Received: 12 July 2025 | Revised: 9 October 2025 | Accepted: 25 November 2025

DOI: 10.1002/ecy.70317

DATA PAPER

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

WABAD: A world annotated bird acoustic dataset for passive acoustic monitoring

Cristian Pérez-Granados[1,2] | Jon Morant[2] | Kevin F. A. Darras[3] | Oscar H. Marín-Gómez[4] | Irene Mendoza[5,6] | Miguel A. Muñoz-Mohedano[6,7] | Eduardo Santamaría-García[6] | Giulia Bastianelli[6] | Alba Márquez-Rodríguez[6,8] | Michał Budka[9] | Gerard Bota[1] | José M. De la Peña-Rubio[10] | Eladio L. García de la Morena[11] | Manu Santa-Cruz[12] | Pablo de la Nava[13] | Mario Fernández-Tizón[14] | Hugo Sánchez-Mateos[15] | Adrián Barrero[16,17] | Juan Traba[16,17] | Tomasz S. Osiejuk[9] | Patrick J. Hart[18] | Amanda K. Navine[18] | Andrés F. Montoya Muñoz[4] | Carlos B. de Araujo[19] | Gabriel L. M. Rosa[20] | Ingrid M. Denóbile Torres[21] | Ana L. Camargo Catalano[21] | Cássio Rachid Simões[21,22] | Diego Llusia[16,17,23] | Manuel B. Morales[16,17] | Pablo Acebes[16,17] | Juan A. Medina[24] | Nicholas Brown[16,25] | Christos Astaras[26] | Ilias Karmiris[26] | Elizabeth Navarrete[26] | Maxime Cauchoix[27] | Luc Barbaro[28] | David Funosas[29] | Dominik Arend[30] | Sandra Müeller[30] | Fernando González-García[31] | Alberto González-Romero[31] | Christos Mammides[32] | Michaelangelo Pontikis[33] | Giordano Jacuzzi[34] | Julian D. Olden[34] | Sara P. Bombaci[35] | Gabriel Marcacci[36] | Alain Jacot[36] | Juan P. Zurano[19] | Elena Gangenova[19] | Diego Varela[19] | Facundo Di Sallo[19] | Gustavo A. Zurita[19] | Andrey Atemasov[37] | Junior A. Tremblay[38] | Vincent Lamarre[38] | Anja Hutschenreiter[39] | Alan Monroy-Ojeda[40] | Mauricio Díaz-Vallejo[41,42] | Sergio Chaparro-Herrera[43] | Robert A. Briers[44] | Renata Sousa-Lima[45] | Thiago Pinheiro[45] | Wigna C. da Silva[45] | Alice Calvente[46] | Anamaria Dal Molin[47] | Alexandre Antonelli[48,49,50] | Svetlana Gogoleva[51,52] | Igo Palko[52] | Hiếu Vũ Trọng[52] | Marina H. Lage Duarte[53,54] | Natalia dos Santos Saturnino[54] | Samuel R. Silva[54] | Ana Rainho[55] | Paula Lopes[55,56] | Karl-L. Schuchmann[57,58] | Marinêz I. Marques[57] | Ana S. de Oliveira[57] |

For affiliations refer to page 3

Nick A. Littlewood[59] | Mao-Ning Tuanmu[60] | Yi-Ru Cheng[60] | Hsuan Chao[60] | Sebastian Kepfer-Rojas[61] | Andrea L. Aguilera[62] | Lluís Brotons[1,63,64] | Mariano J. Feldman[1] | Louis Imbeau[65] | Pooja Panwar[66] | Aaron S. Weed[67] | Anant Deshwal[68] | Raiane Vital da Paz[69,70] | Carlos Salustio-Gomes[69,70] | Dorgival D. Oliveira-Júnior[69,70] | Cicero S. Lima-Santos[69,70] | Mauro Pichorim[69,70] | Wuyuan Pan[71] | Eben Goodale[72] | Alfredo Attisano[73] | Jörn Theuerkauf[73] | Esther Sebastián-González[2,74]

Correspondence
Cristian Pérez-Granados
Email: cristian.perez@ctfc.cat

Handling Editor: Kathryn L. Cottingham

Abstract

Under the current global biodiversity crisis, there is a need for automated and noninvasive monitoring techniques that can gather large amounts of data cost-effectively at various ecological scales, from local to large spatial scales. These data can then be analyzed to inform stakeholders and decision-makers. One such technique is passive acoustic monitoring, which is commonly coupled with automatic identification of animal species based on their sound. Automated sound analyses usually require the training of sound detection and identification algorithms. These algorithms are based on annotated acoustic datasets which mark the occurrence of sounds of species inside sound recordings. However, compiling large annotated acoustic datasets is time-consuming and requires experts, and therefore, they normally cover reduced spatial, temporal, and taxonomic scales. This data paper presents WABAD, the World Annotated Bird Acoustic Dataset for passive acoustic monitoring. WABAD is designed to provide the public, the research community, and conservation managers with a novel and globally representative annotated acoustic dataset. This database includes 5047 min of audio files annotated to species-level by local experts with the start and end time and the upper and lower frequencies of each identified bird vocalization in the recordings. The database has a wide taxonomic and spatial coverage, including information on 91,931 vocalizations from 1192 bird species recorded at 72 recording sites in 29 recording locations (mainly countries) and distributed across 13 biomes. WABAD can be used, for example, for developing and/or validating automatic species detection algorithms, answering ecological questions, such as assessing geographical variations on bird vocalizations, or comparing acoustic diversity indices with species-based diversity indices. The dataset is published under a Creative Commons Attribution 4.0 International license that permits redistribution and reuse on the condition that the original work is properly credited.

KEYWORDS
animal vocalizations, automated sound recorder, autonomous recording units, birds, human expert annotation, passive acoustic monitoring, song, soundscape

AFFILIATIONS

[1]Biodiversity Conservation and Management Programme, Forest Science and Technology Center of Catalonia (CTFC), Lleida, Spain
[2]Department of Ecology, University of Alicante, Alicante, Spain
[3]EFNO, INRAE, Nogent-sur-Vernisson, France
[4]Programa de Biología, Universidad del Quindío, Armenia, Quindío, Colombia
[5]Department of Plant Biology and Ecology, University of Sevilla, Seville, Spain
[6]Doñana Biological Station (CSIC), ICTS-Doñana (Infraestructura Científico-Técnica Singular de Doñana), Seville, Spain
[7]Complex Systems Group, Department of Agroforestry Engineering, ETSIAAB, Universidad Politécnica de Madrid, Madrid, Spain
[8]Institute of Marine Research (INMAR), Centro CEIMAR, International Campus of Excellence in Marine Science (CEIMAR), Universidad de Cádiz, Cádiz, Spain
[9]Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland
[10]Blue Nature Birding and Nature Tours SL, Madrid, Spain
[11]Biodiversity Node, Sector Foresta, Madrid, Spain
[12]Eurofins MITOX B.V., Amsterdam, Netherlands
[13]SEO/BirdLife, Madrid, Spain
[14]Instituto de Investigación en Recursos Cinegéticos, IREC-CSIC-UCLM-JCCM, Ciudad Real, Spain
[15]Iduna Tours, Cáceres, Spain
[16]Terrestrial Ecology Group (TEG-UAM), Departamento de Ecología, Universidad Autónoma de Madrid, Madrid, Spain
[17]Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain
[18]Department of Biology, University of Hawai‘i at Hilo, Hilo, Hawai‘i, USA
[19]Observatorio de Biodiversidad del Bosque Atlántico, Instituto de Biología Subtropical, CONICET-Universidad Nacional de Misiones, Puerto Iguazú, Argentina
[20]Laboratório de Ornitologia e Bioacústica, Universidade Estadual de Londrina, Londrina, Brazil
[21]ConservaSom, João Pessoa, Paraíba, Brazil
[22]Programa de Pós-graduação em Biodiversidade e Evolução, Instituto de Biologia, Universidade Federal da Bahia, Salvador, Brazil
[23]Laboratório de Herpetologia e Comportamento Animal, Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Campus Samambaia, Goiânia, Goiás, Brazil
[24]BUTEO Environmental Initiatives, Mojados, Valladolid, Spain
[25]Department of Life Sciences, Imperial College London, London, UK
[26]Forest Research Institute, ELGO-DIMITRA, Thessaloniki, Greece
[27]Station d’Écologie Théorique et Expérimentale, SETE, CNRS, Moulis, France
[28]Dynafor, INRAE, University of Toulouse, Castanet-Tolosan, France
[29]Université Paul Sabatier - Toulouse III, Toulouse, France
[30]Geobotany, Faculty of Biology, University of Freiburg, Freiburg, Germany
[31]Red Biología y Conservación de Vertebrados, Instituto de Ecología, A.C., Veracruz, México
[32]Nature Conservation Unit, Frederick University, Nicosia, Cyprus
[33]Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus
[34]School of Aquatic and Fishery Sciences, College of the Environment, University of Washington, Seattle, Washington, USA
[35]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA
[36]Swiss Ornithological Institute, Sempach, Switzerland
[37]V.N.Karazin Kharkiv National University, Kharkiv, Ukraine
[38]Environment and Climate Change Canada, Québec, Canada
[39]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Autónoma de México, Morelia, Michoacan, Mexico
[40]Kiekari Bird Observatory, KiekariTerra A.C., Xico, Veracruz, Mexico
[41]Grupo de Ecología y Evolución de Vertebrados, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia
[42]Red de Biología Evolutiva, Instituto de Ecología, A.C., Xalapa, Veracruz, México
[43]Proyecto Atlapetes, Antioquia, Colombia; Laboratorio de Ecología Evolutiva y Urbana, Universidad del Norte, Barranquilla, Colombia
[44]Centre for Conservation and Restoration Science, School of Applied Sciences, Edinburgh Napier University, Sighthill Campus, Edinburgh, UK
[45]EcoAcoustic Research Hub, Biosciences Center, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal, Rio Grande do Norte, Brazil
[46]Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande do Norte, Campus Universitário, Lagoa Nova, Natal, Rio Grande do Norte, Brazil

[47]Department of Microbiology and Parasitology, Biosciences Center, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil
[48]Royal Botanic Gardens, Richmond, Surrey, UK
[49]Gothenburg Global Biodiversity Centre, Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden
[50]Department of Biology, University of Oxford, Oxford, UK
[51]Severtsov Institute of Ecology and Evolution of Russian Academy of Sciences, Moscow, Russian Federation
[52]Southern Branch of the Joint Vietnam-Russia Tropical Science and Technology Research Center, HoChiMinh City, Vietnam
[53]School of Science, Engineering and Environment, University of Salford, Manchester, UK
[54]Graduate Program in Biodiversity and Environment, Pontifical Catholic University of Minas Gerais, Belo Horizonte, Brazil
[55]CE3C-Centre for Ecology, Evolution and Environmental Changes, CHANGE-Global Change and Sustainability Institute & Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
[56]SPEA - Portuguese Society for the Study of Birds, Lisbon, Portugal
[57]Computational Bioacoustics Research Unit (CO.BRA), Institute for Science and Technology in Wetlands (INAU), Federal University of Mato Grosso (UFMT), Cuiabá, Brazil
[58]Ornithology, Zoological Research Museum A. Koenig (ZFMK), Bonn, Germany
[59]Scotland's Rural College, Bucksburn, Aberdeen, UK
[60]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan
[61]Department of Geosciences and Natural Resource Management, University of Copenhagen, Frederiksberg C. Copenhagen, Denmark
[62]Centro de Datos para la Conservación, Centro de Estudios Conservacionistas, Universidad de San Carlos de Guatemala, Guatemala, Guatemala
[63]CREAF, Cerdanyola del Vallès, Spain
[64]CSIC, Cerdanyola del Vallès, Spain
[65]Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda, Québec, Canada
[66]Ecology, Evolution, Environment, and Society, Dartmouth College, Hanover, New Hampshire, USA
[67]US Department of Interior, National Park Service, Northeast Temperate Network, Woodstock, Vermont, USA
[68]Biology Department, Bradley University, Peoria, Illinois, USA
[69]Programa de Pós- Graduação em Ecologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, Brazil
[70]Laboratório de Ornitologia, Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande do Norte, Natal, Brazil
[71]Guangxi Key Laboratory of Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning, Guangxi, China
[72]Department of Health and Environmental Science, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China
[73]Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland
[74]Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", University of Alicante, Alicante, Spain

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Metadata are available as Supporting Information. Due to size, the complete audio files, their annotations, and associated metadata are available in Zenodo at https://doi.org/10.5281/zenodo.17293588.

