vital tool to improve the quality and accessibility of both scientific outputs and scientific process, to bridge the science, technology and innovation gaps between and within countries and to **fulfill the human right of access to science.**

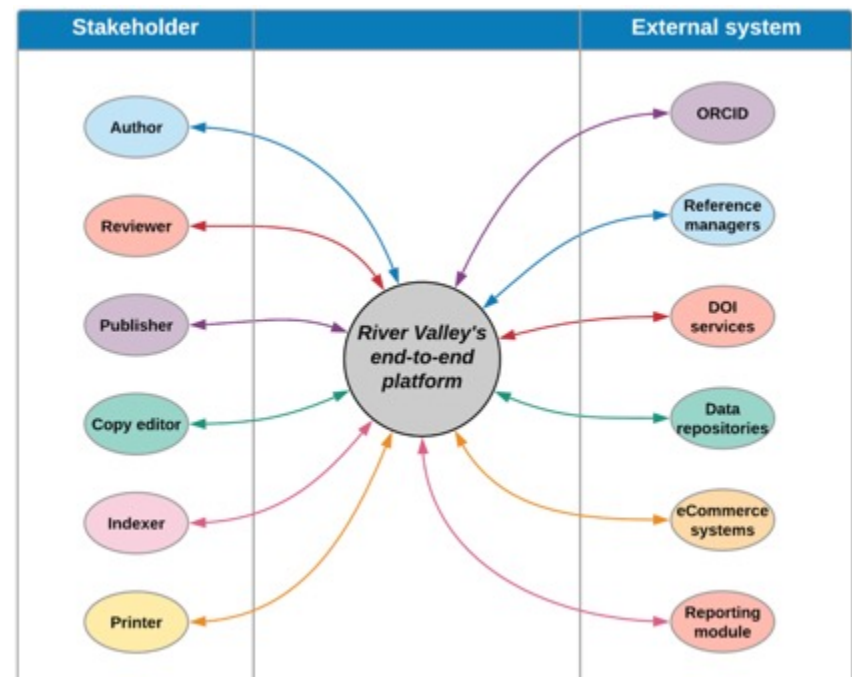
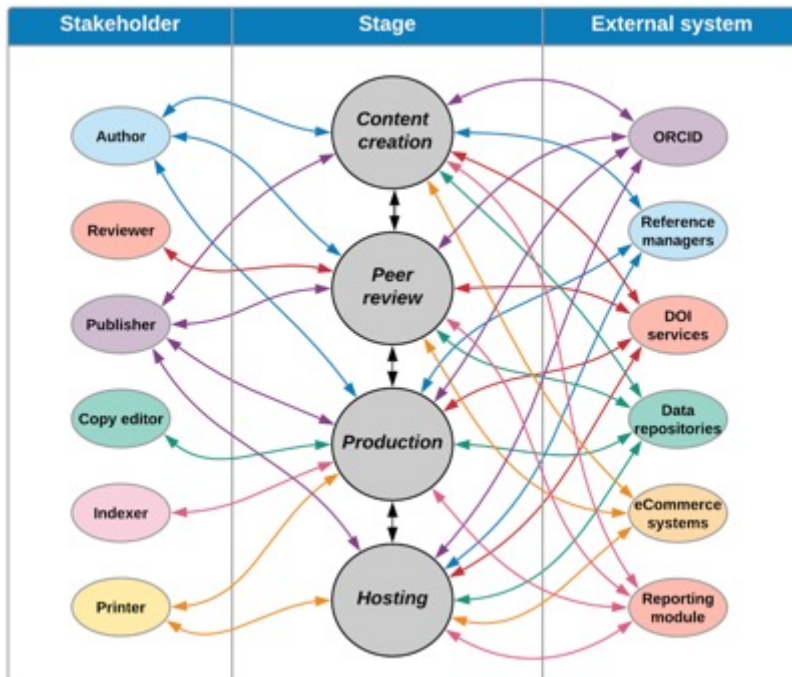
Recommends that Member States apply the provisions of this by taking whatever legislative or other measures may be required to give effect within their jurisdictions to the principles of this Recommendation.

