



## Community guidelines to increase the reusability of marine microfossil assemblage data

Lukas Jonkers<sup>1</sup>, Tonke Strack<sup>1</sup>, Montserrat Alonso-Garcia<sup>2</sup>, Simon D'haenens<sup>3,4</sup>, Robert Huber<sup>1</sup>,  
 Michal Kucera<sup>1</sup>, Iván Hernández-Almeida<sup>5</sup>, Chloe L. C. Jones<sup>6</sup>, Brett Metcalfe<sup>7,8</sup>, Rajeev Saraswat<sup>9</sup>,  
 Lóránd Silye<sup>10</sup>, Sanjay K. Verma<sup>11</sup>, Muhamad Naim Abd Malek<sup>12</sup>, Gerald Auer<sup>13</sup>, Cátia F. Barbosa<sup>14</sup>,  
 Maria A. Barcena<sup>2</sup>, Karl-Heinz Baumann<sup>1,15</sup>, Flavia Boscolo-Galazzo<sup>1</sup>, Joeven Austine S. Calvelo<sup>16</sup>,  
 Lucilla Capotondi<sup>17</sup>, Martina Caratelli<sup>18,19</sup>, Jorge Cardich<sup>20</sup>, Humberto Carvajal-Chitty<sup>21</sup>,  
 Markéta Chroustová<sup>22</sup>, Helen K. Coxall<sup>23</sup>, Renata M. de Mello<sup>24</sup>, Anne de Vernal<sup>25</sup>, Paula Diz<sup>26</sup>,  
 Kirsty M. Edgar<sup>27</sup>, Helena L. Filipsson<sup>28</sup>, Ángela Fraguas<sup>29</sup>, Heather L. Furlong<sup>30</sup>, Giacomo Galli<sup>31,32</sup>,  
 Natalia L. García Chapori<sup>33</sup>, Robyn Granger<sup>34</sup>, Jeroen Groeneveld<sup>35</sup>, Adil Imam<sup>36</sup>, Rebecca Jackson<sup>1</sup>,  
 David Lazarus<sup>37</sup>, Julie Meilland<sup>1</sup>, Marína Molčan Matejová<sup>38</sup>, Raphael Morard<sup>1</sup>, Caterina Morigi<sup>32</sup>,  
 Sven N. Nielsen<sup>39</sup>, Diana Ochoa<sup>2</sup>, Maria Rose Petrizzo<sup>40</sup>, Andrés S. Rigual-Hernández<sup>2</sup>, Marina  
 C. Rillo<sup>41</sup>, Matthew L. Staitis<sup>42</sup>, Gamze Tanik<sup>43</sup>, Raúl Tapia<sup>35</sup>, Nishant Vats<sup>44</sup>, Bridget S. Wade<sup>45</sup>, and  
 Anna E. Weinmann<sup>46</sup>

<sup>1</sup>MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, 28359, Germany

<sup>2</sup>Área de Paleontología, Departamento de Geología, Universidad de Salamanca, Salamanca, 37008, Spain

<sup>3</sup>Data Science Institute, UHasselt – Hasselt University, Diepenbeek, 3590, Belgium

<sup>4</sup>Directorate Research, Library and Internationalisation, UHasselt – Hasselt University, Hasselt, 3500, Belgium

<sup>5</sup>PAGES IPO, University of Bern, Bern, 3012, Switzerland

<sup>6</sup>School of Ocean and Earth Science, University of Southampton, Southampton, SO17 1BJ, United Kingdom

<sup>7</sup>Department of Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, 1081 HV, the Netherlands

<sup>8</sup>Laboratory of Systems and Synthetic Biology, Wageningen University & Research,  
 Wageningen, 6700 EJ, the Netherlands

<sup>9</sup>Micropaleontology Laboratory, Geological Oceanography Division,  
 CSIR – National Institute of Oceanography, Goa, 403004, India

<sup>10</sup>Department of Geology, Babeş-Bolyai University, Cluj-Napoca, 400084, Romania

<sup>11</sup>Palaeontology Division, Central Headquarters, Geological Survey of India, Kolkata, 700016, India

<sup>12</sup>Department of Biological Sciences, Faculty of Science, Universiti Tunku Abdul Rahman (UTAR),  
 Kampar 31900, Perak, Malaysia

<sup>13</sup>NAWI Graz Geocenter, Department of Earth Sciences, University of Graz, Graz, 8010, Austria

<sup>14</sup>Departamento de Geoquímica, Universidade Federal Fluminense, Niterói, 24020-141, Brazil

<sup>15</sup>Department of Geosciences, University of Bremen, Bremen, 28359, Germany

<sup>16</sup>Nannoworks Laboratory, University of the Philippines, Quezon City, 1101, Philippines

<sup>17</sup>Istituto di Scienze Marine, Consiglio Nazionale delle Ricerche (CNR-ISMAR), Bologna, 40129, Italy

<sup>18</sup>CONICET – Consejo Nacional de Investigaciones Científicas y Técnicas,  
 Buenos Aires, C1040 AAH, Argentina

<sup>19</sup>Instituto de Investigación en Paleobiología y Geología,

Universidad Nacional de Río Negro and CONICET, General Roca, R8332, Argentina

<sup>20</sup>CIDIS-Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, 15102, Peru

<sup>21</sup>Departamento de Estudios Ambientales, Laboratorio de Bioestratigrafía,  
 Universidad Simón Bolívar, Caracas, 1086, Venezuela

<sup>22</sup>Czech Geological Survey, Prague, 11821 1, Czech Republic

<sup>23</sup>Department of Geological Sciences, Bolin Centre for Climate Research,  
 Stockholm University, Stockholm, 106 91, Sweden

- <sup>24</sup>PETROBRAS S.A. Headquarters, Rio de Janeiro, 20231-030, Brazil
- <sup>25</sup>Geotop and Department of Earth and Atmospheric Sciences, Université du Québec à Montréal, Montréal, H3C 3P8, Canada
- <sup>26</sup>Centro de Investigación Mariña, XM1, Universidade de Vigo, Vigo, 36310, Spain
- <sup>27</sup>School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, United Kingdom
- <sup>28</sup>Department of Geology, Lund University, Lund, 223 62, Sweden
- <sup>29</sup>Department of Biology, Geology, Physics and Inorganic Chemistry and Research Group in Earth Dynamics and Landscape Evolution (DYNAMICAL), University King Juan Carlos, Móstoles, 28933, Spain
- <sup>30</sup>Department of Earth, Atmosphere and Environment, Northern Illinois University, DeKalb, 60115, USA
- <sup>31</sup>Department of Environmental sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, 30172, Italy
- <sup>32</sup>Department of Earth Sciences, University of Pisa, Pisa, 56126, Italy
- <sup>33</sup>IDEAN – Instituto de Estudios Andinos, University of Buenos Aires and CONICET, Buenos Aires, C1428, Argentina
- <sup>34</sup>Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), Den Burg, 1790 AB, the Netherlands
- <sup>35</sup>Institute of Oceanography, National Taiwan University, Taipei City, 10617, Taiwan (R.O.C.)
- <sup>36</sup>Faculty of Sciences Rabat, Mohammed V University in Rabat, Rabat, 1014 RP, Morocco
- <sup>37</sup>Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, 10115, Germany
- <sup>38</sup>Department of Geology and Paleontology, Faculty of Natural Sciences, Comenius University Bratislava, Bratislava, 842 15, Slovakia
- <sup>39</sup>Instituto de Ciencias de la Tierra, Universidad Austral de Chile, Valdivia, 5110209, Chile
- <sup>40</sup>Department of Earth Sciences, University of Milan, Milan, 20133, Italy
- <sup>41</sup>Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, Wilhelmshaven, 26382, Germany
- <sup>42</sup>Faculty of Science, University of East Anglia, Norwich, NR4 7TJ, United Kingdom
- <sup>43</sup>Department of Geological Engineering, Middle East Technical University, Ankara, 06800, Türkiye
- <sup>44</sup>Micropaleontology Laboratory, Department of Applied Geology, Indian Institute of Technology (Indian School of Mines), Dhanbad, 826004, India
- <sup>45</sup>Department of Earth Sciences, University College London, London, WC1E 6BT, United Kingdom
- <sup>46</sup>Department of Geology and Paleontology, Natural History Museum Vienna, Vienna, 1030, Austria

**Correspondence:** Lukas Jonkers (ljonkers@marum.de) and Tonke Strack (tstrack@marum.de)

Received: 9 September 2024 – Revised: 28 January 2025 – Accepted: 2 March 2025 – Published: 28 May 2025

**Abstract.** Data on marine microfossil assemblage composition have multiple applications. Initially, they were primarily used for (chrono)stratigraphy and palaeoecology, but these data are now also widely used to study evolutionary and ecological processes, such as past biodiversity and its links with environmental dynamics, or to provide a basis for conservation efforts and biomonitoring. The large range of potential applications renders microfossil abundance data ideal for reuse. However, the complexity inherent in taxonomic data, which encompass extant and extinct species, coupled with the inherent intricacies of information on biological communities extracted from sedimentary archives, poses considerable hurdles in reusing marine microfossil data, even when they are publicly available. Here, we present guidelines derived from an online survey conducted within the marine micropalaeontological community, aimed at improving the reusability of microfossil assemblage data. These guidelines advocate for clarity and transparency in the documentation of the methods and the outcome, and we outline the data attributes required for effective reuse of micropalaeontological data. These guidelines are intended for researchers who generate microfossil abundance datasets and for reviewers, editors, and data curators at repositories.