ORCID

Cristian Pérez-Granados https://orcid.org/0000-0003-3247-4182
Jon Morant https://orcid.org/0000-0001-5702-2348
Oscar H. Marín-Gómez https://orcid.org/0000-0002-0743-8350
Miguel A. Muñoz-Mohedano https://orcid.org/0000-0001-6114-4460
Gerard Bota https://orcid.org/0000-0001-9020-7272
Adrián Barrero https://orcid.org/0000-0002-2980-1202
Luc Barbaro https://orcid.org/0000-0001-7454-5765
Sandra Müeller https://orcid.org/0000-0003-4289-755X
Christos Mammides https://orcid.org/0000-0003-1747-175X
Gabriel Marcacci https://orcid.org/0000-0002-2199-0141
Diego Varela https://orcid.org/0000-0003-3123-6756
Gustavo A. Zurita https://orcid.org/0000-0002-1341-3491
Junior A. Tremblay https://orcid.org/0000-0003-4930-0939
Robert A. Briers https://orcid.org/0000-0003-0341-1203

Anamaria Dal Molin https://orcid.org/0000-0003-0633-3888
Karl-L. Schuchmann https://orcid.org/0000-0002-3233-8917
Marinêz I. Marques https://orcid.org/0000-0002-9890-8505
Mao-Ning Tuanmu https://orcid.org/0000-0002-8233-2935
Raiane Vital da Paz https://orcid.org/0000-0001-8951-9463
Mauro Pichorim https://orcid.org/0000-0002-9340-9010
Eben Goodale https://orcid.org/0000-0003-3403-2847
Esther Sebastián-González https://orcid.org/0000-0001-7229-1845

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pérez-Granados, Cristian, Jon Morant, Kevin F. A. Darras, Oscar H. Marín-Gómez, Irene Mendoza, Miguel A. Muñoz-Mohedano, Eduardo Santamaría-García, et al. 2026. "WABAD: A World Annotated Bird Acoustic Dataset for Passive Acoustic Monitoring." *Ecology* 107(2): e70317. https://doi.org/10.1002/ecy.70317

Data Paper published in *Ecology*, an Ecological Society of America journal

Metadata S1

WABAD: A World Annotated Bird Acoustic Dataset for passive acoustic monitoring

Cristian Pérez-Granados; Jon Morant; Kevin F.A. Darras; Oscar H. Marín-Gómez; Irene Mendoza; Miguel A. Muñoz-Mohedano; Eduardo Santamaría-García; Giulia Bastianelli; Alba Márquez-Rodríguez; Michał Budka; Gerard Bota; José M. De la Peña-Rubio; Eladio L. García de la Morena; Manu Santa-Cruz; Pablo de la Nava; Mario Fernández-Tizón; Hugo Sánchez-Mateos; Adrián Barrero; Juan Traba; Tomasz S. Osiejuk; Patrick J. Hart; Amanda K. Navine; Andrés F. Montoya Muñoz; Carlos B. de Araujo; Gabriel L. M. Rosa; Ingrid M. Denóbile Torres; Ana L. Camargo Catalano; Cássio Rachid Simões; Diego Llusia; Manuel B. Morales; Pablo Acebes; Juan A. Medina; Nicholas Brown; Christos Astaras; Ilias Karmiris; Elizabeth Navarrete; Maxime Cauchoix; Luc Barbaro; David Funosas; Dominik Arend; Sandra Müeller; Fernando González-García; Alberto González-Romero, Christos Mammides; Michaelangelo Pontikis; Giordano Jacuzzi, Julian D. Olden, Sara P. Bombaci; Gabriel Marcacci; Alain Jacot; Juan P. Zurano; Elena Gangenova; Diego Varela; Facundo Di Sallo; Gustavo A. Zurita; Andrey Atemasov; Junior A. Tremblay; Vincent Lamarre; Anja Hutschenreiter; Alan Monroy-Ojeda; Mauricio Díaz-Vallejo; Sergio Chaparro-Herrera; Robert A. Briers; Renata Sousa-Lima; Thiago Pinheiro; Wigna C. da Silva; Alice Calvente; Anamaria Dal Molin; Alexandre Antonelli; Svetlana Gogoleva; Igo Palko; Hiếu Vũ Trọng; Marina H. Lage Duarte; Natalia dos Santos Saturnino; Samuel R. Silva; Ana Rainho; Paula Lopes; Karl-L. Schuchmann; Marinêz I. Marques; Ana S. de Oliveira; Nick A. Littlewood; Mao-Ning Tuanmu; Yi-Ru Cheng; Hsuan Chao; Sebastian Kepfer-Rojas; Andrea L. Aguilera; Lluís Brotons; Mariano J. Feldman; Louis Imbeau; Pooja Panwar; Aaron S. Weed; Anant Deshwal; Raiane Vital da Paz; Carlos Salustio-Gomes; Dorgival D. Oliveira-Júnior; Cicero S. Lima-Santos; Mauro Pichorim; Wuyuan Pan, Eben Goodale, Alfredo Attisano, Jörn Theuerkauf, Esther Sebastián-González

Class I. Data Set Descriptors

A. **Data set identity:** A World Annotated Bird Acoustic Dataset for passive acoustic monitoring

B. **Data set identification code:** WABAD

C. Data set description

1. **Originators**: Same as authors

2. **Abstract**:

Under the current global biodiversity crisis, there is a need for automated and non-invasive monitoring techniques that can gather large amounts of data cost-effectively at various ecological scales, from local to large spatial scales. This data can then be analyzed to inform stakeholders and decision makers. One such technique is passive acoustic monitoring, which is commonly coupled with automatic identification of animal species based on their sound. Automated sound analyses usually require the training of sound detection and identification algorithms. These algorithms are based on annotated acoustic datasets which mark the occurrence of sounds of species inside sound recordings. However, compiling large annotated acoustic datasets is time-consuming and requires experts, and therefore they normally cover reduced spatial, temporal and taxonomic scales. This data paper presents WABAD, the World Annotated Bird Acoustic Dataset for passive acoustic monitoring. WABAD is designed to provide the public, the research community, and conservation managers with a novel and globally representative annotated acoustic dataset. This database includes 5,047 minutes of audio files annotated to species-level by local experts with the start and end time, and the upper and lower frequencies of each identified bird vocalisation in the recordings. The database has a wide taxonomic and spatial coverage, including information on 91,931 vocalisations from 1,192 bird species recorded at 72 recording sites in 29 recording locations (mainly countries) and distributed across 13 biomes. WABAD can be used, for example, for developing and/or validating automatic species detection algorithms, answering ecological questions, such as assessing geographical variations on bird vocalisations, or comparing acoustic diversity indices with species-based diversity indices. The dataset is published under a Creative

D. **Keywords/phrases**:

animal vocalisations, autonomous recording units, automated sound recorder, birds, human expert annotation, passive acoustic monitoring, song, soundscape.

Class II. Research origin descriptors

1. **Overall project description:**

Global biodiversity is in crisis, experiencing unprecedented declines in recent decades (Maxwell et al., 2016; Pimm and Raven, 2000). This biodiversity crisis, primarily driven by increasing human activities, poses a significant challenge to conservation and management efforts worldwide (Cowie et al. 2022). As the extent and impact of these human pressures continue to grow, there is an increasing need for monitoring tools that can operate at large spatial and temporal scales to provide accurate data on species distribution, population abundance and trends, and ecosystem health (Henry et al., 2008). Fortunately, technological advances in recent decades have led to the development of several effective, automated, and non-invasive techniques for ecological monitoring, such as environmental DNA analyses, satellite remote sensing, unmanned aerial vehicles, and camera trapping (Lahoz and Magrath 2021). One such tool, whose use has exponentially increased in recent years, is Passive Acoustic Monitoring (PAM, Sugai et al. 2019).

PAM relies on automated (or semi-automated) acoustic sensors that are programmed to record the sounds of the environment (i.e., the ‘soundscape’) either continuously or at specific time intervals, enabling researchers and conservation managers to cover large spatial and temporal scales. PAM has proven to be an effective alternative to traditional survey methods for monitoring vocally active taxa in both terrestrial and aquatic environments (Sugai et al., 2019, Desjonquères et al., 2020). Its applications are diverse, including the detection of threatened or invasive species, monitoring population trends over time, the state of ecosystems or the anthropogenic impacts on animal communities, among others (e.g., Astaras et al. 2017, Gibb et al. 2019, Pieretti and Danovaro 2020, Ross et al. 2023, Alcocer et al. 2023, Bota et al. 2024). The acoustic data collected through PAM can be processed by human visual inspection of spectrograms; however, the large volume of data generated in most projects makes the manual classification of all the data unattainable (Stowell et al., 2019). To address this challenge, the development of deep-learning (DL) algorithms for analysing sound recordings has become a common solution (Kahl et al. 2021, Stowell 2022, Sharma et al. 2023). Training and evaluating DL algorithms for bioacoustics surveys requires annotated acoustic datasets (i.e., datasets in which there are labels indicating what type of sound is present; Stowell 2022, Ventura et al. 2024). Although some annotated, and open or publicly available, acoustic datasets exist, they are often limited in their geographic and taxonomic coverage, reducing their applicability to those sampled areas and species (see e.g., Vidaña-Vila et al. 2017, Gómez-Gómez et al. 2023, Recalde et al. 2023, Jamil et al. 2023, Cañas et al. 2023). However, the growth of DL for ecological applications partly depends on the diversity, quality, and availability of public and standardised annotated datasets.

Birds are the most frequently studied taxon using automated sound recorders, and consequently, most recent advances in bioacoustic surveys have focused on them (e.g., Priyadarshani et al. 2018, Sugai et al. 2019, Kahl et al. 2021, Xie et al. 2023). Indeed, several studies have demonstrated the utility of PAM for estimating bird species richness or inferring population estimates from sound recordings (e.g., Darras et al. 2018a, Pérez-Granados and Traba 2021). Birds also represent the group for which the most extensive bioacoustic resources exist, including publicly available sound repositories such as Xeno-Canto (www.xeno-canto.org), Avibase (Lepage 2021), and the Macaulay Library (www.macaulaylibrary.org). Although these sound archives have enabled the development of high-precision bird recognition models (e.g., BirdNET, Kahl et al. 2021), these portals have not been created as a database for training DL algorithms, and thus their use is partly limited by: i) annotation reliability, ii) the absence of strong annotations (e.g., marking the start and end points of each bird vocalisation in the recording), and iii) the lack of complex soundscapes with overlapping sounds within these benchmark libraries. Therefore, a pre-processing stage and more detailed annotation of the archived recordings in such repositories are needed to train or evaluate DL algorithms (Morfi et al. 2019). However, annotating acoustic datasets with temporal annotations to train or evaluate DL algorithms is a highly time-consuming task that involves a significant amount of manual labour from expert annotators (Morfi et al. 2019), posing challenges for the further implementation of PAM surveys. Moreover, most of the recordings in sound archives have been made with directional microphones, which commonly have higher quality than the omnidirectional ones mounted in most commercially available autonomous recording units (ARUs). Therefore, the performance of the DL algorithms developed using such recordings may decrease when tested with recordings collected using omnidirectional microphones (Wood et al. 2021).