Addressing this with a new journal: (GIGA)byte

Main advantage of workflow is XML from start to end

Traditional workflow: API spaghetti

New: straightforward, single platform



<https://gigabytejournal.com/>



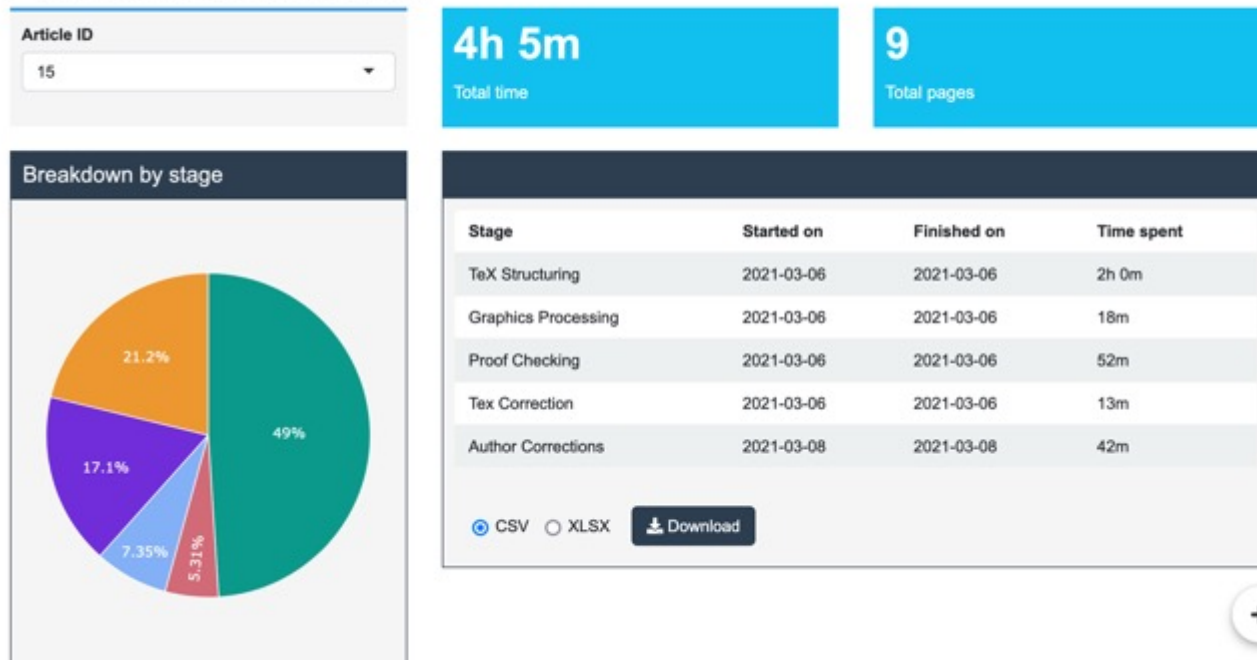
RIVER VALLEY
TECHNOLOGIES

Inclusive features: the cost barrier

Huge time + cost savings from XML-first workflow

DOCX/.latex to XML to paper. Automated production (only needing humans for pagination)

Production time breakdown



End-to-end analytics

Production times in hours not days

Current APC = \$350 USD covers these publication costs (with 10% markup)

Real stats from DARVi (Data Analyser, Reporter and Visualiser) <https://rivervalley.io/darvi/>

Thinking about users: authors, reviewers, readers

What does a GigaByte data paper look like?