A total of 113 researchers evaluated the relevance of about 50 data attributes that might be needed to enable and maximise the reuse of marine microfossil abundance datasets. Each property is ranked based on the survey results. All information is, in principle, considered “desired”. Information that improves the reusability is ranked as “recommended”, and information that is required for reuse is ranked as “essential”. Analysis of a selection of datasets available online reveals a rather large gap between data properties deemed essential by survey participants and what is actually contained in publicly available microfossil assemblage datasets. While the survey

indicates that the micropalaeontological community values good data stewardship, improving data reusability still requires new efforts to incorporate all the essential information. The guidelines presented here are intended as a step in that direction. Determining the optimal forms and formats for data sharing are obvious next steps the community needs to take.

## 1 Introduction

The remains of many groups of marine microorganisms can, under the right conditions, be preserved in sediments. This applies to organisms from different parts of the tree of life, with different ecologies and (ecological) functions. Among these groups, abundant among the extant plankton and benthos, are foraminifera, coccolithophores, diatoms, dinoflagellates, and Radiolaria. Fossil remains of these organisms are abundant in marine sediments, and the resulting record of their diversity and abundance through time extends hundreds of millions of years into the Phanerozoic (Georgescu, 2018). Microfossil occurrence data have hence been successfully applied towards stratigraphic dating since the 19th century (Georgescu, 2018). Soon thereafter, the potential of microfossils for the reconstruction of past environments was recognised (Schott, 1937). Indeed, much of what we know about past climate and the state of the ocean is derived from the analysis of microfossil assemblages (Imbrie and Kipp, 1971). Their ubiquitousness and their spatially and temporally continuous fossil record also render microfossils ideal for studying trends and patterns of evolution (Ezard et al., 2011; Finkel et al., 2005; Lazarus et al., 2014; Lowery et al., 2020; Yasuhara et al., 2020). For the same reasons, microfossil assemblages from sediments can provide information about (past) biodiversity (Yasuhara et al., 2017), which can be used for biomonitoring (Schönfeld et al., 2012), or to establish a baseline of natural biodiversity variability (Jonkers et al., 2019; Smith et al., 2023a) that can also be used to evaluate conservation efforts (Dietl et al., 2015; Yasuhara et al., 2012). Consequently, microfossil assemblages offer crucial information that contextualises the current anthropogenic biodiversity and climate crises within a long-term perspective (Crichton et al., 2023; Hess et al., 2020; Jonkers et al., 2019; Schmidt, 2018).

Microfossil assemblage data are relatively easy to generate in terms of technical needs, but their production requires time and specialised taxonomic expertise. Thanks to the long history of application of microfossils within both academia and industry, a wealth of data exist that are archived and shared to different degrees and in different ways. The value of merging individual datasets into syntheses with extensive coverage in space and/or time has long been recognised, notably for calibrating transfer functions and achieving quantitative environmental reconstructions with sufficient spatial or temporal coverage (CLIMAP project members, 1976). Such synthesis is greatly facilitated through data sharing, as it alleviates

the workload for individual researchers. Thus, data sharing not only increases the transparency of scientific research, but also enables the community to explore new questions and move the field forward (Finnegan et al., 2024; Schiebel et al., 2018; Smith et al., 2023a). Data sharing not only benefits science, but also the individual scientists in the form of community recognition and additional citations (Christensen et al., 2019; Colavizza et al., 2020).

Good scientific practice now stipulates that research data need to be findable, accessible, interoperable, and reusable, or FAIR (Wilkinson et al., 2016), with funding agencies, institutions, and publishers increasingly requiring compliance with these principles. The FAIR principles are not just applicable to data but to other digital (Barker et al., 2022) and real-world components of research (European Commission, Directorate-General for Research and Innovation, 2018). The marine micropalaeontological community has a long history of sharing research data (CLIMAP project members, 1976; Jonkers et al., 2024; MARGO project members, 2009), which is the prerequisite for any further reuse within the framework of FAIR data principles (Wilkinson et al., 2016). However, microfossil assemblage data are complex and, like other palaeodata, require extensive metadata to facilitate their reuse (Khider et al., 2019). The complexity stems in part from taxonomic issues, such as the use of different and evolving taxonomic concepts, or from inconsistent taxonomic practices, and it is exacerbated in palaeontological datasets that include extinct species (Schlagintweit and Simmons, 2022). Consequently, merging datasets generated under different taxonomic frameworks is often challenging, particularly in the absence of visual reference images or when the species concepts are poorly defined. Additional complexity of marine microfossil assemblage data arises from the use of different methodological approaches to acquire them. Thus, in the absence of extensive metadata, marine microfossil data remain difficult to reuse, despite being findable, accessible, and somewhat interoperable.

Moreover, insufficient (capacity for) quality control at data repositories and certain traditions within the field mean that many currently available datasets contain errors that make it difficult to reuse these datasets. For instance, practices such as insufficiently documented taxonomic lumping and reporting relative abundances (instead of raw counts) have led to inaccuracies in many available foraminifera datasets, where species' relative abundances often do not sum to unity (Strack et al., 2023). Although many of these errors can be corrected, this process requires taxonomic expertise and ad-

ditional processing that is not easily automated, hindering reusability.

Data sharing and reuse thus remain hard work for data generators and data reusers. The hurdles to both could be reduced using community-endorsed guidelines that stipulate what and how microfossil data should be reported. Although standards for biodiversity information (e.g. Wieczorek et al., 2012) and palaeoclimate data (Khider et al., 2019) exist, they lack the specificity needed for microfossil data. Some researchers have also proposed guidelines for reporting assemblage data for specific microfossil groups, but they are not comprehensive, as data stewardship was not the focus (Brummer and Kučera, 2022; Schönfeld et al., 2012). The International Ocean Discovery Program (IODP) and its predecessors routinely collected micropalaeontological data on expeditions, providing basic templates for recording taxonomy, (qualitative) abundance, and preservation of different microfossil groups to support consistent data collection and sharing of reusable data. Even though these programmes, in some respects, led the way in increasing the reusability of microfossil (and other) data, the reporting and sharing framework are exclusively used for data collected within these programmes, and microfossil data collected elsewhere remain poorly standardised. In addition, some data synthesis products have used internal (meta)data standards (Fenton et al., 2021; Lazarus, 1994). However, such standards differ among databases and have not (yet) been widely adopted by the marine micropalaeontological community. The limited use of existing standards by the community may be attributed, in part, to their lack of involvement in setting the standards. Guidelines designed and endorsed by the research community are thus needed to increase the reusability of marine microfossil assemblage data.

This article proposes such metadata and data guidelines for marine microfossil assemblage data, based on an extensive survey within the marine micropalaeontological community. The guidelines that result from this survey represent the first step in this process by gathering and analysing the information the community deems necessary for microfossil assemblage data reuse. We see the development of a standard format or the extension of existing formats as a next step in the process of increasing data reusability that would need additional input from the research community. Developing a common format would also need to be accompanied by a standardisation of vocabularies and ontologies or the development of thesaurus(es), a process that also requires community input and the participation of international data repositories. However, it is important to remember changing the format of a dataset is relatively easy as long as it is consistently formatted. In that sense, the exact format in which data are stored presents a smaller hurdle to reusability than datasets that lack important information. This is because changing the format of a dataset can be much more easily automated than sourcing relevant information from published and unpublished sources. Consequently, our guidelines con-

centrate on detailing what information about marine microfossil assemblages should be reported. Presented here are community-endorsed data guidelines that serve data generators and synthesisers alike, aiding in both data synthesis and quality control at data repositories.