Currently, there are a few high-quality annotated acoustic datasets for birds, but they are limited in their taxonomic scope (e.g., Vidaña-Vila et al. 2017 and Recalde et al. 2023, which cover seven and one bird species, respectively) or geographic coverage (e.g., Lostanlen et al. 2018 and Morfi et al. 2019, which cover flight calls of nocturnally migrating birds recorded in New York and bird vocalisations recorded in Spain and France, respectively, see also Gómez-Gómez et al. 2023; Jamil et al. 2023). Therefore, although birds are the group with the most extensive acoustic resources available, there is a notable lack of high-quality, globally representative, and openly accessible annotated acoustic datasets. This absence represents a significant barrier to progress in this field, preventing researchers from leveraging the full potential of PAM to address pressing questions in avian ecology and conservation. The development of novel acoustic datasets could contribute to the creation of new DL algorithms for automated bird identification or refine existing ones, such as BirdNET (Kahl et al. 2021) or Nighthawk (Van Doren et al. 2023). In fact, although the latest version of BirdNET covers about 6,000 bird species (https://github.com/birdnet-team/BirdNET-Analyzer), there is a need for annotated acoustic datasets to extend its current spatial and taxonomic coverage. Moreover, they could also help answer several other ecological questions, such as assessing geographic variation in bird songs or elucidating whether soundscape indices are related to bird species richness, to name a few.

1. **Identity**: WABAD: A World Annotated Bird Acoustic Dataset for passive acoustic monitoring

2. **Originators**: Same as authors. A detailed list of the responsible for each particular recording site, annotations, associated metadata, and the publications related to each recording site, when available, can be found in Table 1.

Table 1. List of the 72 recording sites, the contact email of the coordinator for each recording site, and the publications associated with each recording site. See the complete references list in section *Publications and Results*. Recording sites are identified by their acronyms. * Datasets with weak annotations (not annotated at vocalisation level).

Recording site	**Contact**	**Reference**
ARD	zuranojp@gmail.com	Gangenova et al. 2025
BAM	tomasz.osiejuk@amu.edu.pl	
BERB*	kevin.darras@inrae.fr	Darras et al. 2018b
BIAL	tomasz.osiejuk@amu.edu.pl	
BMT	klschuchmann@googlemail.com	Pérez-Granados & Schuchmann 2020, 2021
BOLIN	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
BRCAS	yirucheng@gmail.com	
BRE	andresf.montoyam@uqvirtual.edu.co	
BUR	gabriel.marcacci@vogelwarte.ch	
CARI	ohmarin1@uniquindio.edu.co	
CAT	l.brotons@creaf.uab.cat	
CB	mariano.feldman@ctfc.cat	Feldman et al. 2023
CLH	r.briers@napier.ac.uk	
COU	david.funosas-i-planas1@univ-tlse3.fr	Funosas et al. 2024
CRAT*	ohmarin1@uniquindio.edu.co	Marín-Gómez & MacGregor-Fors 2019, Marín-Gómez et al. 2020a
CRUZ	pjhart@hawaii.edu	Hart et al. 2021
DEVA	aattisano@miiz.waw.pl	
DONG	viverricula@gmail.com	
DUNAS	sousalima.renata@gmail.com	Somervuo et al. in press
DYOM	cmammides@gmail.com	Mammides et al. 2025a, 2025b
EFFOR*	kevin.darras@inrae.fr	

Recording site	Contact	Reference
EFFOU*	kevin.darras@inrae.fr	
EMP	raianevital11@gmail.com	
EVROS	christos.astaras@elgo.gr	
FEU	Junior.Tremblay@ec.gc.ca	
FNCA	cabarau@gmail.com	
GLEN	Nick.Littlewood@sruc.ac.uk	
GTLU	skro@ign.ku.dk	
HAG	david.funosas-i-planas1@univ-tlse3.fr	Funosas et al. 2024
HAK	pjhart@hawaii.edu	
HAR	david.funosas-i-planas1@univ-tlse3.fr	
HONDO	esther.sebastian@ua.es	
HUAP	cmammides@gmail.com	Mammides et al. 2025a
JUNCA	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
KAR	michal.budka@amu.edu.pl	
KIB	michal.budka@amu.edu.pl	
LIM	ohmarin1@uniquindio.edu.co	
MABI	pooja.panwar.gr@dartmouth.edu	
MAPIMI	fernando.gonzalez@inecol.mx	
MARTI	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
MILLAN	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
MONTEB	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
MOPU	diego.llusia@uam.es	Santana et al. 2022, García 2022
NAV	gerard.bota@ctfc.cat	Bota et al. 2023
NL	mariano.feldman@ctfc.cat	Feldman et al. 2023
OESF	gioj@uw.edu	
OLIV	diego.llusia@uam.es	Franco et al. 2022, Pérez et al. 2023
OIO	amrainho@ciencias.ulisboa.pt	
PETI	m.h.l.duarte@salford.ac.uk	Duarte et al. 2015

Recording site	Contact	Reference
PGF	aattisano@miiz.waw.pl	
PINA	diego.llusia@uam.es	Medina 2022
PITI	cristian.perez@ctfc.cat	Manzano-Rubio et al. 2022
POZO	irene.mendoza@ebd.csic.es	Marques-Rodríguez et al. 2024, 2025
PUUL	pjhart@hawaii.edu	
QR	a.hutschenreiterdd@gmail.com	
RBA	raianevital11@gmail.com	
RFP	raianevital11@gmail.com	
RGU	raianevital11@gmail.com	
RME	raianevital11@gmail.com	
ROTOK	sbombaci@rams.colostate.edu	Bombaci et al. 2018
SAL	andresf.montoyam@uqvirtual.edu.co	
SBN	fernando.gonzalez@inecol.mx	
SCHF	dominik.arend@bio.uni-freiburg.de	
SCHG	dominik.arend@bio.uni-freiburg.de	Müller et al. 2022
SD	mariano.feldman@ctfc.cat	Feldman et al. 2023
SITH	christos.astaras@elgo.gr	Spatharis et al. 2024
SLOB	atemasov@gmail.com	
SPMCO	mauriciodiazva@gmail.com	Díaz-Vallejo et al. 2023
TAM	ohmarin1@uniquindio.edu.co	
UNI	andresf.montoyam@uqvirtual.edu.co	Marín-Gómez 2022
VER	ohmarin1@uniquindio.edu.co	
VIL	cristian.perez@ctfc.cat	

3. **Period of study**: Recordings were gathered from 2007 to 2024.

4. **Objectives**:

To address a critical gap in the current landscape of bioacoustics datasets, we introduce WABAD – the World Annotated Bird Acoustic Dataset. WABAD is an annotated dataset of bird vocalisations compiled through a collaborative effort that includes bird species annotations. WABAD is designed to provide the public, research community, and conservation managers with a globally representative annotated dataset for bird monitoring. Our primary goal is twofold: (1) to compile a standardised, rigorously annotated, and openly accessible dataset of bird vocalisations from a wide range of habitats and regions; and (2) to provide detailed metadata and standardised annotations that promote a culture of open data and facilitate collaboration and research across diverse applications in ecology, including the development and evaluation of DL algorithms, behavioural and evolutionary studies, and habitat assessments.

5. **Abstract**: Same as above

6. **Sources of funding**: CP-G. acknowledges the support from the project 2021-SGR 00302 funded by Departament de recerca i universitats de la generalitat de Catalunya. ESG received the grant RYC2019-027216-I funded by MCIN/AEI/10.13039/501100011033 and by ESF Investing in your future. BIOMON was funded by the European Union's Horizon Europe programme under grant agreement 101090273. JM was supported by Generalitat Valenciana and European Fund (CIAPOST/199-2022). MDV and SCH received the IdeaWild support DIAZCOLO0720-00. LB has received support from Biodiversa+ TABMON PCI2024-153427 project funded by MCIN/AEI/10.13039/501100011033 and the EU "NextGenerationEU"/PRTR. SCHG and SCHF has been funded by the DFG Priority Program 1374 "Biodiversity- Exploratories" (512414116, 252306891). Doñana datasets (BOLIN, MARTI, MILLAN, MONTEB, JUNCA, and POZO) have received support from BIRDeep project, which is funded by MICIU/AEI/10.13039/501100011033 and the 'European Union

NextGenerationEU/PRTR'. IM received a grant PID2020-115129RJ-I00 from MCIN/AEI/10.13039/501100011033. GB was funded by a grant PTA2021-020336-I from MCIN/AEI/10.13039/501100011033. AMR received a JAE-Intro grant from CSIC (ref. JAEINT23 EX 0243)". MC, DF and LB have received support from ADEME as part of the PSI-BIOM project and from the National Research Agency through a Junior Professor Chair NeoSensation to MC (ANR-23-CPJ1-0174-01), the PARMENIDE project (INRAE Biosefair), the SpatialTreep project (ANR-21-CE03-0002) coordinated by Thierry Feuillet and the Terra Forma project (ANR-21-ESRE-0014). Dataset SBN and MAPIMI have received funding from Proyectos Estratégicos: Conservación de Äreas Naturales Protegidas (2022-2024), and from Red Biología y Conservación de Vertebrados, both from Instituto de Ecología, A. C. AA and JT have received support from Narodowe Centrum Nauki, grant 2022/45/B/NZ8/03740. The project that collected the Dataset DUNAS has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 856506) and RSSL has received a research fellowship from CNPq (process number 311533/2022-1). SITH dataset: data collection was funded within the framework of the Single RTDI State Aid Action "Research – Create – Innovate" (grant T1EΔK-04488) and was co-funded by the European Union's European Regional Development Fund (ERDF) and national funds of the Hellenic Republic (Greece) via the Operational Programme Competitiveness, Entrepreneurship and Innovation 2014-2020 (EPAnEK). MOPU, OLIV and PINA have received support from the Comunidad de Madrid (2020-T1/AMB-20636, Atracción de Talento Investigador, Spain, PI: DLL), a research project (REMEDINAL-TE, P2018/EMT-4338; Madrid Regional Government and EU Social Fund, PI: DLL) and a project of the Fundación Universidad Autónoma de Madrid (066205, FUAM, Spain; PI: MM). Dataset OIO was funded by Portuguese Funds through the FCT – Foundation for Science and Technology, I.P., within the scope of the project EcoPestSuppression (doi:10.54499/PTDC/ASP-AGR/0876/2020) and a PhD Grant to PL (022.14253.BD). Dataset MABI was funded by the USDOI National Park Service I&M program under agreement P21AC10555. Dataset BRCAS received funding from Biodiversity Research Center, Academia Sinica. Datasets BERB, EFFOR, and EFFOU were funded by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the collaborative German—Indonesian research project CRC990. We are grateful to Irfan Fitriawan, Wichyanan Limparungpatthanakij and Iwan Fadlurrahman for the work during the annotation process. Dataset OESF received support from the Washington State Department of Natural Resources. Dataset QR received funding from the Rufford Foundation and DGAPA-UNAM.

B. Specific subproject description

1. Site description

a. **Site type:** Data has been gathered at 72 sites in 29 recording locations (mainly countries, see Figure 1, Table 2). Most of the data (59.7% of the recording sites, n = 43) were collected in the northern hemisphere, with northern data coming mainly from Europe (38.6% of the recording sites, n = 27), North America (14.3% of the total, n = 10), and to a lesser extent, Asia (8.6% of the total, n = 6). Nonetheless, the database includes data from six continents, with several recording sites in Central and South America (28.6% of the total, n = 20, Figure 1).

b. **Geography**: Worldwide. The specific position of the recordings can be found in Figure 1 and Table 2.

c. **Habitat**: The collected data also represents a broad range of biomes. The most represented biome (following Olson et al. 2001) was Tropical and Subtropical Moist Broadleaf Forests (27 sites), while Mediterranean Forests were also widely represented, with 10 sites. The dataset also includes sites in Temperate Broadleaf and Mixed Forests (8 sites), Boreal Forests/Taigas (7 sites), wetlands (7 sites) and Tropical and Subtropical Dry Broadleaf Forests (7 sites). The less represented biomes were Montane Grasslands and Savannas (2 sites), Tropical and Subtropical Grasslands or Savannas and Shrublands (2 sites), Deserts and Xeric Shrublands (1 site) and Temperate Grasslands (1 site). Specific biomes of each recording site can be found in Table 2.