Data Release: a short, updatable, description of a research dataset



(GIGA)byte

Gigabyte, 2020, 1–8

DOI: 10.1093/gigabyte/gba001
Manuscript in Preparation
Paper

PAPER

Template for GigaByte Journal Data Report Submissions

First Author^{1,*}, Second Author^{2,*}, Third Author² and Fourth Author^{2,*}

¹First Institution and ²Second Institution

*doi:10.1093/gigabyte/gba001
*Contributed equally

Abstract

The abstract of the manuscript should cover a presentation of the interest or relevance of these data for the broader community; a very brief preview of the data type(s) produced, the methods used, and information relevant to data validation. As well as the potential uses of these data and implications for the field. Please minimize the use of abbreviations and do not cite references in the abstract. As this article type is focussed on describing a dataset, conclusions or interpretive insights are not required.

Classifications: Classification 1; Classification 2; Classification 3 (Minimum three key words from the GigaScience/GigaDB Classifications List)

Discoverability & credit: Highlights and help to contextualize openly available datasets to encourage reuse.

Sharing: All data can be linked to the Data Release via GBIF, GigaDB or other data DOIs or accessions.

Data, not analysis: Incentivizes and allows more rapid releases of data before subsequent detailed analysis has been carried out. Or in coordination with publication of an analysis paper.

Simple: Structure = *Context, Methods, Data Validation and QC, Reuse Potential, Data Availability*

Submit via:



Overleaf
L^AT_EX

Integrates with preprints:



<https://gigabytejournal.com/data-release-description>

Bringing this together: Open Science for Public Health



Access to data on vectors and vector-borne diseases is improved through the release of a special issue publication of a series of data papers

26 June 2022 | News release | Reading time: 2 min (552 words)

Publication of a series of 11 papers with data on vectors that transmit vector-borne diseases is a significant advance in the availability of such data in an easy and open access format. Published by the journal *GigaByte* by GigaScience Press, in partnership with the Global Biodiversity Information Facility and supported by TDR, the papers improve knowledge on vectors and pave the way forward in data sharing.



Aedes aegypti, vector of arboviral diseases caused by viruses such as Dengue, Chikungunya and Zika.
Picture from Pascal Gaborit, Pasteur Institute of French Guiana, in the Institut Pasteur photothèque

Currently affecting about half of the world's population, vector-borne diseases are transmitted by arthropod (insect and tick) vectors such as mosquitoes and flies; they include diseases such as malaria, dengue and leishmaniasis. While some diseases such as malaria show a decline in numbers of fatalities, others, such as those caused by arboviruses like dengue, chikungunya and Zika viruses, show a worrying increase in number of cases (although improved case management has led to a decrease in fatalities). Factors such as climate change and urbanization have the potential to further affect the impact of these diseases, so it is vital that research is ongoing to understand more about the vectors and to ensure that data are shared widely with researchers and policy-makers in an open access, freely available and discoverable format.

Additional content of submissions

Papers presenting sampling methods and protocols, diverse data types, examples working with citizen scientists and indigenous communities in the Amazon, and papers showcasing many novel technical features...



End product: Open Science for Public Health

Data published from the *GigaByte Vectors* of human disease series includes:

Vectors

Mosquitos Sand flies Ticks Triatominae

GBIF mediated data

>500,000 Occurrence Records Specimens >675,000

More than **50 Countries**

Data Openness
CC0 Public Domain

Data Types

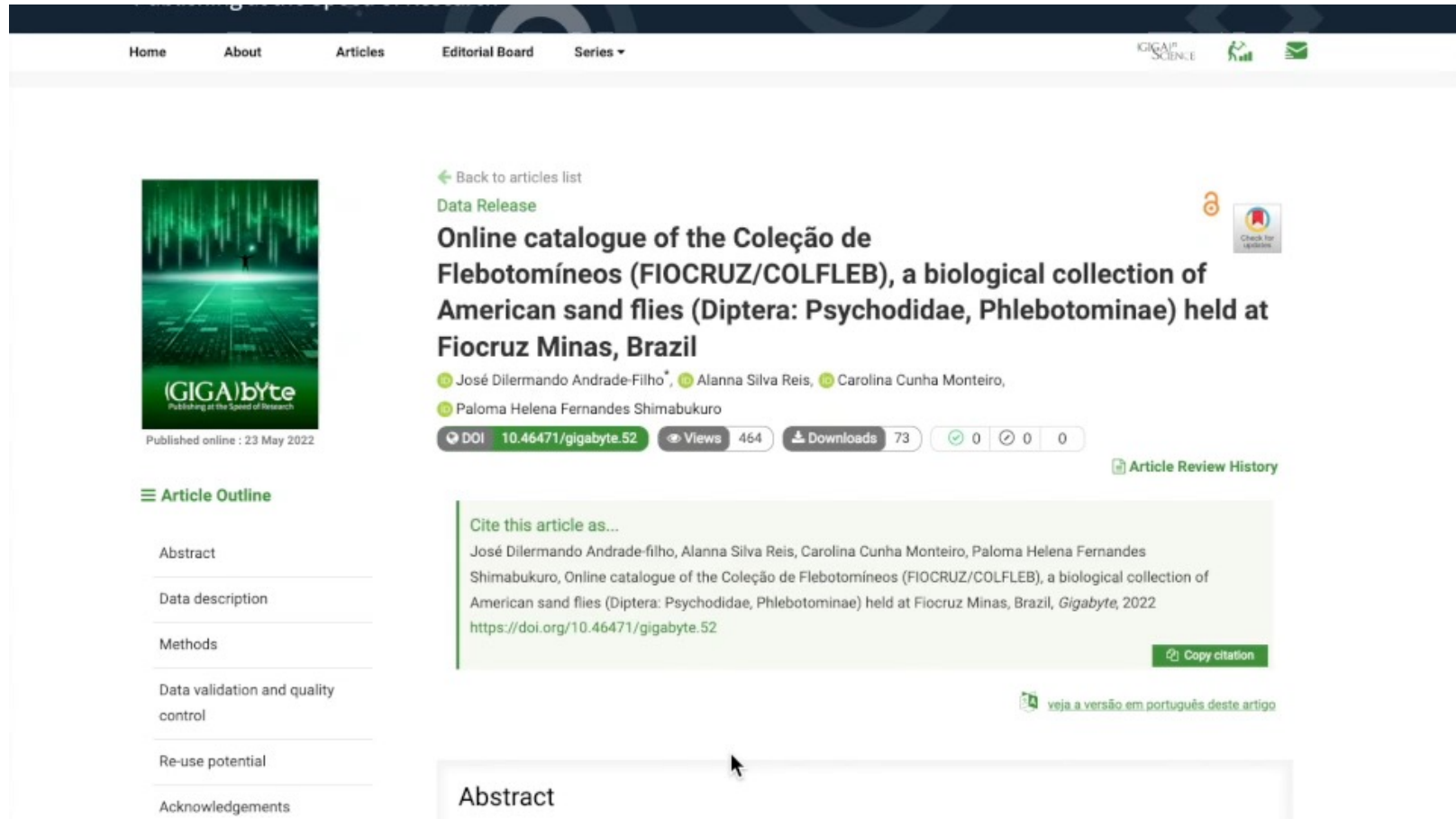
Observations Specimens Imaging Data DNA Barcodes Citizen Science

PHASE II IS OPEN FOR SUBMISSIONS...submit by 30 April 2023.

https://doi.org/10.46471/GIGABYTE_SERIES_0002

End product: Open Science for Public Health

Multilingualism for better accessibility, understandability and trust



The screenshot shows the article page for 'Online catalogue of the Coleção de Flebotomíneos (FIOCRUZ/COLFLEB), a biological collection of American sand flies (Diptera: Psychodidae, Phlebotominae) held at Fiocruz Minas, Brazil'. The page includes a navigation menu, a cover image, author information, DOI, views/downloads, and a citation box.