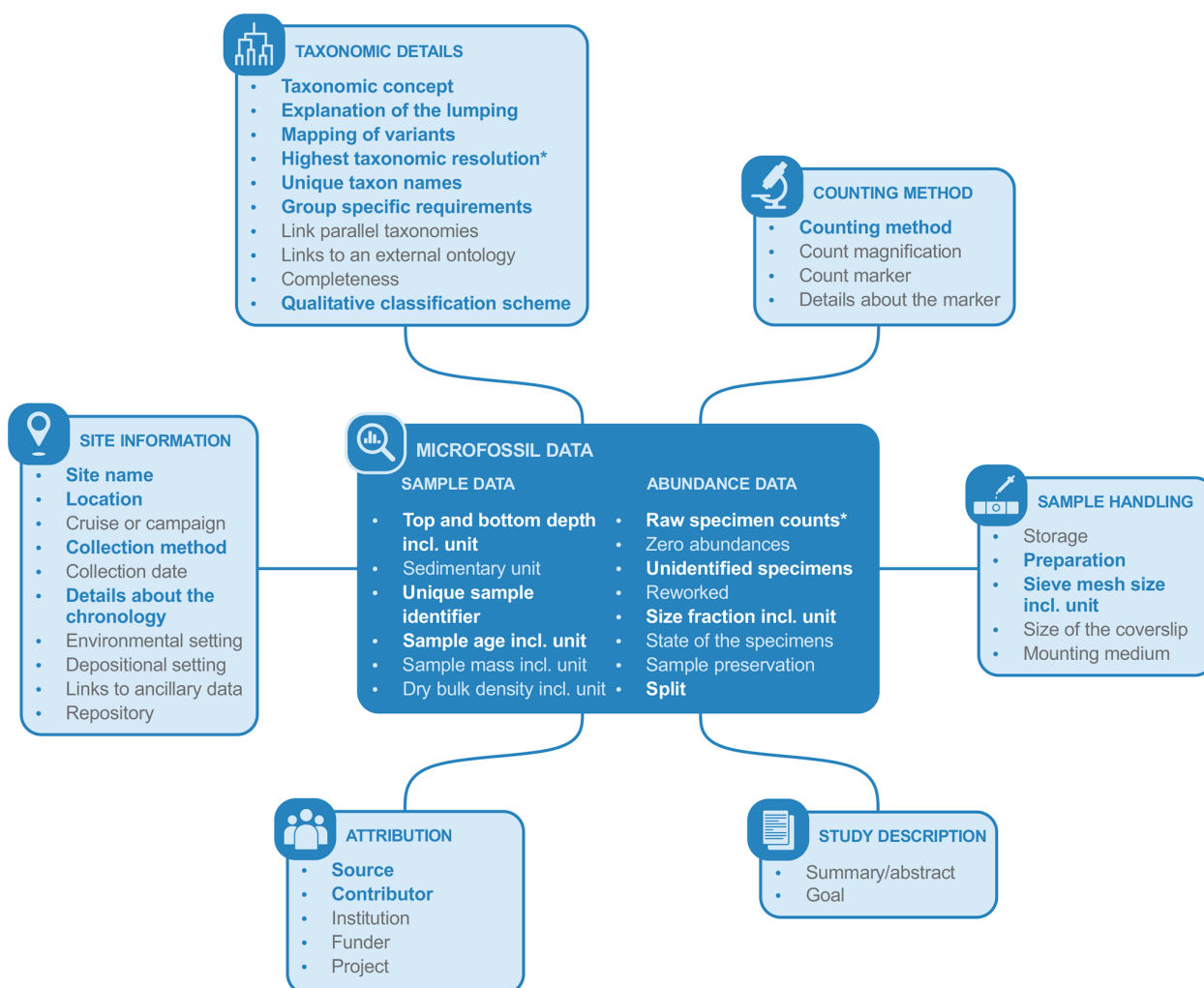
## 2 Methods

To develop guidelines for marine microfossil assemblage data, we designed a survey with input from the (academic) community to assess which information pertaining directly to marine microfossil data is necessary for its reuse. We realise that such data do not exist in isolation and that different aspects of micropalaeontological datasets should be linked when data are shared (Felden et al., 2023; McKay and Emile-Geay, 2016). As for all palaeodata, this holds particularly true for chronological and methodological information. Since clear and community-endorsed guidelines for chronological data exist, for instance, for reporting radiocarbon ages (Millard, 2014), we focus here on developing community-endorsed guidelines specific for the reporting of compositional data on marine microfossil assemblages.

The survey questions were designed by initially asking selected researchers working on different microfossil groups to list types of raw data or data attributes relevant to the interpretation of the assemblage data. The focus on raw data is because of the central aim of the survey to derive guidelines to improve data reusability rather than direct reproducibility, where the latter often requires a description (script) of how inferred variables were derived from the raw data. A total of 49 data attributes potentially relevant for the reuse of microfossil data were defined and subsequently grouped thematically (Fig. 1, Table A1 in Appendix A). At the centre of this grouping are the microfossil data, with their specific information about samples and microfossil abundances. These central attributes may vary among samples. The remaining aspects are specific to the core, the outcrop, or the methodology used for the analyses in a given study and generally do not vary by sample. The survey questions and the (anonymised) responses are freely available (see Data availability statement).

Survey participants were then asked to rank the importance of each of the selected data attributes and presented with two simple yes/no questions. In the survey, questions were accompanied by explanatory text and examples. The ranking scheme of these data aspects follows the categories proposed in previous work (Khider et al., 2019). In principle, all attributes are desired, and attributes that, when lacking, would prevent reuse are classed as essential. Any attribute that increases the value and reusability of these data (or the dataset) are classed as recommended. Participants were asked to consider the importance of data attributes for (hypothetically) searching or filtering datasets. Answering the questions was voluntary, and questions could be left unan-





**Figure 1.** Attributes of marine microfossil data of which the relevance for data reusability was assessed in the survey. Attributes are grouped thematically, with “microfossil data” containing sample-specific information, while the remaining groups contain site-specific information that generally does not vary by sample. Attributes highlighted in blue/white bold font were ranked as “essential” by the survey respondents, and those marked with an asterisk were surveyed using yes/no questions.

swered. Following blocks of thematically organised questions, participants had the opportunity to add free text with comments and suggestions. This led to the addition of a single attribute (split; see below) in the guidelines. The survey ended with questions about the demographics of the participants. Participants were able to remain anonymous if they wanted to.

The survey was implemented in Google and Microsoft forms to allow global access. Researchers were invited to take part in the survey at the Forums 2023 conference, and the survey was advertised through diverse communication channels (e.g. mailing lists, websites), such as Past Global Changes (PAGES; <https://pastglobalchanges.org/>, last access: 14 May 2025), The Micropalaeontological Society (<https://www.tmsoc.org/>, last access: 14 May 2025), the Cushman Foundation for Foraminiferal Research (<https://cushmanfoundation.org/>, last access: 14 May 2025), and the International Nannoplankton Association (<https://ina.tmsoc.org/>, last access: 14 May 2025), as well as personal networks. The survey was open from June to November 2023. Most answers arrived on Wednesdays.

We did not calculate average scores for each evaluated data aspect, as the distance between the three categories is not equal; i.e. the likelihood of voting “recommended” over “desired” is not the same as choosing “essential” over “recommended”. Instead, we assigned an attribute to a single category when it received > 50 % of the votes. This absolute majority threshold was chosen to avoid obscuring patterns in the response (e.g. when the responses were distributed nearly evenly, e.g. 34 %, 33 %, 33 %). When two adjacent rankings together received > 70 % of the votes, aspects were given an intermediate label. The threshold was set to 70 % to reduce

the effect of the relatively low number of responses. The recommended category was assigned in all other cases, i.e. when respondents were divided in their answers in near-equal proportions.

Additionally, we compared results between researchers working on different microfossil groups and between experienced and early career researchers (ERs and ECRs, respectively). The distinction between ERs and ECRs was set at 5 years after obtaining a PhD. We acknowledge that this is only a crude approximation, as it ignores any time spent outside of microfossil-related research. We also note that the delineation among researchers working on different microfossil groups may not be as sharp as suggested here. While all researchers were asked to indicate which group they mainly work on, some may work on multiple groups. As the number of responses from researchers working on groups other than foraminifera was low, we combined the answers from those working on coccolithophores, diatoms, dinoflagellate cysts, ostracods, Radiolaria, and other groups. Thus, we compared the two categories of career stage (i.e. ER and ECR;  $n = 81$  and  $32$ , respectively) across three categories of microfossil groups (i.e. planktonic foraminifera, benthic foraminifera, and other groups;  $n = 44$ ,  $40$ , and  $29$ , respectively). To deal with the ranked and categorical nature of these data, we used adjacent-category logistic regression to test if researchers were more likely to choose “recommended” over “desired” and “essential” over “recommended”. We used this method rather than ordinal logistic regression because tests showed that the assumption of proportional odds does not hold.

We checked five datasets for different taxonomic groups (benthic and planktonic foraminifera, coccolithophores, diatoms, dinoflagellates, ostracods, and Radiolaria) to provide a first-order assessment of the degree to which marine microfossil assemblage data already available in the public domain meet the community guidelines laid out here. We only assessed datasets available on the open-access data-sharing platform PANGAEA (Felden et al., 2023) and selected the first five datasets from the search results. The inclusion of data aspects was evaluated on both standard (meta)data fields and on free text information associated with the files.

### 3 Demographics of the participants

In total, 113 researchers took part in the survey. The median number of answers per question was 112, indicating that most respondents answered the majority of the questions. However, some questions received fewer answers, with a minimum of 95 (Figs. 3–10). The number of respondents and the high proportion of questions answered suggest that our results are likely representative of the views of the marine micropalaeontological community. It should be noted, however, that the high response rate also risks the views of non-experts being imposed on experts, for instance, via answers to questions highly specific to a certain microfossil

group that nearly all survey participants provided. This bias is illustrated clearly in the questions specific to certain taxonomic groups (e.g. about using cyst or motile taxonomy for dinoflagellate cysts), which were answered by considerably more participants than the number of researchers who indicated their primary group of interest as dinoflagellate cysts. However, as the number of microfossil-group-specific questions was low and taxonomic expertise does not equate to expertise in data stewardship, we do not view this as a major concern.

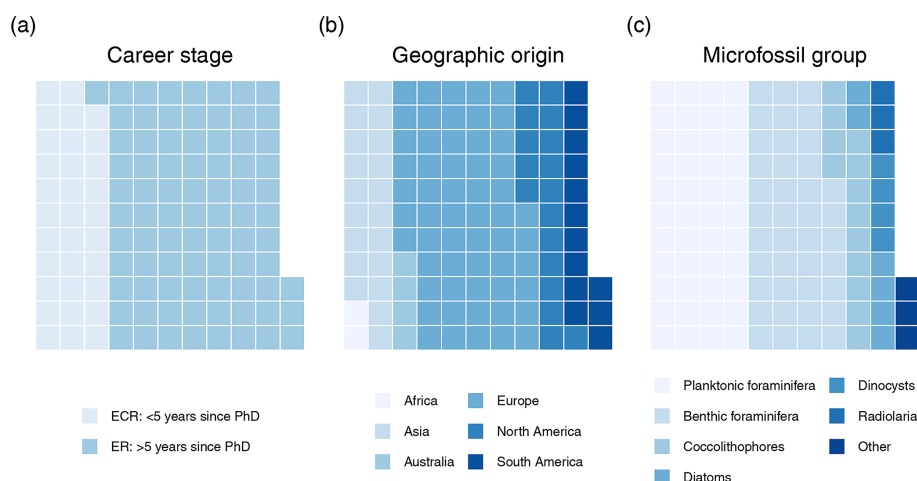
A total of 72 % of the respondents indicated “experienced researcher” as their career stage (Fig. 2). While respondents work on all six inhabited continents, the majority work in Europe (50 %), which may reflect a (networking) bias from the true geographic distribution of the community (Fig. 2). The majority of the respondents primarily work on foraminifera (74 %), with slightly more working on planktonic (39 %) than benthic (35 %) foraminifera. Beside the researchers working on coccolithophores (12 %), all other microfossil groups are represented by fewer than 10 respondents in total (Fig. 2). The dominance of foraminiferal workers within the survey is broadly consistent with (albeit a little higher than) the balance of taxonomic expertise making up the membership of The Micropalaeontological Society (TMS; personal communication with TMS president) and thus may be a true reflection of the community, rather than of the network of the lead authors.