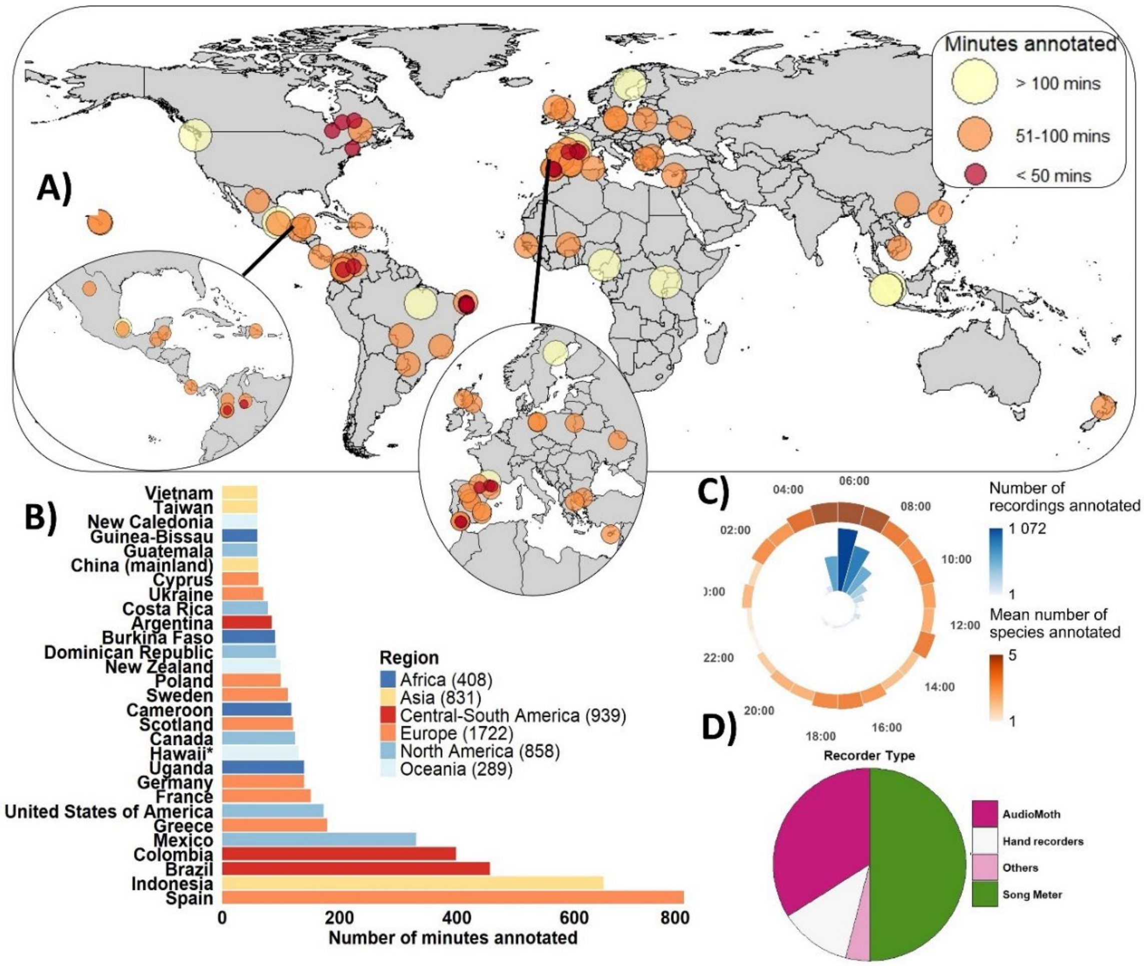

Figure 1: Panel a. Global mapping of recording sites. Colours and size of dots refer to the number of minutes annotated per recording site. The small circles show the position of recording sites in Europe and Central America. Panel b. Number of annotated minutes per recording location and region. Colours of the recording sites refer to different regions. * Hawai'i is part of the United States of America, but it is listed separately since it is biogeographically associated with Oceania. Panel c: Circular histogram detailing the total number of recordings annotated per recording hour and the mean number of species annotated at each recording time. Panel d: Pie chart showing the proportion of different types of audio recorders used across recording sites.

Table 2. Description of the 72 recording sites where the recordings were collected, including the recording site ID, recording site name, recording location (mainly countries), biome (adapted from Olson et al. 2001, BioScience), latitude and longitude (geographic coordinates in decimal degrees).

Site ID	Recording site	Recording location	Biome	Latitude	Longitude
ARD	San Jorge Private Reserve, Misiones province	Argentina	Tropical and Subtropical Moist Broadleaf Forest	-25.7905	-54.2778
BAM	Bamenda Highlands. Big Babanki	Cameroon	Tropical and Subtropical Moist Broadleaf Forest	6.0903	10.3017
BERB	Berbak National Park	Indonesia	Tropical and Subtropical Moist Broadleaf Forest	-1.2837	104.2642
BIAL	Bialowieza forest district	Poland	Temperate Broadleaf and Mixed Forest	52.7453	23.7733
BMT	Pantanal Baia das Pedras	Brazil	Wetland	-16.5139	-56.3936
BOLIN	Torre Palacio. Doñana	Spain	Wetland	36.9905	-6.44262
BRCAS	Tengjhih, Kaohsiung	Taiwan	Tropical and Subtropical Moist Broadleaf Forest	23.0650	120.7636
BRE	Reserva Natural Bremen-La Popa	Colombia	Tropical and Subtropical Moist Broadleaf Forest	4.6753	-75.6136
BUR	Ouagadougou	Burkina Faso	Tropical and Subtropical Grasslands & Savannas and Shrublands	12.5036	-1.7205
CARI	Vereda Caribabare. Tame. Arauca	Colombia	Tropical and Subtropical Dry Broadleaf Forest	6.2831	-71.7778
CAT	Solsona	Spain	Mediterranean Forests. Woodlands and shrublands	41.9919	1.509824
CB	Eeyou Istchee James Bay	Canada	Boreal forest/Taiga	49.3877	-79.1380
CLH	Eastern Scotland	Scotland	Temperate Broadleaf and Mixed Forest	55.9247	-3.2315
COU	Couserans	France	Montane Grasslands and Savannas & Boreal Forests/Taiga	42.7969	1.0767

Site ID	Recording site	Recording location	Biome	Latitude	Longitude
CRAT	Parque Ecologico Macuiltepetl	Mexico	Tropical and Subtropical Moist Broadleaf Forest	19.5489	-96.9203
CRUZ	Las Cruces Biological Station	Costa Rica	Tropical and Subtropical Moist Broadleaf Forest	8.7853	-82.9589
DEVA	Domaine de Déva	New Caledonia	Tropical and Subtropical Dry Broadleaf Forest	-21.5891	165.3762
DONG	Dong Nai Province	Vietnam	Tropical and Subtropical Moist Broadleaf Forest	11.4414	107.4008
DUNAS	Parque Estadual Dunas de	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-5.8118	-35.1917
DYOM	Western Cyprus	Cyprus	Mediterranean Forests. Woodlands and shrublands	34.9193	33.2008
EFFOR	EFForTS Riparian plots	Indonesia	Tropical and Subtropical Moist Broadleaf Forest	-1.8326	103.2019
EFFOU	EFForTS Upland core plots	Indonesia	Tropical and Subtropical Moist Broadleaf Forest	-1.7773	102.9124
EMP	Mata do Jiqui - EMPARN	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-5.8117	-35.1918
EVROS	Evros River Delta National	Greece	Mediterranean Forests. Woodlands and shrublands	40.7971	26.0527
FEU	Labrieville. Québec	Canada	Boreal Forests/Taiga	49.2500	-69.7200
FNCA	Floresta Nacional de Carajás	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-6.179221	-50.156768
GLEN	Glencrippesdale	Scotland	Temperate Broadleaf and Mixed Forest	56.6800	-5.7000
GTLU	Mayan Biosphere Reserve	Guatemala	Tropical and Subtropical Moist Broadleaf Forest	17.4828	-90.0569
HAG	Haute Garonne	France	Temperate Broadleaf and Mixed Forest	43.2161	0.8808
HAK	Hakalau NFWR	Hawaii (USA)	Tropical and Subtropical Moist Broadleaf Forest	19.7942	-155.3216
HAR	Haute Ariège	France	Montane Grasslands and Savannas & Boreal Forests/Taiga	42.6191	1.7644
HONDO	El Hondo Natural Park	Spain	Wetland	38.1826	-0.7544
HUAP	Huaping National Nature	China (mainland)	Tropical and Subtropical Moist Broadleaf Forest	25.5891	109.8936
JUNCA	Juncabalejo. Doñana	Spain	Wetland	36.9361	-6.3783
KAR	Ecopark Käringberget	Sweden	Boreal Forests/Taiga	64.0578	18.6402
KIB	Kibale National Park	Uganda	Tropical and Subtropical Moist Broadleaf Forest	0.5664	30.3725

Site ID	Recording site	Recording location	Biome	Latitude	Longitude
LIM	Los Limones. Monte Plata	Dominican Republic	Tropical and Subtropical Dry Broadleaf Forest	18.9581	-69.7111
MABI	Marsh-Billings-Rockefeller NHP	USA	Temperate Broadleaf and Mixed Forest	43.6312	-72.5179
MAPIMI	Reserva de la Biosfera Mapimi	Mexico	Deserts and Xeric Shrublands	26.6883	-103.7500
MARTI	Caño de Martinazo. Doñana	Spain	Wetland	37.028209	-6.437997
MILLAN	Cancela Millán. Doñana	Spain	Wetland	37.019064	-6.6025
MONTEB	Monteblanco entrada. Doñana	Spain	Mediterranean Forests. Woodlands and shrublands	37.020790	-6.554525
MOPU	Monte Público. Zarzalejo	Spain	Mediterranean Forests. Woodlands and shrublands	40.5270	-4.176811
NAV	Navarra	Spain	Temperate Broadleaf and Mixed Forest	43.0165	-1.7748
NL	Eeyou Istchee James Bay	Canada	Boreal forest/Taiga	51.8760	-76.0650
OESF	Olympic Experimental SF	USA	Temperate Coniferous Forest	47.71518	-124.16975
OIO	Oio Region. Guinea-Bissau	Guinea-Bissau	Tropical and Subtropical Dry Broadleaf Forest	12.3808	-15.2043
OLIV	Ontigola Toledo & Villaconejos	Spain	Mediterranean Forests. Woodlands and shrublands	40.0115	-3.5770
PETI	Estação Ambiental de Peti	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-19.8992	-43.3686
PGF	Parc des Grandes Fougères	New Caledonia	Tropical and Subtropical Moist Broadleaf Forest	-21.6222	165.7612
PINA	Tierra de Pinares. Valladolid	Spain	Mediterranean Forests. Woodlands and shrublands	41.4582	-4.71367
PITI	Pitillas Lagoon	Spain	Wetlands	42.4116	-1.5861
POZO	Pozo de Santa Olalla. Doñana	Spain	Mediterranean Forests. Woodlands and shrublands	36.981868	-6.482437
PUUL	Pu'u Lā'au. Hawai'i	Hawaii (USA)	Tropical and Subtropical Dry Broadleaf Forest	19.8620	-155.5431
QR	Yucatán peninsula (Quintana Roo)	Mexico	Tropical and Subtropical Dry Broadleaf Forest	18.7106	-88.3923