Home About Articles Editorial Board Series

GIGABYTE SCIENCE

← Back to articles list

Data Release

Online catalogue of the Coleção de Flebotomíneos (FIOCRUZ/COLFLEB), a biological collection of American sand flies (Diptera: Psychodidae, Phlebotominae) held at Fiocruz Minas, Brazil

José Dilermando Andrade-Filho*, Alanna Silva Reis, Carolina Cunha Monteiro, Paloma Helena Fernandes Shimabukuro

DOI 10.46471/gigabyte.52 Views 464 Downloads 73

Article Review History

Cite this article as...
José Dilermando Andrade-filho, Alanna Silva Reis, Carolina Cunha Monteiro, Paloma Helena Fernandes Shimabukuro, Online catalogue of the Coleção de Flebotomíneos (FIOCRUZ/COLFLEB), a biological collection of American sand flies (Diptera: Psychodidae, Phlebotominae) held at Fiocruz Minas, Brazil, *Gigabyte*, 2022
<https://doi.org/10.46471/gigabyte.52>

veja a versão em português deste artigo

Abstract

<https://doi.org/10.46471/gigabyte.52>

Notes on the vectors series

The data papers submitted should describe datasets with the following criteria:

- Data has clear relevance for research on vectors of human vector-borne diseases
- Dataset contains more than 5,000 records that are new to GBIF.org in 2022/23 with high-quality data and metadata
- Data is dedicated to the public domain under an open CC0 designation
- Deadline for submission is 30 April 2023





Notes on the vectors series

Data deposition is key, and supported by GBIF helpdesk and GigaDB curators

- Authors should start by preparing the dataset and publishing it through GBIF.org before writing
- Support from health@gbif.org for questions on publishing data through GBIF, data standards, etc.
- GigaDB team (database@gigasciencejournal.com) on hand to help with additional supporting data
- GigaDB curators will also help review process by providing a data audit for each submission

Issues flagged in data audit process



- Should use the GBIF Data Validator before submission (check these issues)
- Double check license is CC0 (can fix after submission)
- Curators then look for inconsistencies between statistics in GBIF data and paper
- Randomly select 5-10 fields per dataset for deeper inspection
- Generally consistent, but few examples with inconsistencies in metadata caused by conversion problems and non-ASCII characters
- What's an "Observation" (specimens/sampling events/etc.) can be defined by the author (but nice to aim for richer)

Open peer review inc. audit checklist

Submission ID: DRR-202202-03

Title: **AIMSurv: First pan-European harmonized surveillance of *Aedes* invasive mosquito species of relevance for human vector-borne diseases.**

Brief summary of manuscript:

The datanote described the coalition of multiple european partners coming together to survey and monitor the distribution of mosquitoes across europe. It is well written and describes the context and collection methods very well.

In general, the GBIF dataset accompanying this manuscript is well annotated and very close to complete for the time period covered.

The manuscript alludes to additional data being collected for at least part of the study "The more ambitious RP sampling included...the daily or weekly record of meteorological parameters (maximum, minimum, average temperature) per site, collected using data loggers or local weather stations." Those data are not available in the GBIF dataset. Are they available elsewhere? If so, can a PID be provided for access to those data? If they are not currently hosted publicly anywhere, then a GigaDB dataset can be created to host them.

Number of GBIF datasets included in the manuscript - 1

Major comments (Author action required):

- 1 - The GBIF dataset needs to be made CC-0
- 2 - provide access to the additional data collected for the RP sampling

See Curation checklist below for additional information.

Minor comments (Author action suggested):

- 1 - if the missing GPS coordinates are available they should be added to the GBIF dataset.
- 2 - consider adding more than 1 contributor to the GBIF dataset

See Curation checklist below for additional information.

Curation checklist

GBIF dataset name: **AIMSurv *Aedes* Invasive Mosquito species harmonized surveillance in Europe. AIM-COST Action**

Name used in MS :The dataset is not referred to be name in the MS

GBIF UUID = 03269e13-84ae-430f-890e-f11069413e36

GBIF DOI = <https://doi.org/10.15470/vs3677>

Alternate URI(s) =

GBIF dataset date = **Published 23-MAR-2022.**

Dataset description: .