Since the size of the global community working on/with marine microfossil assemblages is unknown, it is impossible to calculate the exact proportion of the community that participated in the survey. We can, however, compare the number of participants to that of a similar survey among palaeoclimate researchers (Khider et al., 2019), which almost certainly constitutes a larger community, given that microfossils can be considered a subset of palaeoclimate research. Approximately 135 respondents took part in the polls and survey that led to the development of the Paleoclimate Community reporting Standard (PaCTS), yet the average answering rate was below 50 % (Khider et al., 2019), which to a large degree reflects the wide scope of PaCTS, where researchers only answered questions within their area of expertise.

It is difficult to establish whether the respondents are an accurate reflection of the entire marine micropalaeontological community. However, the survey data show participants from various career stages and geographic regions, and the major marine microfossil groups are represented. The results presented here thus likely provide a reasonable reflection of the attitudes of the marine micropalaeontological community towards data stewardship.

### 4 Community guidelines

The primary goal of the guidelines is to enhance the reusability of microfossil assemblage data and not explicitly to en-



**Figure 2.** Demographics of the survey respondents ( $n = 113$ ). Each square represents one respondent. ECR: early career researcher; ER: experienced researcher. The microfossil group assignment reflects the primary (taxonomic) expertise of the respondents.

sure reproducibility of particular studies. The guidelines are designed with the anticipation that new datasets will encompass all the information deemed crucial by the community for reuse, thereby eliminating the need for researchers to seek essential metadata and information from other sources, e.g. from publications or, worse, from the original authors. The main focus of the guidelines is to ensure reusability of raw data, and we encourage authors to describe their workflow for derived data in an accompanying paper or associated code.

Below we report on the results of the survey. The assessed data attributes are described here in more detail and cover information about (1) the study site, (2) the original study, (3) attribution, (4) sample handling, (5) the counting method, (6) taxonomic details, (7) sample data, and (8) assemblage data (Fig. 1, Table A1). In this section, the attributes are ranked according to the average voting score within each category. Additional aspects mentioned by multiple respondents in the free text fields are added where applicable. Collectively, the results of the survey represent the core of the guidelines, which target both quantitative and qualitative data on marine microfossil assemblages from sediments.

## 4.1 Site information

This set of information pertains to the details of the study site (core or outcrop) from which the analysed samples were taken. In principle, this information can (should) be linked across multiple datasets from the same archive (Fig. 3).

### 4.1.1 Essential

Location (latitude, longitude, and water depth/elevation including units), collection method, site name, and details about the chronology are all considered to be essential metadata for the archive. Ideally, details regarding the chronol-

ogy should conform to data standards for these types of data and allow reproduction or updating of the chronology. At the minimum, the methodology used for the construction of the chronology should be described.

### 4.1.2 Recommended–essential

Data aspects categorised as recommended–essential are information about the cruise or campaign during which the archive was obtained/sampled, the collection date, the location where the archive is stored (repository; to allow access to the core and/or the samples), and a description of the environmental and depositional settings.

### 4.1.3 Recommended

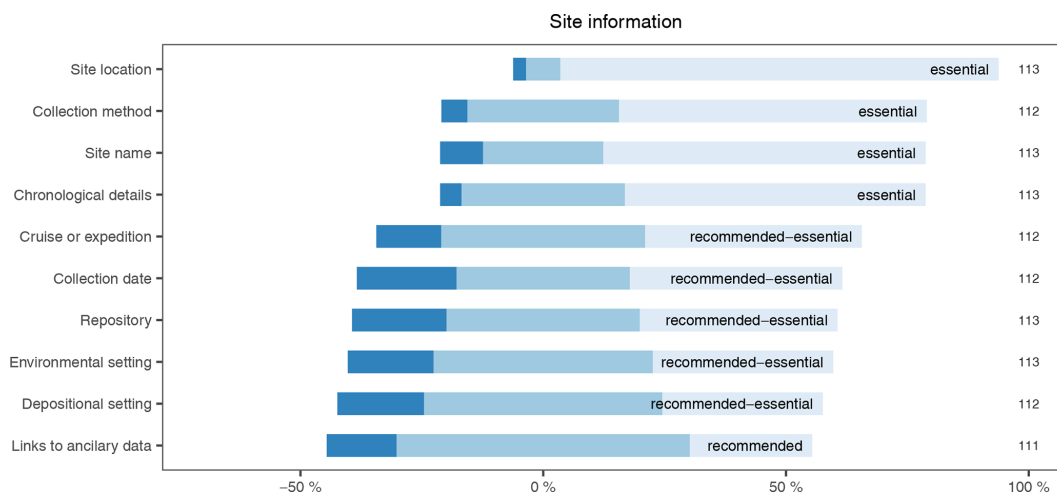
Inclusion of links to ancillary data measured in the same archive that help contextualise the microfossil assemblage data, e.g. radiocarbon ages or foraminifera oxygen isotope ratios, or any other data pertaining to the same archive is recommended if such data are available. Note that some repositories (such as PANGAEA) link datasets from the same site.

## 4.2 Study description

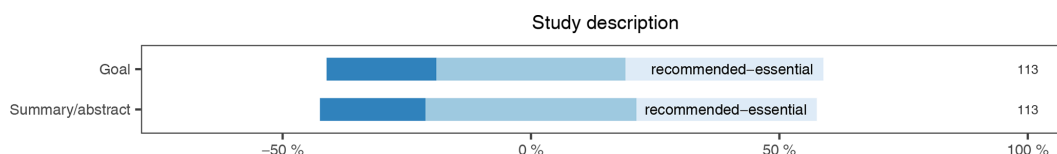
These data aspects explain the context in which data were collected and the main results of the study. Both the goal (reason) of the analysis and a summary of the results or abstract of the study are recommended–essential (Fig. 4).

### 4.3 Attribution

Attribution details are generally not directly relevant for addressing scientific questions (Fig. 5). Still, they are crucial for determining the source of these data and to provide necessary credit to the data generators upon reuse, especially



**Figure 3.** Community-derived guidelines for data properties related to the sampling site (sediment core or outcrop). Survey results are displayed as a Likert diagram showing the proportion of votes for each category (“desired”, “recommended”, and “essential” as dark, mid-, and light blue, respectively) centred on the midpoint of the proportion of votes for the recommended category. The consensus ranking is indicated in each bar, and the number of respondents is shown in grey on the right-hand side.



**Figure 4.** Community ranking of study description information relating to microfossil assemblage data. Colours and layout as in Fig. 3.

when datasets are included in a larger synthesis (Smith et al., 2023b). Attribution information can also be useful to measure the impact or success of research efforts.

#### 4.3.1 Essential

An indication of the source of these data, be it reference to a publication or a persistent identifier (e.g. DOI) for the dataset itself, is essential. Information about the contributor of these data, who can, at least in theory, be contacted if additional information is needed, is also essential.

#### 4.3.2 Recommended

Recommended information is a specification of the institution or laboratory where these data were generated. This information, apart from being relevant to the institute, can also be used to assess the taxonomic concepts used to generate species counts, as taxonomic schools are to some degree laboratory-specific or laboratories may have certain traditions of how to collect data.

#### 4.3.3 Desired–recommended

Information about the project within which these data are generated is ranked between desired and recommended.

#### 4.3.4 Desired

Survey participants ranked information about the data collection funder as desired. This ranking probably reflects the fact that this information is of limited value for scientific reuse of data. However, funding guidelines likely require inclusion of this information.

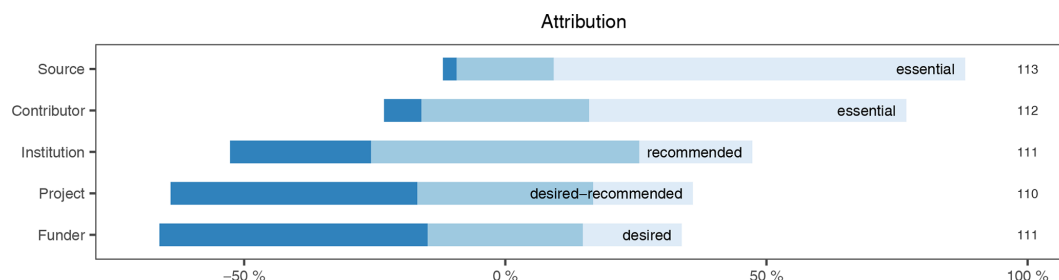
### 4.4 Sample handling

Information on how sediment samples were handled prior to the counting of the microfossils may provide useful information. Assessing the influence of certain preparation methods on the integrity of the samples might help to assess the reliability of the microfossil data to some degree. Survey results regarding sample handling are shown in Fig. 6.