Site ID	Recording site	Recording location	Biome	Latitude	Longitude
RBA	Mata Rio Baldum	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-6.20197	-35.2276
RFP	RPPN Fazenda Pacatuba	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-7.0407	-35.1541
RGU	RPPN REBIO Guaribas	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-6.7178	-35.1818
RME	RPPN Mata Estrela	Brazil	Tropical and Subtropical Moist Broadleaf Forest	-6.3961	-35.0094
ROTOK	Rotokare Scenic Reserve	New Zealand	Temperate Broadleaf and Mixed Forest	-39.4542	174.4091
SAL	Reserva Natural La Montaña	Colombia	Tropical and Subtropical Moist Broadleaf Forest	4.6325	-75.4594
SBN	Santuario del Bosque de Niebla	Mexico	Tropical and Subtropical Moist Broadleaf Forest	19.3453	-96.9392
SCHF	Schorfheide-Chorin	Germany	Temperate Broadleaf and Mixed Forest	52.8731	13.9028
SCHG	Schorfheide-Chorin	Germany	Temperate Grasslands	53.0010	13.9002
SD	Eeyou Istchee James Bay	Canada	Boreal forest/Taiga	52.5440	-72.0780
SITH	Sithonia Penninsula. Chalkidiki	Greece	Mediterranean Forests. Woodlands and shrublands	40.1602	23.7744
SLOB	North-Eastern Ukraine	Ukraine	Temperate Coniferous forest	50.0600	35.2300
SPMCO	San Pedro de los Milagros	Colombia	Tropical and Subtropical Grasslands & Savannas and Shrublands	6.4258	-75.5358
TAM	Bosque el Algarrobo, Tamara.	Colombia	Tropical and Subtropical Dry Broadleaf Forest	5.7533	-72.2108
UNI	Jardín Botánico Univ. del Quindío	Colombia	Tropical and Subtropical Moist Broadleaf Forest	4.5539	-75.6611
VER	Alto Quindío. EL Vergel. Quindío	Colombia	Tropical and Subtropical Moist Broadleaf Forest	4.5914	-75.5908
VIL	Villena shrublands	Spain	Mediterranean Forests. Woodlands and shrublands	38.5603	-0.8778

2. Experimental or sampling design

Audio recordings

The WABAD project started in March 2023 as a collaborative effort. First, the idea of the project was shared with a reduced number of researchers with similar interests, but later, it was shared with researchers working in larger acoustic networks, such as the "Worldwide Soundscapes project" (see Darras et al., 2025) and during the "5th World Ecoacoustic Congress" (Madrid, Spain, July 2024). The data was compiled by E. Sebastián-González and C. Pérez-Granados after an open call for data that was advertised in specialist mailing lists and by using a snowball approach (i.e., asking colleagues working in bioacoustics to share the call with their contacts). Additionally, and after a first preliminary analysis of the collected data and projects , focal publication searches of studies related to annotated acoustic datasets were conducted using Google Scholar and the corresponding authors were offered to join the project, especially for data collected in initially underrepresented regions. However, the spatial distribution depends on the availability of data from local researchers in each area, so our dataset is biased geographically and it cannot be considered representative of the global diversity of species and biomes.

Overall, 5,047 minutes of recordings were gathered during the project (Table 3). The selected sites and audio recordings were chosen by the contributors based on a larger probability of annotating a larger number of vocalisations and bird species (e.g., selected recordings were typically from the early morning hours, when bird activity was highest). Originally, the coordinator of each recording location (Table 1) was instructed to annotate a minimum of 60 minutes per recording site. Although this was not possible for 23 of the recording sites (31.9% of the total), we decided to keep these recording sites in WABAD, since they might be useful for further projects. All acoustic recordings included in WABAD were obtained passively using unattended recorders, which operated on scheduled or continuous recording modes and were not triggered by sound amplitude thresholds. At each site, one or more autonomous recording units were deployed. The devices most used included different versions of the Song Meter recorders from Wildlife Acoustics (48.6% of the total, n = 35) and AudioMoth recorders from Open Acoustic Devices (36.1% of the total, n = 26, Hill et al., 2018), all equipped with omnidirectional microphones. Three other recording sites used different autonomous recording units that were also equipped with omnidirectional microphones (4.2% of the total). Nonetheless, eight recording sites used handheld recorders mounted with directional microphones, although left unattended (with no human presence) in the field to record the soundscapes for several hours (11.1 % of the total, Figure 1 and Table 3). All recordings collected in a recording site always used the same recorder brand (Table 3). When more than one recorder was placed in the same habitat type and separated by less than 30 km, the collected recordings were considered part of a single recording site. In such cases, the site coordinate was estimated as the centroid of the recording site. Although this simplification may not reflect the exact geometric center, it provides a consistent spatial reference for data aggregation and analysis. There was not a minimum requirement regarding the temporal separation of the annotated files to be included in the project.

The original audio files are provided in .wav format, allowing future users to either process the original audio or resample them. Although the sampling rate varied among some datasets, most of the datasets were recorded using a standard frequency of 44.1 (18 datasets, 25% of the total) or 48 kHz (27 datasets, 37,5% of the total, see Table 3), and all datasets were recorded using a sampling rate high enough to ensure that all bird vocalisations could be recorded. Contributors also provided the metadata regarding their recording site (e.g., geographic coordinates, biome, Table 2) and recording parameters (e.g., sampling rate, recorder model, Table 3).

Table 3. Information on the sampling design for each of the 72 recording sites. We specified the recording site ID, recorder brand (and the microphone when not the default one used), whether the microphone was omnidirectional, the sampling rate, the year of the recordings, the number of minutes annotated at each site, and the number of different recorders within each site ID.

Site ID	**Recorder (+ microphone)**	**Omnidirectional**	**Sampling rate**	**Recording date**	**Min. annotated**	**Recorders**
ARD	Song Meter SM4	Yes	48 kHz	2021	84	17
BAM	Song Meter SM3	Yes	22 kHz	2015	118	1
BERB	SM2BAT+	Yes	22 kHz	2013	180	5
BIAL	Song Meter Mini	Yes	48 kHz	2021	100	1
BMT	Song Meter SM2	Yes	32 kHz	2014, 2021	60	1
BOLIN	AudioMoth	Yes	32 kHz	2023	93	1
BRCAS	Song Meter SM4	Yes	44.1 kHz	2022	60	1
BRE	Marantz PMD 222 + Sennheiser ME 66/k6	No	44.1 kHz	2007	49	2
BUR	AudioMoth	Yes	48 kHz	2022	90	6
CARI	Tascam DR100 MKII + Sennheiser ME 66/k6	No	44.1 kHz	2015	62	1
CAT	Song Meter Mini	Yes	44.1 kHz	2021	59	1
CB	Song Meter SM4	Yes	24 kHz	2018, 2019	35	6
CLH	AudioMoth	Yes	192 kHz	2023	61	1
COU	Song Meter Mini	Yes	24 kHz	2022	13	7

Site ID	Recorder (+ microphone)	Omnidirectional	Sampling rate	Recording date	Min. annotated	Recorders
CRAT	Song Meter SM4	Yes	48 kHz	2017	130	1
CRUZ	Song Meter SM2	Yes	44.1 kHz	2012	78	6
DEVA	AudioMoth	Yes	32 kHz	2024	30	1
DONG	Song Meter SM4	Yes	48 kHz	2021	60	1
DUNAS	AudioMoth	Yes	48 kHz	2023	60	1
DYOM	Song Meter Mini	Yes	48 kHz	2023	62	1
EFFOR	SM2BAT+	Yes	24 kHz	2016	200	3
EFFOU	SM2BAT+	Yes	24 kHz	2016	270	1
EMP	Song Meter SM2	Yes	48 kHz	2022	14	2
EVROS	AudioMoth	Yes	48 kHz	2023	80	4
FEU	BarLT (by Frontier Labs)	Yes	44.1 kHz	2019	69	21
FNCA	AudioMoth	Yes	48 kHz	2022	190	17
GLEN	AudioMoth	Yes	32 kHz	2023	60	5
GTLU	AudioMoth	Yes	48 kHz	2022	60	1
HAG	SM4	Yes	24 kHz	2019-2021	122	62
HAK	Song Meter SM4	Yes	44.1 kHz	2021-2022	70	3
HAR	Song Meter Mini	Yes	24 kHz	2022	16	8
HONDO	Song Meter Mini	Yes	44.1 kHz	2023	72	1
HUAP	Song Meter Mini	Yes	44.1 kHz	2023	61	1

Site ID	Recorder (+ microphone)	Omnidirectional	Sampling rate	Recording date	Min. annotated	Recorders
JUNCA	AudioMoth	Yes	32 kHz	2023	30	1
KAR	Song Meter Mini	Yes	48 kHz	2021	112	1
KIB	Song Meter Mini	Yes	48 kHz	2021	140	1
LIM	Tascam DR100 MKII + Sennheiser ME 66/k6	No	48 kHz	2014	92	1
MABI	SwiftOne	Yes	32 kHz	2022	47	8
MAPIMI	AudioMoth	Yes	48 kHz	2023	64	1
MARTI	AudioMoth	Yes	32 kHz	2023	48	1
MILLAN	AudioMoth	Yes	32 kHz	2023	20	1
MONTEB	AudioMoth	Yes	32 kHz	2023	48	1
MOPU	Song Meter SM4	Yes	44.1 kHz	2018	55	1
NAV	AudioMoth	Yes	48 kHz	2023	58	1
NL	Song Meter SM4	Yes	24 kHz	2018-2019	10	1
OESF	Song Meter Mini	Yes	32 kHz	2023	126	2
OIO	AudioMoth	Yes	48 kHz	2022	60	2
OLIV	AudioMoth	Yes	32 kHz	2020	71	4
PETI	Song Meter SM2+	Yes	44.1 kHz	2012	60	5
PGF	AudioMoth	Yes	32 kHz	2024	30	1
PINA	Song Meter SM4	Yes	44.1 kHz	2022	60	3

Site ID	Recorder (+ microphone)	Omnidirectional	Sampling rate	Recording date	Min. annotated	Recorders
PITI	AudioMoth	Yes	16 kHz	2022	48	1
POZO	AudioMoth	Yes	32 kHz	2023	69	1
PUUL	Song Meter SM4	Yes	44.1 kHz	2017	60	2
QR	AudioMoth	Yes	48 kHz	2022	65	5
RBA	AudioMoth	Yes	48 kHz	2022	15	2
RFP	Song Meter SM2	Yes	48 kHz	2022	31	3
RGU	Song Meter SM2	Yes	48 kHz	2022	13	1
RME	Song Meter SM2	Yes	48 kHz	2022	13	3
ROTOK	Olympus DM-620	No	320 kbps	2017	99	28
SAL	Marantz PMD 222 + Sennheiser ME 66/k4	No	48 kHz	2009	57	1
SBN	AudioMoth	Yes	48 kHz	2023	72	1
SCHF	AudioMoth	Yes	48 kHz	2023	60	3
SCHG	Soundscape Explorer Terrestrial (Lunilettronik Cooperativa)	Yes	48 kHz	2016	80	5
SD	Song Meter SM4	Yes	24 kHz	2018-2019	10	2
SITH	Song Meter Mini	Yes	48 kHz	2022	99	1
SLOB	Tascam + Sennheiser ME 66/k6	No	44.1 kHz	2023	70	1
SPMCO	AudioMoth	Yes	32 kHz	2021	62	1

Site ID	**Recorder (+ microphone)**	**Omnidirectional**	**Sampling rate**	**Recording date**	**Min. annotated**	**Recorders**
TAM	Tascam DR100 MKII + Sennheiser ME 66/k6	No	44.1 kHz	2015	29	1
UNI	AudioMoth	Yes	44.1 kHz	2023	83	1
VER	Marantz PMD 222 + Sennheiser ME 66/k5	No	44.1 kHz	2008	57	1
VIL	Song Meter Mini	Yes	22 kHz	2023	56	1

Audio Annotation

For data standardisation, we developed a detailed annotation protocol. The protocol was designed to provide strong annotations, although in a few datasets, the participants provided weak annotations (Table 1) - i.e., they annotated bird vocalisations with long annotations that encompassed multiple bird vocalisations but also silences in the recording. Nonetheless, most datasets provided strong annotations (95% of the total, see specific datasets with strong labels in Table 1). Annotations were generated by expert ornithologists, familiar with the local avifauna, who were trained - if required - and assisted in using the annotation software. All people involved during data collection and audio annotation processes, on each recording site, were invited to participate as co-authors in this project. The ornithologists examined the audio file spectrograms and provided annotations for bird vocalisations by selecting the spectral view and auditory inspections in Raven Pro 1.6. (Bioacoustic Program 2024), using the default configuration parameters (Window type= Hann, DFT size= 512 samples, brightness= 50, contrast = 50). However, the experts were free to adjust the parameters of the spectral view at their convenience for the species identification. We did not discriminate between the different vocalisation types of a bird species (i.e., song or call), and just provided the identification of the species vocalising.