Contact author: only 1 author is listed as all 3 things (originator, contact and metadata provider), for future proofing of contact I would like to see at least 2 individuals listed.

For more see series overview/umbrella commentary in *GigaScience*



GigaScience, 2022, 0, 1-0
DOI: 10.1093/gigascience/giac114
Commentary

Publishing data to support the fight against human vector-borne diseases

Scott C Edmunds^{1,2}, Florence Fouque³, Kyle A Copas⁴, Tim Hirsch⁵, Dmitry Schigel⁶, Paloma Helena Fernandes Shimabukuro^{4,5}, José Diêrmando Andrade-filho^{4,5}, Catalina Marceló⁴, Carlos Andrés Morales⁷, Maria Camila Lesmes^{4,8}, Patricia Fuya⁹, Sergio Méndez⁹, Horacio Cadena⁹, Álvaro Avila-Díaz⁹, Erika Santamaría¹⁰, Žilko Južnič-Zonta¹¹, Roger Eritja¹², John R.B. Palmer¹², Frederic Bartumeus^{10,11,13}, Maurício dos Santos-Conceição¹⁴, Samira Chahad-Ehlers¹⁵, Cássio Lázaro Silva-Inácio¹⁶, Ana Leuch Lozovei¹⁴, Andrey José de Andrade¹⁷, Sara Paull¹⁸, Miguel Ángel Miranda¹⁹, Carlos Barceló¹⁹, Francis Schaffner²⁰, Alessandra Della-Torre²⁰, Dimitri Erensens²¹, Wouter Dekoninck²², Guy Hendrickx²⁴, Wim Van Bortel²⁵, Isza Deblauwe²⁵, Nathalie Smitz²⁶, Veezie Versteirt²⁷, Rodrigo Espindola Godoy²⁸, Andreia Fernandes Brilhante²⁹, Soledad Ceccarelli^{30,31}, Agustín Balsalobre^{30,31}, María Eugenia Vicente³⁰, Rachel Curtis-Robles³², Sarah A. Hamer³², José Manuel Ayala Landa³³, Jorge E. Rabinovich^{30,31} and Gerardo A. Marti^{30,31}