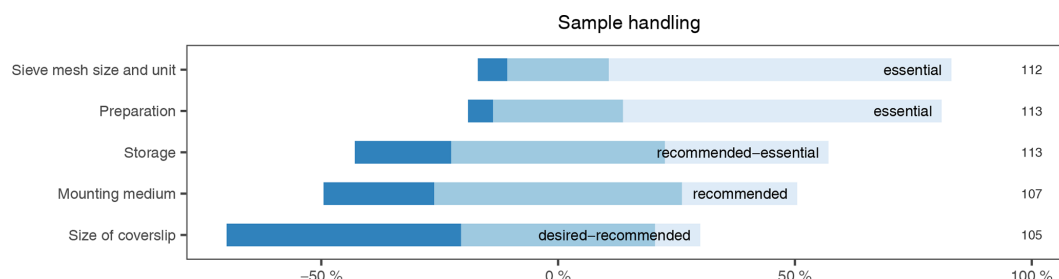
#### 4.4.1 Essential

The initial sieve mesh size (including an SI unit) used in the first step of sample preparation for counting is essential. Note that this size may differ from the actual size fraction of the counted microfossils (see sample data below). A description of the preparation of the sample is also considered to be essential, as the way samples are prepared may affect the microfossils. This information should describe any physical





**Figure 5.** Community ranking of attribution information relating to microfossil assemblage data. Colours and layout as in Fig. 3.



**Figure 6.** Community ranking of information about the handling of samples used for microfossil abundance analysis. Colours and layout as in Fig. 3.

(oven/freeze drying, wet/dry sieving, filtering, settling, centrifuging, etc.) and chemical treatment of the samples (e.g. for sediment dispersal or microfossil isolation). When applicable, this description should also cover (rose bengal) staining to determine the freshness of organic material.

#### 4.4.2 Recommended–essential

As post-sample dissolution may affect (calcareous) microfossil assemblages, a description of how (e.g. at which temperature) and how long samples were stored prior to analysis is recommended–essential (Dunkley Jones and Bown, 2007; Self-Trail and Seefelt, 2005).

#### 4.4.3 Recommended

For samples mounted on coverslips, including information about the mounting medium is recommended.

#### 4.4.4 Desired–recommended

For such samples, the size of the coverslip is desired–recommended.

### 4.5 Counting method

Like information about sample handling, information about the counting method may help to assess the reliability of the assemblage data and provide information useful for merging different datasets. The survey results are summarised in Fig. 7.

#### 4.5.1 Essential

A description of the counting method is essential. The description should, for instance, state what kind of microscope (light microscope, scanning electron microscope, etc.) was used or whether specimens were identified manually or using automated image recognition.

#### 4.5.2 Recommended–essential

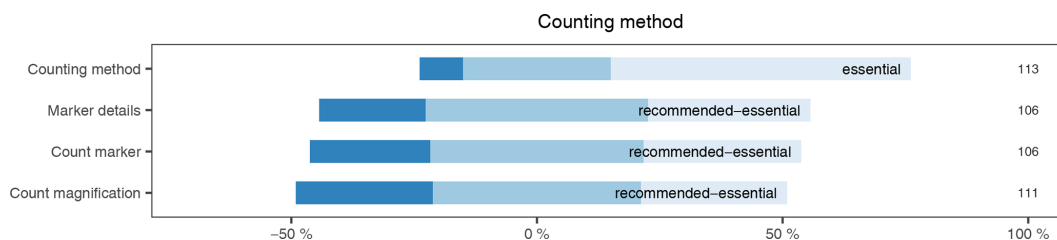
Counting magnification; information about whether or not and what kind of count marker was used; and, where applicable, details about the marker (batch number, amount used) all increase the value of a dataset and are considered recommended–essential.

### 4.6 Taxonomic details

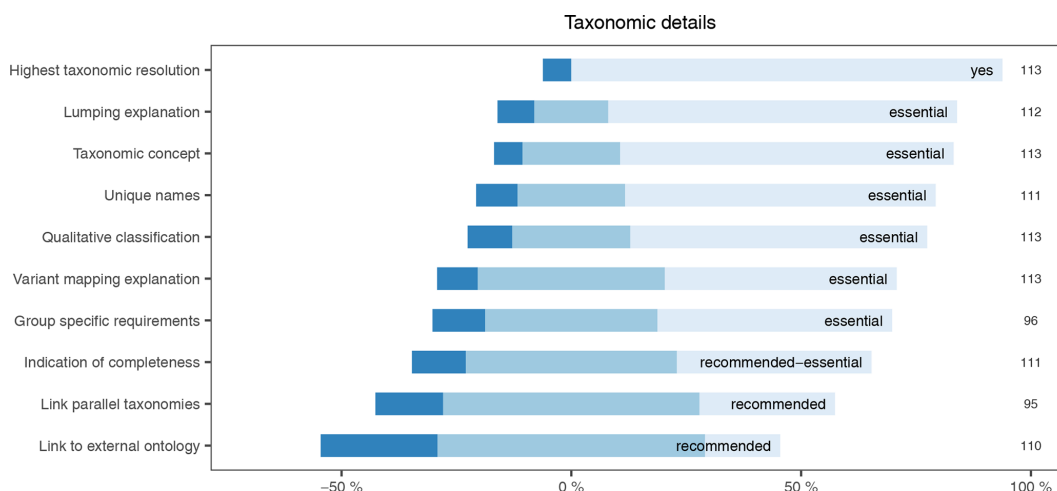
For the purpose of these guidelines, information about the taxonomy is among the most important attributes to tackle the complexity of microfossil assemblage data and increase their potential for reuse. It is important to clarify the taxonomic guides used for species identification. The survey results with regards to the taxonomic details are summarised in Fig. 8.

#### 4.6.1 Essential

An explanation of the taxonomic concept used for the study is essential for the reuse of microfossil data, as this aids the harmonisation of datasets generated by different researchers



**Figure 7.** Community ranking of details on the counting method of microfossil assemblages. Colours and layout as in Fig. 3.



**Figure 8.** Community ranking of taxonomic details of microfossil assemblage data. Colours and layout as in Fig. 3.

and the updating of data when taxonomic concepts change. Such an explanation should, ideally, cite the work(s) on which the taxonomy is based and, if relevant, explain departures from the cited concept. Taxonomic concepts should also be illustrated using images where possible. Several survey respondents, those working with taxonomically complex groups or with extinct species in particular, indicated the need for the inclusion of imagery in datasets. Linking the abundance data to image files illustrating the individual specimens analysed is not yet a common practice because of the challenges related to required storage space, long-term preservation of these images, and related costs. However, availability of such image datasets would allow true reproducibility and even updating of the classification of individual specimens as taxonomic insights proceed.

An overwhelming majority of the survey participants (94 %) agreed that species count data should be archived the way they were counted, i.e. at the highest taxonomic resolution and excluding summed taxa when the constituent taxa have been counted separately. For instance, reporting the abundance of the planktonic foraminifera species *Globigerinoides ruber* is not necessary and hinders reusability if the abundance of the subspecies *G. ruber ruber* and *G. ruber albus* is already provided. Data archived in this way are more easily reused, since lumping of taxa is trivial, whereas removing lumped abundance data increases the workload and

the possibility of errors down the line. To ensure the reproducibility of a particular analysis that requires lumping of taxa that were counted separately, the lumping can be explained in an accompanying paper or code.

If taxa were counted together because they were or could not be distinguished, an explanation of the lumping needs to be included, or the lumping needs to be made obvious in the dataset. For instance, if the planktonic foraminiferal species *G. ruber* and *Globigerinoides elongatus* were counted together, their abundance should be reported as “*Globigerinoides ruber* and *Globigerinoides elongatus*”, or it should be made clear that the counts of *G. ruber* include individuals of *G. elongatus*. The former option is clearest and most likely to eliminate errors upon reuse of these data. The explanation of the lumping is especially important when updating the taxonomy of legacy datasets, as the example above illustrates. The species *G. elongatus* has only recently been split from *G. ruber* (Aurahs et al., 2011), and virtually all planktonic foraminifera abundance legacy datasets only include *G. ruber*, which in most cases represents the abundance of the two species combined. Taxonomic concepts can also change over time, as the holotype and paratype materials in museum collections are digitally imaged and become available. For example, *Globigerina praebulloides* was frequently reported in early Miocene assemblages; however, investigations of the holotype showed this species to be a junior synonym of *Glo-*

*bigerinella obesa* (Spezzaferri et al., 2018), requiring a new name for the form previously reported as *Globigerina prae-bulloides*.

The mapping of variants onto the parent taxon needs to be clear from the dataset or explained in the description, especially when using informal taxonomy. Moreover, fully spelled out (genus and species) and unique taxon names should be used to avoid confusion. This requirement may seem odd, but an astonishing number of datasets of planktonic foraminifera stored at PANGAEA contain the same taxon names more than once (Strack et al., 2023), complicating their reuse considerably. When abundances are indicated only qualitatively, the qualitative classification scheme should be clearly described, e.g. clarifying whether “a” stands for “absent” or “abundant”. Finally, some groups have taxonomic group-specific requirements; for instance, when reporting dinoflagellate cyst assemblage composition, it should be made clear whether the taxonomy used refers to the cyst or motile stage of the organisms.