The audio annotation process consisted of opening a complete audio file in Raven Pro 1.6, and then marking with a box the portion of the file where a bird vocalisation occurred. In four recording sites the coordinators used different audio software but also marked with a box the portion of the file with a bird vocalizing. These annotations were then transformed to Raven Pro format. The box was limited such that its extent included the whole bird vocalisation (without harmonics if possible) and mark the starting and ending time (± 0.2 s of error), as well as the maximum and minimum frequency of each vocalisation (i.e., strong annotation). Two bird vocalisations of the same species were allowed to be annotated together when they were separated by less than one second; otherwise, a new annotation was created. Separate annotations were created for overlapping calls of different species. We did not apply any quality criteria (e.g., considering only sounds of high signal-to-noise ratio or high amplitude) to include bird vocalisations in WABAD; the only criterion was that the expert was sure about the species' identity. The annotation protocol was designed so that each recording was annotated by a single expert ornithologist. Nonetheless, we acknowledge that even experienced annotators may occasionally disagree on the species identity of a given sound (e.g. van Osta et al. 2023). Therefore, WABAD may contain potential annotation errors inherent to the process, which should be considered when using the dataset for training or evaluating machine learning models. Sounds uttered by other taxa (i.e., non birds), humans, or human activities were not included. Nonetheless, we would like to encourage further efforts on compiling annotated soundscapes to include vocalisations of a wider range of taxa, as well as to annotate, with a different category, uncertain vocalisations. For each vocalisation, the annotators provided the scientific name of the bird species following the Clements list (Clements et al., 2021). Such nomenclature is the same as the one used in BirdNET v2.4 (Kahl et al., 2021), the most updated version at publishing time, which will facilitate further use of our data. Once the annotation of the entire audio file was completed, the annotations were exported in .txt format with the exact same name as the audio file.

The total number of annotations (i.e., vocalisations) and different species annotated per region (as shown in Figure 1b) is as follows: Africa, 6,455 annotations and 169 species annotated; Asia, 8,990 annotations and 179 species annotated; Europe, 29,766 annotations and 182 species annotated; North-Central America, 16,174 annotations and 219 species; Oceania, 8,867 annotations and 71 species annotate; and South America, 21,679 annotations and 432 species annotated.

Data Preprocessing

To standardise the data generated by the different contributors, we took the following steps. First, we checked and modified, whenever required, the file names to provide exactly the same names (besides the extension format) for both recordings and annotations, as in the following example:

"*SITE*"_"*Date*"_"*Time*".wav

where "*SITE*" is a 3-6 letter acronym (all in capital letters) identifying the recording site and whose metadata information can be found in Tables 2 and 3. "*Date*" refers to the year, month, and day when the recording was collected, and "*Time*" refers to the hour (24h format, local time of each site), minute and second of the starting time of the recording. For example, the file "HONDO_20231104_170500.wav", identifies a recording (due to *.wav* format) made in the HONDO site, recorded on the 4 of November 2023 whose starting recording time was five minutes past five of the afternoon. Hours are always expressed as local time. Likewise, the file "HONDO_20231104_170500.txt" refers to the audio annotations made to the recording cited above.

Class III. Data set status and accessibility

A. **Status**

1. **Latest update:** 8 October June 2025
2. **Latest archive date:** 8 October 2025
3. **Metadata status:** 8 October 2025
4. **Data verification:** 8 October 2025

B. **Accessibility**

1. **Storage location and medium:** The original dataset, along with any updated versions and complementary material, is too large to be included as supporting information with the submission, but can be freely accessed in Zenodo at https://doi.org/10.5281/zenodo.17293588. The data are provided for public use and can be used for research purposes. The vocalizations annotated in WABAD can also be accessed, as 3-second segments, in Hugging Face at https://huggingface.co/datasets/DBD-research-group/WABAD.
2. **Contact persons:** Cristian Pérez-Granados[1] & Esther Sebastián-González[2],

   [1] Biodiversity Conservation and Management Programme, Forest Science and Technology Center of Catalonia (CTFC), 25280 Lleida, Spain. cristian.perez@ctfc.cat.

   [2] Department of Ecology. University of Alicante, Ctra. San Vicente del Raspeig sn, 03690 San Vicente del Raspeig, Alicante, Spain. esther.sebastian@ua.es
3. **Copyright restrictions**: The dataset is published under a Creative Commons Attribution 4.0 International license that permits redistribution and reuse on the condition that the original work is properly credited.
4. **Proprietary restrictions**: Please cite the *Ecology* data paper and the Zenodo release accordingly when using this data.

Class IV. Data structural descriptors

A. **Data set file**

1. **Identity:**

The WABAB includes 4,302 raw audio files (5,047 minutes annotated) and their associated 4,302 annotation files. Audio files are provided in .wav format and have not been standardized (i.e. to have the same number of channels or sampling frequency) or transformed, thus allowing future users to either process the original audio files or transform them (e.g., resample, downsample, stereo to mono). Nonetheless, we provide metadata information of each audio file regarding the recording site, type of recorder and recording parameters employed during recording. Audio and annotation folders are organised within zipped recording site folders. Within each recording site folder, there are three subfolders, one folder with the audio recordings ("Recordings") and the other two folders with the audio annotations in two different formats ("Raven Pro annotations" and "Audacity annotations"). We provide the audio annotations in two different formats aiming to allow users to open the annotations in Raven Pro, which is a commercial software (Bioacoustics software 2014), but includes a free lite version (Raven Lite), and in Audacity (https://audacityteam.org/), which is a free audio software. Both software are commonly used by the public and researchers and allow users to create, open and modify annotations of audio files. The WABAD has also four separate files: 1) one .csv file ("Metadata.csv", comma separated with decimals as points) with all information available regarding metadata collected for each recording site, 2) a .txt file ("README.txt"), explaining all the files regarding the metadata collected for each recording site, such as recording schedule, geographic coordinates, biome, or brand of recorders. Such information can be also found in Tables 1, 2, and 3 of the present document. In WABAD there is also 3) a .csv file ("Pooled annotations.csv", comma separated with decimals as points) with all annotations into a single file to facilitate searching annotations, such as looking for a target species, recording site, country, or biome, and 4) a pdf. file ("Species list.pdf") with a list showing the number of annotations and recordings where each species appears in WABAD (species ordered alphabetically).

2. **Size**:

The WABAD comprises 4,302 raw audio files and their associated 4,302 annotation files. A total of 5,047 minutes are annotated, which include 91,931 annotations of 1,192 different bird species belonging to 642 genus. The number of species annotated per recording location (mainly countries) together with the number (and percentage respect to the total) of species not included in BirdNET v2.4. are shown in Table 4. This comparison reveals that a considerable proportion of species included in WABAD are not yet included in BirdNET, especially for five recording locations from Africa, Asia, and Oceania where more than 10% of WABAD species are not yet included, emphasizing the potential of

WABAD to support the development of new species recognition algorithms. The number of annotations greatly varied between species. Indeed, the 100 species with the largest number of annotations (8.4% of the species in WABAD) represent 51,951 out of the 90,662 annotations (57.0% of the total). The number of annotations for the 20 species with the largest number of annotations and two graphs showing how the number of annotations are distributed among species can be found in Figure 2. The complete list with the number of annotations and recordings where each species is annotated can be found in the "*Species list*" file, in Zenodo at https://doi.org/10.5281/zenodo.17293588.

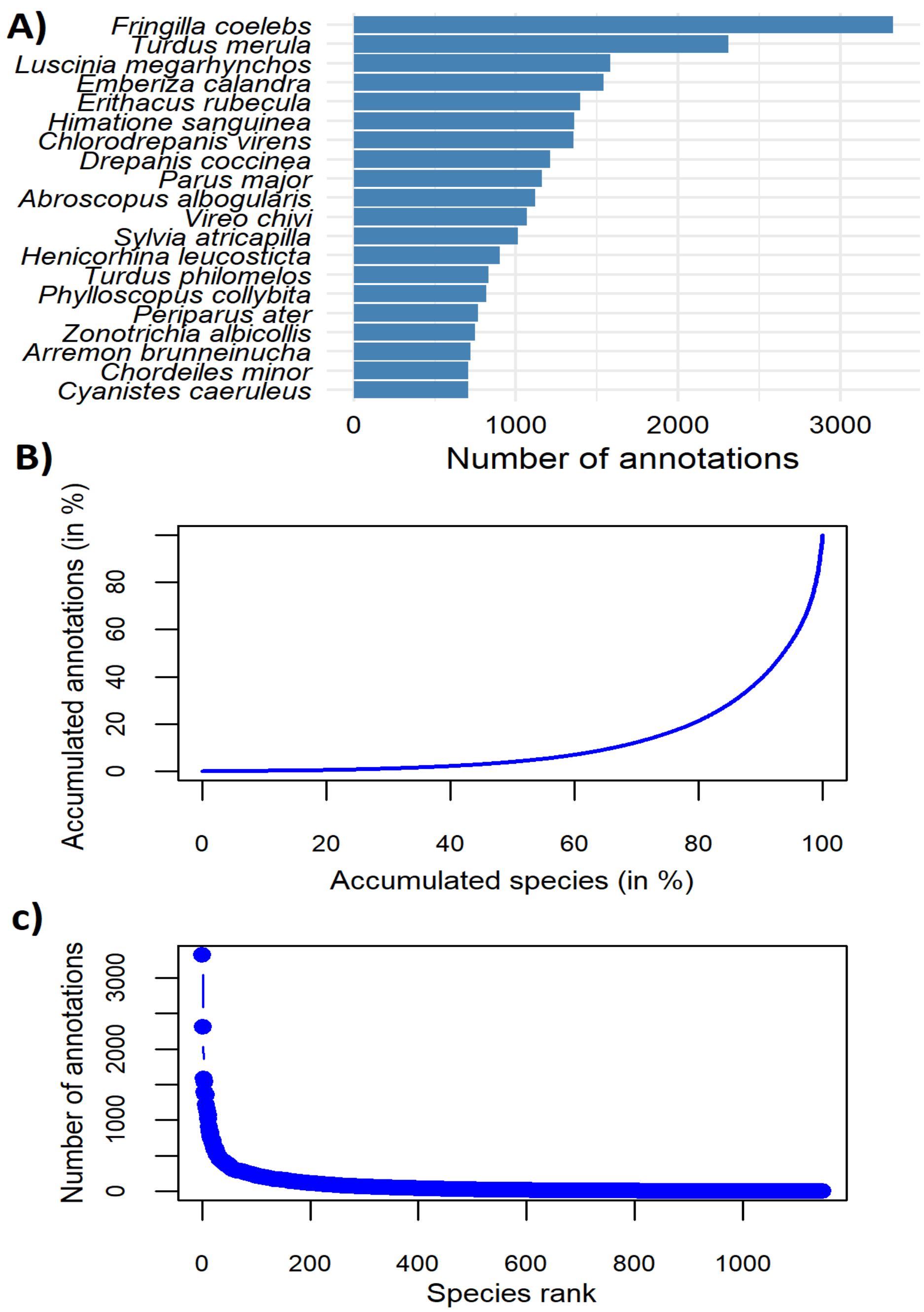

Figure 2. Distribution of the number of annotations across the dataset. Figure 2a shows the number of annotations for the 20 most common species in WABAD. Figure 2b illustrates the cumulative distribution of annotations across species. This curve reveals that a small number of species account for a large proportion of the annotations. The rank-abundance plot (figure 2c) shows the relationship between species rank and the number of annotations. Species are ranked from most to least abundant number of annotations, illustrating a steep decline in the number of annotations as rank increases.