- ¹GigaScience Press, BGI Hong Kong Tech Co Ltd., Rm A 20F Kings Wing Plaza 2, 1 On Kwan Street, Shek Mun, Sha Tin, NT, Hong Kong SAR
²Special Programme for Research & Training in Tropical Diseases (SPRD), World Health Organization, Avenue Appia 20, 1211 Geneva 27, Switzerland
³GRIFF Secretariat Universitetsparken 15, DK-1100 Copenhagen Ø, Denmark
⁴Colectivo de Flebotomíneos (FIOCRUZ-COLFLEB), Instituto René Rachou, Fiocruz Minas Avenida Augusto de Lima, 1715 - Barro Preto, 30190009, Belo Horizonte, Brazil
⁵Grupo de Estudos em Leishmanioses, Instituto René Rachou, Fiocruz Minas Avenida Augusto de Lima, 1715 - Barro Preto, 30190009, Belo Horizonte, Brazil
⁶Grupo de Entomología, Instituto Nacional de Salud, Bogotá, Colombia
⁷Secretaría de Salud Departamental del Cauca, Popayán, Colombia
⁸Universidad de Ciencias Aplicadas y Ambientales, Bogotá, Colombia
⁹Programa de Estudio y Control de Enfermedades Tropicales PECECT, Medellín, Colombia
¹⁰Centre d'Estudis Avançats de Blanes (CEAB-CSIC), C/Ítaca a la Cala St. Francesc 14, 17300 Blanes, Girona, Spain/Centre d'Estudis Avançats de Blanes (CEAB-CSIC), C/Ítaca a la Cala St. Francesc 14, 17300 Blanes, Girona, Spain
¹¹Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Edifici C Campus de, 08193 Bellaterra, Barcelona, Spain
¹²Departament de Ciències Polítiques i Socials, Universitat Pompeu Fabra, Plaça de la Mercè, 10-12, 08002 Barcelona, Spain
¹³Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Passeig de Lluís Companys, 08010 Barcelona, Spain
¹⁴Basic Pathology Department, Federal University of Paraná, Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba, PR 81531-980, Brazil
¹⁵Genetics and Evolution Department, Federal University of São Carlos, Rodovia Washington Luis, km 235 SP-310, São Carlos, SP 13565-905, Brazil
¹⁶Post-graduate Programme in Entomology, Zoology Department, Federal University of Paraná, Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba, PR 81531-980, Brazil
¹⁷National Ecological Observatory Network, Battelle, 1685 38th St, Boulder, CO 80501, USA
¹⁸Applied Zoology and Animal Conservation group, University of the Balearic Islands (UIB), Ctra Valdemossa km 7.5, 07122 Palma, Spain
¹⁹Francis Schaffner Consultancy, Lörracherstrasse 50, 4125 Riehen, Switzerland
²⁰Dep. Public Health and Infectious diseases, University Sapienza, Piazzale Aldo Moro 5, 00185 Roma, Italy
²¹Research Institute for Nature and Forest (INBO), Havenlaan 88 b73, 1000, Brussels, Belgium
²²Royal Belgian Institute for Natural Sciences (RBINS - BopCo & Scientific Heritage Service), Vautierstraat 29, 5000, Brussels, Belgium
²³Avia-GIS NV, Raschoflei 31, 2980, Zoersel, Belgium
²⁴Unit Entomology, Dept. of Biomedical Sciences, Institute of Tropical Medicine (ITG), Nationalstrasse, 115, 2000, Antwerpen, Belgium
²⁵Royal Museum for Central Africa (RMCA - BopCo), Leuvensesteenweg 17, 3080 Tervuren, Belgium
²⁶Agency for Nature and Forests, (ANR), Havenlaan 88 b75, 1000, Brussels, Belgium
²⁷Independent Researcher, Brazil
²⁸Universidade Federal do Acre, Departamento de Ciências da Saúde e Educação Física, Universidade Federal do Acre, Distrito Industrial, Rio Branco, 69920900, Brazil
²⁹Centro de Estudios Parasitológicos y de Vectores (CEPAVE-CCF-La Plata-CONICET-UNLP), La Plata, Buenos Aires 1900, Argentina
³⁰Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
³¹College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA
³²Museo del Instituto de Zoología Agrícola (MIZA), Maracay, Venezuela
³³Corresponding author: Scott Edmunds, BGI Hong Kong Tech Co Ltd., Hong Kong, NT HONG KONG. E-mail: scott@gigasciencejournal.com

Abstract

Vector-borne diseases are responsible for more than 17% of human cases of infectious diseases. In most situations, effective control of debilitating and deadly vector-borne diseases (VBDs), such as malaria, dengue, chikungunya, yellow fever, Zika and Chagas requires up-to-date, robust and comprehensive information on the presence, diversity, ecology, bionomics and geographic spread of the organisms

Received: October 21, 2022. Accepted: October 24, 2022
© The Author(s) 2022. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

<http://dx.doi.org/10.1093/gigascience/giac114>

Association of Learned & Professional Society Publishers Innovation in Publishing Award Winner 2022



<http://gigasciencejournal.com/blog/gigabyte-alpsp-innovation-award/>

Thanks to TDR/WHO for support of this datasets on vectors of human diseases series

Due to this very generous sponsorship the article processing fee (normally \$350 USD) will be waived for the first 15 papers that are accepted and meet the series criteria.



Many thanks to our partners



(GIGA)byte

For further questions contact: editorial@gigabytejournal.com



Submit by 30th April 2023:

<https://gigabytejournal.com/>