#### 4.6.2 Recommended–essential

Survey participants also ranked an indication of the taxonomic completeness of the counts highly. Even though the completeness of count data can be estimated if the abundance of unidentified specimens is reported, information about whether all species present in the sample (set) have been reported markedly increases the value of a dataset, as it is essential information to assess biodiversity.

#### 4.6.3 Recommended

For microfossil groups with resting stage assemblages, participants recommend including an explanation or reference to how to link parallel taxonomies (i.e. the link between resting and motile stage taxonomies), which is important information required when datasets with different taxonomies are merged. Similarly, including links to an external ontology for the taxon names increases the value of microfossil assemblage data, as such links can facilitate automated harmonisation of datasets with different taxonomies and automated updating of legacy data. It should be noted that some repositories already include such links. For instance, PANGAEA aims to include AphiaIDs that link taxa to entries in the World Register of Marine Species (<https://www.marinespecies.org/>, last access: 14 May 2025).

### 4.7 Sample data

The following section covers aspects and properties of microfossil data that vary by sample and includes recommendations on how to report abundance data. The survey results are split into information relating to the (physical) samples themselves (Fig. 9) and to the abundance data (Fig. 10).

#### 4.7.1 Essential

For each sample in a dataset, its bottom and top depth, including a unit, should be reported. Alternatively, the mid-depth and thickness of the sample should be provided. When available, the age of the sample and its unit are also essential. We note that age–depth relationships are uncertain and subject to change. Linking to information needed to update chronologies (e.g. radiocarbon ages, tie points to reference curves, timescale, or zonation) is, therefore, recommended.

When available, a unique sample identifier, e.g. an IODP sample ID that allows updating composite depth scales or an International Generic Sample Number (IGSN; formerly International Geo Sample Number), is also essential.

#### 4.7.2 Recommended–essential

Reporting the dry sample mass, including a unit, increases the value of these data, as this allows the calculation of microfossil concentrations or accumulation rates. In addition, if sedimentary (or lithostratigraphic) units have been distinguished within the core or section, information about the specific unit to which the sample belongs adds significant value.

#### 4.7.3 Desired–recommended

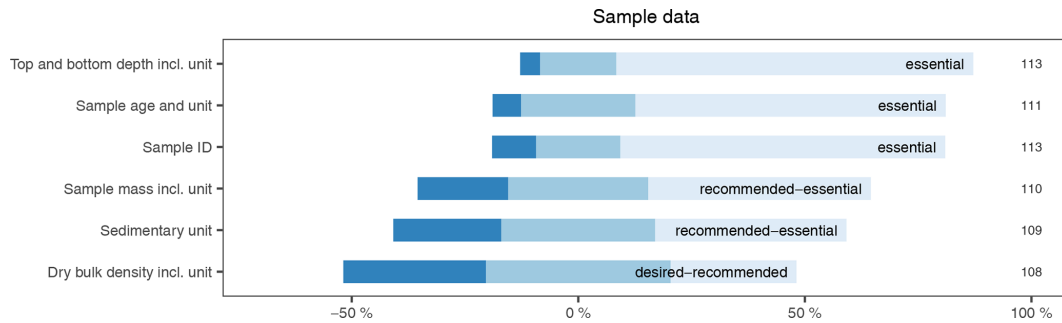
Survey respondents ranked information about the dry bulk density, including a unit, of the sample in between desired and recommended.

### 4.8 Abundance data

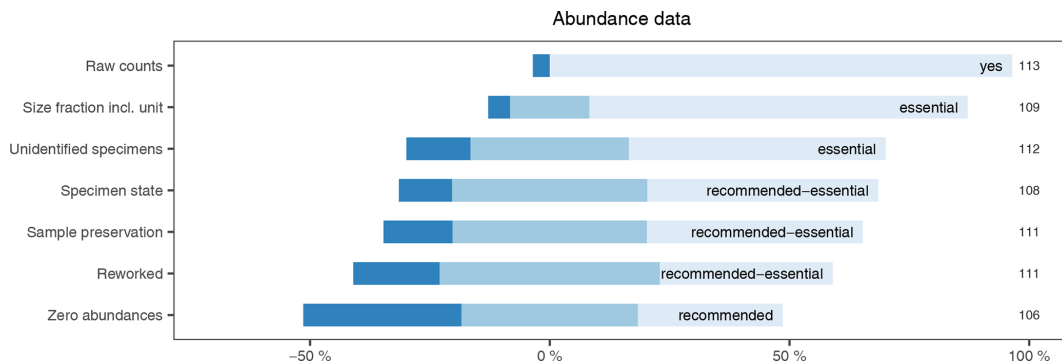
#### 4.8.1 Essential

For the microfossil abundance data, 96 % of the respondents agreed that raw specimen counts, rather than relative abundances, should be reported. The advantage of raw count data is the possibility of quality assurance and estimation of confidence limits based on the total number of specimens counted. The availability of raw counts also allows the averaging or combining of samples and, importantly, reduces the risk of rounding and other errors. There is no need to report relative abundance together with the raw counts, since this calculation is trivial and reporting the same data twice increases required storage space and requires additional processing steps upon reuse. In many cases, reporting counts is sufficient to obtain the desired information (e.g. when only relative abundances are needed for the analysis). Raw counts are usually reported as integers, but it should be explained if – and how – fragments of microfossils have been counted. When concentrations or accumulation rates can be calculated based on the available data, information about the split fraction is also essential and preferable over specimen concentrations or accumulation rates.

Survey participants also ranked reporting of the size fraction (minimum, maximum, and a unit) in which the micro-



**Figure 9.** Community ranking of sample information of microfossil assemblage data. Colours and layout as in Fig. 3.



**Figure 10.** Community ranking of abundance data aspects of microfossil assemblage data. Colours and layout as in Fig. 3.

fossils were counted and the number of unidentified specimens as essential. The former is needed, since the species composition of assemblages may vary as a function of the size (e.g. Patterson and Fishbein, 1989; Peeters et al., 1999), complicating the comparison of assemblages from different size ranges. Reporting unidentified specimens allows assessment of the taxonomic completeness and quality of a dataset.

#### 4.8.2 Recommended–essential

Information about the state of the specimens, e.g. whether they were stained or not, and whether the sample contains reworked specimens is ranked between recommended and essential. If samples were stained, the methodology should be clearly described and the strategy for distinguishing stained from unstained should be explained. Similarly, an explanation of how the state or reworking was assessed needs to be supplied to interpret these data. Information about the sample preservation (e.g. degree of dissolution or mechanical break-up, affected by secondary precipitation) ranked at the same level. Ideally, the description of the state of preservation should be created on the basis of measurable and clearly described criteria (e.g. Broerse et al., 2000; Dittert et al., 1999).

#### 4.8.3 Recommended

Reporting the absence of species using zero abundances, rather than not reporting the species, is recommended. The absence of a species from a dataset could either mean that the species was not observed or that it was not counted. Since the real absence of a species can be relevant information, the abundance of species not reported in a dataset is, out of necessity, often interpreted as meaning zero abundance. However, this ambiguity can be avoided easily, particularly when reporting assemblage data of groups with relatively few species.

### 5 Difference in data stewardship attitudes among career stages and microfossil groups

The microfossil data reporting guidelines that followed from the survey represent the consensus view of the diverse marine micropalaeontological community. Whilst the main goal of the survey was to establish these guidelines, the design of the survey allowed us to test for differences in the attitude towards data stewardship among researchers with different levels of experience (split at 5 years since obtaining a PhD degree) and among researchers working on different microfossil groups (planktonic and benthic foraminifera and others).

The analysis indicates that, irrespective of the focal fossil group, experienced researchers are 1.3 times more likely to choose “desired” over “recommended” than early career researchers ( $p = < 0.01$ ). There is, however, no significant difference in the odds of choosing “essential” over “recommended” between these career groups ( $p = 0.29$ ), indicating that, even though the two groups generally agree on what aspects or properties of microfossil assemblage datasets are essential for reuse, early career researchers favour the inclusion of more information. This difference could have several reasons. It is possible that experienced researchers have over the years learned to work with datasets with fewer attributes and are pessimistic about obtaining the long list of recommended and essential data aspects for legacy data. On the other hand, early career researchers may be more aware of the FAIR data principles and more likely to take a big data approach to micropalaeontology and, for these reasons, favour datasets with more information. Notwithstanding, these results offer hope for future datasets to contain richer metadata and be more easily reusable for a wide range of applications.

Irrespective of the career stage and compared to researchers working with benthic foraminifera, those working on planktonic foraminifera are 1.3 times more likely to choose “desired” over “recommended” ( $p = 0.01$ ) and 0.8 times less likely to choose “essential” over “recommended” ( $p = < 0.01$ ). Researchers working on other microfossil groups showed no significant difference in voting behaviour compared to those working on benthic foraminifera ( $p = 0.16$  and  $0.41$ , respectively). The tendency of planktonic foraminifera researchers to have fewer demands on the metadata might reflect the low number of species and hence lower taxonomic complexity of planktonic foraminifera compared to other microfossil groups, along with the types of applications for which their sedimentary assemblages are most often used.