Table 4. Number of annotated species per recording location, grouped by region (in alphabetical order). For each recording location, it also shows the number (and the proportion, in brackets) of annotated species that are not currently included in BirdNET v2.4. Recording locations where more than 10% of the annotated species are not included in BirdNET are shown in bold.

Region	**Recording location**	**Annotated species**	**BirdNET mising (%)**
Africa	Burkina Faso	45	1 (2%)
Africa	**Cameroon**	**42**	**16 (38%)**
Africa	Guinea-Bissau	41	0 (0%)
Africa	**Uganda**	**74**	**10 (14%)**
Asia	**China (mainland)**	**26**	**3 (12%)**
Asia	Taiwan	27	0 (0%)
Asia	Vietnam	61	2 (3%)
Central-South America	Argentina	88	0 (0%)
Central-South America	Brazil	251	2 (1%)
Central-South America	Costa Rica	29	0 (0%)
Central-South America	Colombia	165	3 (2%)
Central-South America	Guatemala	34	0 (0%)
Central-South America	**Dominican Republic**	**15**	**2 (13%)**
Europe	Cyprus	16	1 (6%)
Europe	France	73	0 (0%)
Europe	Germany	45	0 (0%)
Europe	Greece	39	1 (3%)
Europe	Poland	36	0 (0%)
Europe	Scotland	15	0 (0%)
Europe	Spain	138	0 (0%)
Europe	Sweden	26	0 (0%)
Europe	Ukraine	32	0 (0%)
Europe	United Kingdom	24	0 (0%)
North America	Canada	55	0 (0%)
North America	Mexico	83	1 (1%)
North America	United States of America	41	0 (0%)
Oceania	Hawaii	22	0 (0%)
Oceania	**New Caledonia**	**33**	**15 (45%)**

3. **Format and storage mode:**

Each acoustic dataset is uploaded in a separate folder, which includes the audio files (in .wav format) and the corresponding annotation files (.txt format), both in Raven Pro and Audacity formats.

4. **Header information**:

See Variable Information below.

B. **Variable information**

The information of the annotation files varies among whether the annotations were prepared for Raven Pro or Audacity.

Raven Pro annotations contain eight columns with the following information (Figure 3):

<u>Selection</u>: ID of the selection

<u>View</u>: Whether the information comes from the spectrogram or from the waveform

<u>Channel</u>: Channel of origin of the recordings

<u>Begin Time (s)</u>: Start second of the selection

<u>End Time (s)</u>: End second of the selection

<u>Low Freq (Hz)</u>: Lowest frequency of the selection

<u>High Freq (Hz)</u>: Highest frequency of the selection

<u>Species</u>: Species scientific name

The annotations in Raven Pro can be used to add and export additional acoustic parameters according to the users' analytical needs.

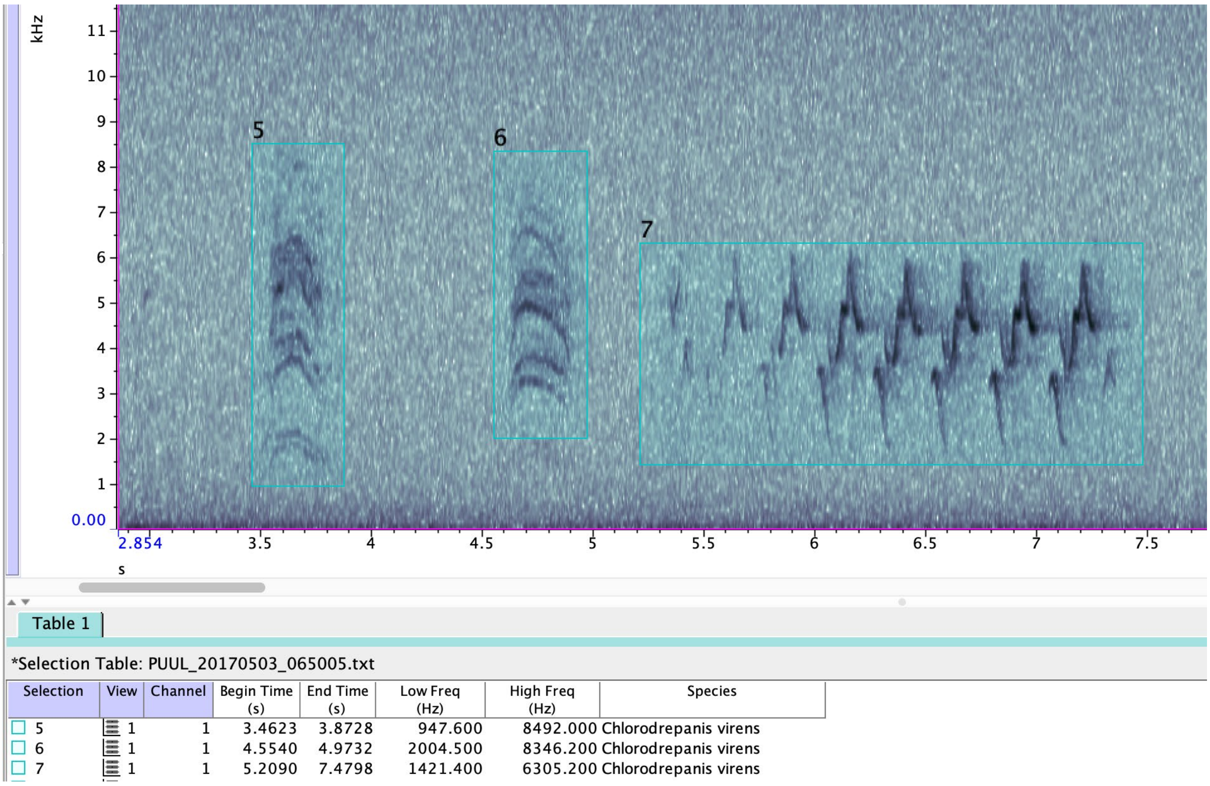

Figure 3. Annotations made with Raven: the top part is a 4 seconds long representation of a spectrogram of the sounds in a recording, whereas the bottom part shows the manual annotations identifying all bird sounds present in the recording. The annotations encapsulate each bird vocalisation based on their frequency and time. Vocalisations of the same species that are less than 1 second apart could be contained by a single annotation. Each annotation contains an ID number (Selection), whether the information is for the spectrogram or the waveform view (View), the begin and end time, the low and high frequencies and the scientific name of the species based on the Clements list (Clements et al. 2021).

Each annotation in Audacity is in two consecutive columns, the first column refers (following the prior nomenclature) to the Begin Time (in s), the End Time (in s), and the Species, and in the next row, which starts with a “\”, is the Low Freq (in Hz) and the High Freq (Hz, see Table 5).

Table 5: Example of the structure of a file containing annotations in Audacity. There are no headings in Audacity format, the first row refers to the time measures (in seconds) and the species scientific name, while the second row shows the frequency limits. This file example (without headings) contains two vocalisations of different species. The first annotation belongs to the species *Nyctibius griseus*, which was annotated from the second 15.142 to 17.650 and from a low frequency of 125.5 Hz to 1933.1 Hz, and a second annotation of *Nyctibius grandis* between the second 39.707 and 42.547 of the recording, and with a low frequency of 500.6 Hz and with a High Frequency of 2184.1.

15.142	17.650	*Nyctibius griseus*
\	125.5	1933.1
39.707	42.547	*Nyctibius grandis*
\	500.6	2184.1

Class V. Supplemental descriptors

A. Data acquisition

1. **Data forms or acquisition methods**: See *Experimental or sampling design* section.
2. **Data entry verification procedures**: To standardise the annotations provided by different annotators, we first verified that the bird species name was added to a new column called “Species” and modified the column name if not. Secondly, we ensured that the scientific names provided by the experts followed the Clements list (Clements et al. 2021) nomenclature. To verify the nomenclature, we used a script in R (v 4.4.0) that compares the annotation names provided by the experts with the list of bird scientific names in Clements et al. (2021). Non-matching species names were checked and changed appropriately.

B. Quality assurance/quality control procedures: The coordinator of each recording site was responsible to check that the data met the requirements described in "Audio Annotation" and "Data Preprocessing" sections. Additionally, a subset of the files sent by each of the collaborators (approximately 30% of the files on each dataset) was opened in Raven Pro by CPG and ESG to ensure compliance with the annotation and file preparation protocols. This review included checking adherence to bounding box protocols, separation of vocalisations separated by more than 1 second, including only bird sounds and the requested variables, proper file formatting, and correct application of species names and metadata. Datasets were included in WABAD once full compliance, on such subset of files, was confirmed. Otherwise, the respective coordinators of each dataset were informed to correct the files, and CPG and ESG re-reviewed the new files to ensure compliance.

C. Computer programs and data-processing algorithms: To verify the nomenclature used by expert annotators and that in Clements et al. (2021), we used a script in R (v 4.4.0) to match both nomenclature and changed it, whenever needed, to follow the Clements list.

D. Archiving

1. **Archival procedures:** The complete audio files, their annotations and associated metadata are too large to be included as supporting information with the submission, but are available in Zenodo at https://doi.org/10.5281/zenodo.17293588.

E. Publications and results:

Part of the recordings of this dataset have been used for the following publications and preprints as listed in Table 1:

Bombaci, S., L. Pejchar, and J. Innes. 2018. Fenced sanctuaries deliver conservation benefits for most common and threatened native island birds in New Zealand. Ecosphere 9:e02497.

Bota, G., Manzano-Rubio, R., Catalán, L., Gómez-Catasús, J., and Pérez-Granados, C. 2023. Hearing to the unseen: AudioMoth and BirdNET as a cheap and easy method for monitoring cryptic bird species. *Sensors* 23:7176.

Darras, K. F. A., Rahman, D., Sugito, W., and others. 2018. Birds of primary and secondary forest and shrub habitats in the peat swamp of Berbak National Park, Sumatra [version 2; peer review: 2 approved]. F*1000Research* 7:229.

Díaz-Vallejo, M., Chaparro-Herrera, S., Lopera-Salazar, A., Castaño-Díaz, M., Correa, R., and Parra, J. L. 2023. Use of acoustic monitoring to estimate occupancy of the Antioquia Brushfinch (*Atlapetes blancae*), a critically endangered species, in San Pedro de los Milagros, Antioquia. *Journal of Field Ornithology* 94:2.

Duarte, M. H. L., Sousa-Lima, R. S., Young, R. J., Farina, A., Vasconcelos, M., Rodrigues, M., and Pieretti, N. 2015. The impact of noise from open-cast mining on Atlantic forest biophony. *Biological Conservation* 191:623–631.

Feldman, M. J., Mazerolle, M. J., Imbeau, L., and Fenton, N. J. 2023. Beaver activity and red squirrel presence predict bird assemblages in boreal Canada. *Ornitholog*y 140:ukad009.

Franco, L, Morales, M. B., Acebes, P., and Llusia, D. 2021. Caracterización de la diversidad acústica de aves en olivares con distinto grado de intensificación agraria. Master Thesis. Universidad Autónoma de Madrid (UAM), Spain.

Funosas, D., Barbaro, L., Schillé, L., Elger, A., Castagneyrol, B., and Cauchoix, M. 2024. Assessing the potential of BirdNET to infer European bird communities from large-scale ecoacoustic data. *Ecological Indicators* 164:112146.

Gangenova, E., Varela, D., Zurano, J. P., Di Sallo, F., Costa, S., De Araujo, C. B., & Zurita, G. A. (2025). Forest corridors preserve biodiversity in tree plantation landscapes of the Southern Atlantic Forest: a multi-taxa approach using passive sampling methods. Forest Ecology and Management, 580, 122522.