## 6 Do available datasets adhere to the guidelines?

Of the 49 data attributes covered in the survey, only 10 are ranked recommended or below, rendering the guidelines as ambitious. To provide a first-order assessment of how well existing microfossil assemblage datasets adhere to the new guidelines, we analysed 35 randomly selected datasets (see Methods). On average, data attributes ranked as essential for reuse were included in just over half of the datasets assessed (Fig. 11). Importantly, five of these essential data aspects (site name, site location, collection method, source, contributor) are required or standard elements of datasets at PANGAEA, where the sample set was taken from. The assessment clearly emphasises the value of data repositories that curate and quality-check these data. The mismatch between what these guidelines indicate as essential data attributes and what is (not) included in the assessed datasets also highlights a large gap between what the community thinks should be

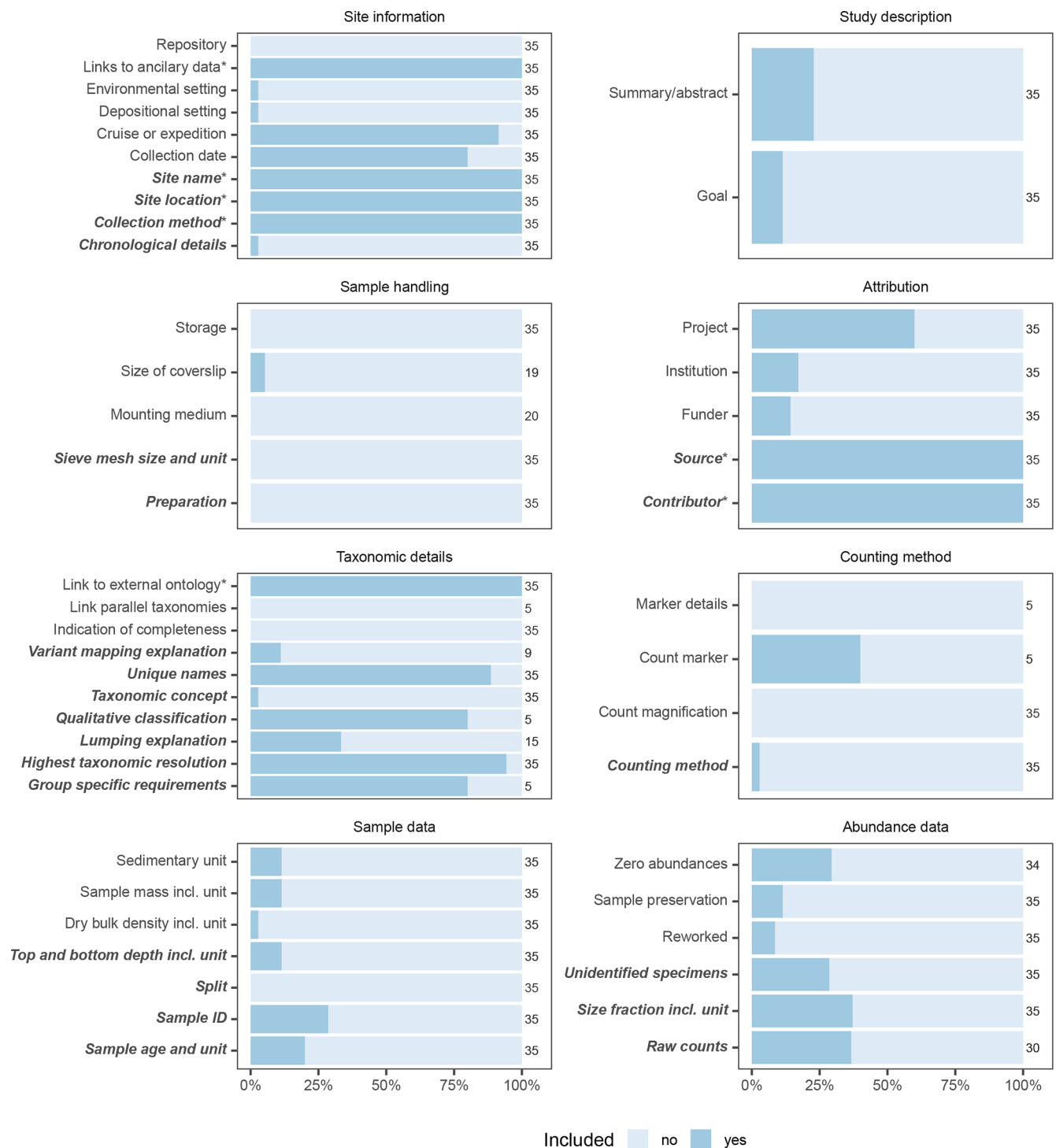
included in a dataset and what is actually included in them. The most likely reason for this mismatch is that datasets are generally published in parallel with a publication, in which (more) details about the abundance data are provided than in the published standalone dataset. As a result, the information needed for reuse is to some degree available (though not necessarily freely accessible) but not as part of the dataset itself. This distribution of information complicates the reuse of microfossil abundance data. Whatever the exact reasons for the mismatch, adherence to its own guidelines to improve the reusability of data will require a cultural change in how we as a community archive and share data.

## 7 A more reusable example

To provide an example of a marine microfossil abundance dataset that meets the requirements for reuse and includes all essential data aspects, the planktonic foraminifera assemblage data from IODP Site 306-U1314 were upgraded by the data generator. The original dataset (Alonso-Garcia et al., 2011b) was deposited at PANGAEA to allow reproducibility of a study investigating arctic front shifts during the mid-Pleistocene (Alonso-Garcia et al., 2011a). Of the essential data aspects, the originally deposited dataset lacked details on the methodology, chronology, and taxonomy. In addition, the abundance data did not cover all species, were reported as percentages, and contained abundances of lumped taxa/species groups that were not described. Sample depths and sample allocation to hole, core, and section were also unclear or not given.

The upgraded version (Alonso-Garcia et al., 2024) contains the raw count data of all taxa including unidentified specimens needed to assess the taxonomic coverage and completeness of the counts. It provides information about the origin of the samples and an accurate depth and ID assignment. Sample mass and density are included to allow calculation of the planktonic foraminifera concentration and accumulation rate. In addition, the description of the dataset now contains information about the goal of the data collection and a detailed methodology, including details on how to reproduce the analysis presented in Alonso-Garcia et al. (2011a). Minor changes were made to align the taxonomy with current insights and clarify the lumping of taxa previously not recognised as separate species. These changes are indicated in the new dataset as text. Information about the goal of the analysis and the methodological and taxonomic aspects is, for now, supplied as text in the description of the dataset. The dataset now contains all the information deemed essential to reuse the microfossil assemblage data, and the example shows that microfossil assemblage data with sufficient information for proper reuse can be archived with relatively little additional effort.





**Figure 11.** First-order assessment of adherence to guidelines of selected microfossil datasets from PANGAEA. Data attributes ranked as essential for reuse are highlighted in italics and bold; items with an asterisk are default variables at PANGAEA. Not all properties are applicable to each dataset, so the number of datasets considered for the calculation is given on the right in grey.

## 8 The way forward

### 8.1 Meeting the guidelines

The survey results clearly indicate that the micropalaeontological community values good data stewardship and sets the bar high for datasets to meet the requirements deemed essential for reuse. Crucially, survey participants highlighted the importance of methodological aspects related to the data generation process that is essential to enable reuse of microfossil abundance data. However, despite the importance of extensive metadata, the preliminary assessment of the completeness of legacy data indicates a clear mismatch between expectations of what information a reusable dataset should contain and the reality of the (meta)data provided. This mismatch may be because we assessed only legacy data, and new(er) datasets may adhere more closely to the community guidelines as the importance of good data stewardship is now more widely recognised by the community and encouraged, or even required, by publishers and funders. However, it is also possible that, historically, data archiving and sharing were not regarded as important as the data generation process and that data were archived at the last minute at the end of the publication of a study without (much) consideration of the FAIR data principles. A similar mismatch between community expectations about data standards and the information actually contained in datasets seems to hold for palaeoclimate data at large (Khider et al., 2019). PaCTS was framed as an aspirational data standard because of the sheer number of data properties the palaeoclimate community considered essential for data reuse. Despite the apparent gap between the guidelines described here and the sampled microfossil assemblage datasets, we believe that, although the guidelines are ambitious, they are not onerous and should not remain aspirational. We recommend that new datasets submitted to repositories, at a minimum, include the items that the guidelines specify as essential information for marine microfossil assemblages, but the inclusion of desired and recommended attributes is encouraged. The example given in Sect. 7 shows that this is possible, but meeting the guidelines will require dedicated effort and, to some degree, a change in how we approach data stewardship.

Many institutes have, or are setting up, internal rules for research data management that generally adhere to the FAIR principles. Funding agencies also increasingly demand clear and FAIR data management plans. The guidelines presented here could help individual researchers to develop data management plans and increase the reusability of their datasets, as they provide clear guidance on what aspects render a microfossil abundance dataset reusable. The guidelines also enable repositories that curate datasets (e.g. PANGAEA, Neotoma) and database managers to increase the reusability of microfossil abundance data by ensuring essential data properties are included in datasets, for instance, through issuing data input templates that incorporate these guidelines.

For datasets associated with scientific articles, publishers, editors, and reviewers could explicitly consider the datasets and their reusability in evaluating contributions to the journals. Many publishers already require that research data be made (publicly) available at the review stage. However, availability does not equate to reusability. Being endorsed by the community, the guidelines offer an easy instrument to evaluate and improve the reusability of microfossil abundance datasets during review.