García, D. 2019. Evaluación de índices acústicos de diversidad en comunidades de aves y anuros ibéricos. BSc Thesis. Universidad Autónoma de Madrid (UAM), Spain.

Hart, P. J., Ibanez, T., Paxton, K., Tredinnick, G., Sebastián-González, E., and Tanimoto-Johnson, A. 2021. Timing is everything: Acoustic niche partitioning in two tropical wet forest bird communities. *Frontiers in Ecology and Evolution* 9:753363.

Mammides, C., Wuyuan, P., Huang, G., Sreekar, R., Ieronymidou, C., Jiang, A., Goodale, E., Papadopoulos, H. 2025a. The combined effectiveness of acoustic indices in measuring bird species richness in biodiverse sites in Cyprus, China, and Australia. Ecological Indicators 170, 113105.

Mammides, C., Ieronymidou, C., and Papadopoulos, H. 2024b. An ecoacoustic dataset collected on the island of Cyprus in the Mediterranean Basin biodiversity hotspot. Scientific Data, 12: 461.

Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E., and Pérez-Granados, C. 2022. Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. *Ecological Informatics* 72:101910.

Marín-Gómez, O. H. 2022. Artificial light at night drives earlier singing in a neotropical bird. *Animals* 12:1015.

Marín-Gómez, O. H., and MacGregor-Fors, I. 2019. How early do birds start chirping? Dawn chorus onset and peak times in a Neotropical city. *Ardeola* 66:327–341.

Marín-Gómez, O. H., Dáttilo, W., Sosa-López, J. R., Santiago-Alarcon, D., and MacGregor-Fors, I. 2020. Where has the city choir gone? Loss of the temporal structure of bird dawn choruses in urban areas. *Landscape and Urban Planning* 194:103665.

Márquez-Rodríguez, A., Muñoz-Mohedano, M., Martínez-Jiménez, M.J., Santamaría-García, E., Bastianelli, G. & Mendoza, I. (2024). BIRDeepAudioAnnotations (Revision 4cf0456). Hugging Face. DOI: 10.57967/hf/2801

Márquez-Rodríguez, A., Mohedano-Munoz, M. Á., Marín-Jiménez, M. J., Santamaría-García, E., Bastianelli, G., Jordano, P., & Mendoza, I. (2025). A Bird Song Detector for improving bird identification through Deep Learning: a case study from Doñana. arXiv preprint arXiv:2503.15576.

Medina, J.A. 2022. Estudio de la comunicación acústica de aves en el contexto del cambio global. BSc Thesis. Universidad Española a Distancia (UNED), Spain.

Müller, S., Gossner, M.M., Penone, C., Jung, K., Renner, S.C., Farina, A., Anhäuser, L., Ayasse, M., Boch, S., Haensel, F., Heitzmann, J., Kleinn, C., Magdon, P., Perović, D.J., Pieretti, N., Shaw, T., Steckel, J., Tschapka, M., Vogt, J., Westphal, C., Scherer-Lorenzen, M. 2022. Land-use intensity and landscape structure drive the acoustic composition of grasslands. Agriculture, Ecosystems & Environment. 328: 107845.

Pérez, C., Acebes, P., Franco, L., Llusia, D., and Morales, M. B. 2023. Olive grove intensification negatively affects wintering bird communities in central Spain. *Basic and Applied Ecology* 70:27–37.

Pérez-Granados, C., and Schuchmann, K. L. 2020. Monitoring the annual vocal activity of two enigmatic nocturnal Neotropical birds: the Common Potoo (*Nyctibius griseus*) and the Great Potoo (*Nyctibius grand*is). *Journal of Ornithology* 161:1129–1141.

Pérez-Granados, C., and Schuchmann, K. L. 2021. Passive acoustic monitoring of Chaco Chachalaca (*Ortalis canicollis*) over a year: vocal activity pattern and monitoring recommendations. *Tropical Conservation Science* 14:19400829211058295.

Santana, R., Morales, M. B., and Llusia, D. 2019. Optimización del esfuerzo de muestreo acústico automático. Master Thesis. Universidad Autónoma de Madrid (UAM), Spain.

Somervuo, P., Roslin, T., Fisher, B. L., Hardwick, B., Kerdraon, D., Raharinjanahary, D., Rajoelison, E. T., et al. In press. Less predictable global rhythms in human than 15 wildlife contributions to soundscapes. Nature Ecology and Evolution.

Spatharis, D., Tsaligopoulos, A., Matsinos, Y., Karmiris, I., Pleniou, M., Navarrete, E., Boikou, E., Astaras, C. (2024). Monitoring postfire biodiversity dynamics in Mediterranean pine forests using acoustic indices. Environments 11:277.

F. History of data set usage:

Data has not yet been used by any external user.

Acknowledgements

DLL thanks Catalina Pérez and Laura Franco (site OLIV), and Ruben Santana, David García and Alberto Albiero (site MOPU) for their assistance in fieldwork. He also acknowledges a postdoctoral grant provided by the Comunidad de Madrid (2020-T1/AMB-20636, Atracción de Talento Investigador, Spain) and a research project (the REMEDINAL-TE, P2018/EMT-4338; Madrid Regional Government and EU Social Fund). MM acknowledges a project funded by the Fundación Universidad Autónoma de Madrid (066205, FUAM, Spain). Dataset BRCAS thanks the Tengjhih Research Center at the Taiwan Biodiversity Research Institute for device maintenance support. KD acknowledges the Sounds of Life Huma-Num consortium for training support. OHMG was supported by the graduate Grant provided by the National Council of Science and Technology (CONACYT 417094), as well as the Doctoral Program of the Instituto de Ecología, A.C. (INECOL, Xalapa). RSS-L thanks the Brazilian Council for Scientific and Technological Development (CNPq-Brazil) for her research fellowship (process number 311533/2022-1).

Literature Citations

Alcocer, I., Lima, H., Sugai, L. S. M., & Llusia, D. 2022. Acoustic indices as proxies for biodiversity: a meta-analysis. *Biological Reviews* 97:2209–2236.

Astaras, C., Linder, J. M., Wrege, P., Orume, R. D., & Macdonald, D. W. 2017. Passive acoustic monitoring as a law enforcement tool for afrotropical rainforests. *Frontiers in Ecology and the Environment* 15:233.

Bota, G., Manzano-Rubio, R., Fanlo, H., Franch, N., Brotons, L., Villero, D., ... & Pérez-Granados, C. 2024. Passive acoustic monitoring and automated detection of the American bullfrog. *Biological Invasion*s 26:1269–1279.

Cañas, J. S., Toro-Gómez, M. P., Sugai, L. S. M., Benítez Restrepo, H. D., Rudas, J., Posso Bautista, B., ... & Ulloa, J. S. 2023. A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring. *Scientific Data* 10:771.

Clements, J. F., Schulenberg, T. S., Iliff, M. J., Billerman, S. M., Fredericks, T. A., Gerbracht, J. A., Lepage, D., Sullivan, B. L., & Wood, C. L. 2021. *The eBird/Clements checklist of Birds of the World*: v2021.

Cowie, R. H., Bouchet, P., & Fontaine, B. 2022. The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews* 97:640–663.

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S. L., Mulyani, Y. A., & Tscharntke, T. 2018. Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. J*ournal of Applied Ecology* 55:2575–2586.

Darras, K. F., Rountree, R., Van Wilgenburg, S., Cord, A. F., ... & Wagner, T.C. 2025. Worldwide soundscapes: a synthesis of passive acoustic monitoring across realms. Global Ecology and Conservation, 34(5): e70021.

Desjonquères, C., Gifford, T., & Linke, S. 2020. Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments. *Freshwater Biology* 65:7–19.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution* 10:169–185.

Gómez-Gómez, J., Vidaña-Vila, E., & Sevillano, X. 2023. Western Mediterranean Wetland Birds dataset: A new annotated dataset for acoustic bird species classification. *Ecological Informatics* 75:102014.

Henry, P. Y., Lengyel, S., Nowicki, P., Julliard, R., Clobert, J., Čelik, T., ... & Henle, K. 2008. Integrating ongoing biodiversity monitoring: potential benefits and methods. *Biodiversity and Conservation* 17:3357–3382.

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. 2018. AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolutio*n 9:1199–1211.

Jamil, N., Norali, A. N., Ramli, M. I., Shah, A. K. M. K., & Mamat, I. 2023. SiulMalaya: an annotated bird audio dataset of Malaysia lowland forest birds for passive acoustic monitoring. *Bulletin of Electrical Engineering and Informatics* 12:2269–2281.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics* 61:101236.

Lahoz-Monfort, J. J., & Magrath, M. J. 2021. A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience* 71:1038–1062.

Lepage, D. 2021. Avibase-The World Bird Database. URL https://avibase.bsc-eoc.org/checklist.jsp.

Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. 2018. Birdvox-full-night: A dataset and benchmark for avian flight call detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) 266–270.

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. 2016. Biodiversity: the ravages of guns, nets and bulldozers. *Nature* 536:143–145.

Morfi, V., Bas, Y., Pamuła, H., Glotin, H., & Stowell, D. 2019. NIPS4Bplus: a richly annotated birdsong audio dataset. *PeerJ Computer Science* 5:e223.

Pérez-Granados, C., & Traba, J. 2021. Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis* 163:765–783.

Pieretti, N., & Danovaro, R. 2020. Acoustic indexes for marine biodiversity trends and ecosystem health. *Philosophical Transactions of the Royal Society B* 375:20190447.

Pimm, S. L., & Raven, P. 2000. Extinction by numbers. *Nature* 403:843–845.

Ross, S. R. J., O'Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., ... & Burivalova, Z. 2023. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology* 37:959–975.

Priyadarshani, N., Marsland, S., & Castro, I. 2018. Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology* 49:jav-01447.

Recalde, N. M., Estandía, A., Pichot, L., Vansse, A., Cole, E. F., & Sheldon, B. C. 2024. A densely sampled and richly annotated acoustic data set from a wild bird population. *Animal Behaviour* 211:111–122.

Sharma, S., Sato, K., & Gautam, B. P. 2023. A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques. *Sustainability* 15:7128.

Stowell, D. 2022. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* 10:e13152.

Sugai, L. S. M., Silva, T. S. F., Ribeiro Jr, J. W., & Llusia, D. 2019. Terrestrial passive acoustic monitoring: review and perspectives. *BioScienc*e 69:15–25.

Van Doren, B. M., Farnsworth, A., Stone, K., Osterhaus, D. M., Drucker, J., & Van Horn, G. 2024. Nighthawk: acoustic monitoring of nocturnal bird migration in the Americas. *Methods in Ecology and Evolution* 15:329–344.

van Osta, J. M., Dreis, B., Meyer, E., Grogan, L. F., & Castley, J. G. 2023. An active learning framework and assessment of inter-annotator agreement facilitate automated recogniser development for vocalisations of a rare species, the southern black-throated finch (*Poephila cincta cincta*). *Ecological Informatics*, 77, 102233.

Ventura, T. M., Ganchev, T. D., Pérez-Granados, C., de Oliveira, A. G., de S. G. Pedroso, G., Marques, M. I., & Schuchmann, K. L. 2024. The importance of acoustic background modelling in CNN-based detection of the Neotropical White-lored Spinetail. *Bioacoustic*s 33:103–121.

Vidaña-Vila, E., Navarro, J., & Alsina-Pagès, R. M. 2017. Towards automatic bird detection: An annotated and segmented acoustic dataset of seven Picidae species. *Data* 2:18.

Wood, C. M., Kahl, S., Chaon, P., Peery, M. Z., & Klinck, H. 2021. Survey coverage, recording duration, and community composition affect observed species richness in passive acoustic surveys. *Methods in Ecology and Evolution* 12:885–896.

Xie, J., Zhong, Y., Zhang, J., Liu, S., Ding, C., & Triantafyllopoulos, A. 2023. A review of automatic recognition technology for bird vocalizations in deep learning. *Ecological Informatics* 73:101927.