### 8.2 Improving the reusability of legacy data

Apart from providing guidelines for new microfossil datasets, the guidelines can also be used to retrospectively increase the reusability of legacy data. Unpublished microfossil abundance data (either so-called “dark data” that sit on a local hard drive or server, data of which the existence can be inferred from the availability of quantitative reconstruction, or data that are only shown in a publication in graphical form) can be made available at community-recognised repositories according to the FAIR data principles. Such legacy datasets can directly be made compliant with the guidelines presented here. However, legacy data that are in the grey zone, e.g. only available as a supplement to a paper and hence not necessarily accessible, easily interoperable, or freely reusable, can, in addition to being made FAIR, also be upgraded to meet the community guidelines presented here. The same holds for legacy data that are truly in the public domain. Individual researchers can upgrade legacy datasets for small numbers of datasets (e.g. their own), but larger (synthesis) projects may require community efforts (Smith et al., 2023a), as retrieving (meta)data from other sources is cumbersome. Some information, e.g. count data, may be lost entirely, and upgrading legacy datasets to meet the community guidelines completely may not be possible. For reasons of transparency, upgrades of legacy data that were already accessible should clearly refer to the original data and preserve the original taxonomy to allow future updating or revision. The use of (external) ontologies may markedly reduce the workload associated with such revisions.

### 8.3 Future perspectives

The survey results and the resulting data guidelines for marine microfossil assemblages are not set in stone but reflect the state of an ongoing discussion about data stewardship. As research demands develop and regulations change, so too will the frameworks surrounding microfossil metadata. In this context, the guidelines outlined in this paper should be seen as an attempt to initiate a formal community-led process of improving data standards for marine microfossil assemblages.

To move forward with increasing the reusability and interoperability of microfossil data, the community needs to define a standardised vocabulary or ontology. To do so re-

quires us to reach an agreement on names used for the various data aspects and properties and also on the values or form they can have or be described with. Such standardised vocabulary should at least cover the aspects and properties of microfossil abundance datasets essential for reuse. Still, a more ambitious standardisation should cover all relevant aspects of the data. Ideally, the new ontology should complement or extend existing standardisation efforts (e.g. <https://www.tdwg.org/standards/>, last access: 14 May 2025) in order to integrate microfossil abundance data into a larger framework of earth science and biodiversity information and also make such data reusable outside the micropalaeontological community. Many of the data attributes identified as essential relate to the methodology used to generate these data. The lack of such provenance information also affects other disciplines, and a cross-disciplinary approach may therefore be necessary to solve these problems. Existing approaches, such as PROV-O (<https://www.w3.org/TR/prov-o/>, last access: 14 May 2025), which is well suited to extending existing metadata formats as recommended by the Data on the Web Best Practices Working Group (<https://www.w3.org/TR/dwbp/#provenance>, last access: 14 May 2025), could be used for this purpose.

Perhaps, ultimately, the marine micropalaeontological community can, together with dedicated repositories, design or agree on a common format for the data that allows searching and filtering by all relevant (meta)data. A standardised format should be flexible to allow changes in data requirements and accommodate the complexity of microfossil assemblage data. Ideally, microfossil data should be incorporated into a framework that links the microfossil abundances to other data from the same archive. Such ancillary data may include chronological information required to assess or update time series, other (palaeo)data to put the microfossil data into an environmental perspective, or image material. Fortunately, several data formats that can accommodate palaeo-data are available and, in theory, flexible enough to accommodate the metadata and ancillary data demands for microfossil assemblage data, e.g. LiPD (McKay and Emile-Geay, 2016), Darwin Core (Wieczorek et al., 2012), SOD (Lazarus et al., 2018), and ABCDEFG (Petersen et al., 2018). Therefore, this process does not need to start from scratch.

## Appendix A

**Table A1.** Microfossil abundance data attributes that were evaluated in the community survey and their resulting ranking. The table presents (partially fictional) examples from across microfossil groups of how each attribute can be reported; it does not prescribe permitted values or vocabularies (see Sect. 8.3). Please refer to Sect. 4 for an extended description of the attributes. References to Table A2 are provided in italics.

Attribute group	Attribute	Description	Example	Ranking
Site information	Location	Latitude, longitude, and elevation (including a unit) of the study site	Latitude: 56.36; longitude: $-27.89$ ; elevation: $-2800.3$ m	Essential
Site information	Collection method	Method used to collect the samples	Advanced piston coring	Essential
Site information	Site name	Name of the study site	IODP Site 306-U1314	Essential
Site information	Details about the chronology	Details of how the chronology of the site was established	The chronology of the time series presented here is based on tuning the $\delta^{18}\text{O}$ of benthic foraminifera ( <i>Cibicidoides wuellerstorfi</i> , <i>Cibicidoides</i> sp., <i>Melonis pompilioides</i> ) of Site U1314 with the LR04 global stack.	Essential
Site information	Cruise or campaign	Name of the cruise or campaign during which the study material was collected	IODP Expedition 306 (North Atlantic Climate 2)	Recommended–essential
Site information	Collection date	Date at which the study site was sampled	23 October 2004	Recommended–essential
Site information	Repository	Repository where the archive and/or samples are stored	Core: Bremen Core Repository; samples: University of Salamanca	Recommended–essential
Site information	Environmental setting	Environmental setting of the study site	Pelagic	Recommended–essential
Site information	Depositional setting	Depositional setting of the study site	Sediment drift	Recommended–essential
Site information	Links to ancillary data	Links to other data measured in the same archive	<a href="https://doi.org/10.1234566789">https://doi.org/10.1234566789</a>	Recommended
Study description	Goal	A description of the goal for which the microfossil assemblages were analysed	Planktonic species assemblage data were generated to assess changes in the position of the Arctic front on millennial timescales across the mid-Pleistocene transition.	Recommended–essential
Study description	Summary/abstract	A summary or abstract of the findings of the study	Surface water conditions were inferred using planktonic foraminifer assemblages between Marine Isotope Stages 19 and 11 (ca. 800–400 ka). Factor analysis of the planktic foraminifer assemblages suggests... ( <i>incomplete example</i> )	Recommended–essential

Table A1. Continued.

Attribute group	Attribute	Description	Example	Ranking
Attribution	Source	The source of the data so the dataset can be identified	<a href="https://doi.org/10.1234566789">https://doi.org/10.1234566789</a>	Essential
Attribution	Contributor	The person(s) who contributed the data to the repository	Alonso-Garcia, Montserrat	Essential
Attribution	Institution	Institution where these data were generated	University of Salamanca, Spain	Recommended
Attribution	Project	Project during which these data were generated	Multidisciplinary research consortium on gradual and abrupt climate changes and their impacts on the environment (GRACCIE)	Desired–recommended
Attribution	Funder	Funding agency of the project or research during which these data were generated	Ministerio de Educación, Formación Profesional y Deportes, grant no. CSD2007-00067	Desired
Sample handling	Initial sieve mesh size	The mesh size of the sieve used for initial washing of the sediment. Note that this size may differ from the actual size fraction of the counted microfossils.	63 µm	Essential
Sample handling	Preparation	Preparation of the samples prior to microfossil counting	Prior to counting, the sediment samples were washed by wet-sieving over a 63 µm mesh with deionised water, dried overnight in an oven at 40 °C, and sieved again over a 150 µm mesh retaining both the coarse (> 150 µm) and fine fractions (63–150 µm).	Essential
Sample handling	Storage	A description of how the core and/or the samples were stored before analysis	The sediment core was stored at 4 °C prior to sampling; once sampled, samples were kept refrigerated at 4 °C until processing.	Recommended–essential
Sample handling	Mounting medium	The medium used to mount the prepared sample on a coverslip	Glycerine jelly	Recommended
Sample handling	Size of coverslip	The size (area) of the coverslip including a unit	1 cm <sup>2</sup>	Desired–recommended
Counting method	Counting method	A description of how the microfossil counts were conducted	Counts were made manually using a stereomicroscope on aliquots of the samples containing at least 234 specimens.	Essential
Counting method	Counting magnification	The microscope magnification used for the counting	×250	Recommended–essential
Counting method	Count marker	Which marker fossils were used (if any)	<i>Lycopodium</i>	Recommended–essential
Counting method	Count marker details	Details about the marker (e.g. batch number)	Batch number 123456; mean number of spores 96 660 ± 2123 (1 SD)	Recommended–essential



Table A1. Continued.

Attribute group	Attribute	Description	Example	Ranking
Taxonomic details	Taxonomic concept	A description of the taxonomic concept used to identify the microfossils	John and Smith (2025) except for <i>Tenuitella compressa</i> , which is here considered a synonym of <i>T. fleisheri</i> .	Essential
Taxonomic details	Highest taxonomic resolution	Report the counts of taxa at the taxonomic resolution they were counted.	See example in Table A2 (subspecies reported individually, sum of subspecies excluded).	Essential
Taxonomic details	Explanation of lumping	Clarify how taxa are lumped if they were not counted separately.	<i>Globigerinoides ruber albus</i> includes <i>Globigerinoides elongatus</i> .	Essential
Taxonomic details	Mapping of variants	If variants were counted, clarify how they map on the parent taxon.	Left- and right-coiling specimens of <i>Turborotalita quinqueloba</i> have been counted separately.	Essential
Taxonomic details	Unique taxon names	Use full and unique genus and species (and variant/subspecies) names.	<i>Globigerinoides ruber albus</i> (as opposed to “ruber”)	Essential
Taxonomic details	Qualitative classification scheme	When relevant, provide an unambiguous qualitative abundance classification scheme.	Abundant, frequent, rare, absent (not A, F, R, A)	Essential
Taxonomic details	Group-specific requirements	For dinoflagellate cysts: specify if a cyst or motile taxonomy was used.	The taxonomy of the dinoflagellate cysts refers to their motile stage.	Essential
Taxonomic details	Completeness	An indication of whether all specimens in the samples have been counted	All specimens present in the sample have been counted. Counts of unidentified specimens are included.	Recommended–essential
Taxonomic details	Link parallel taxonomies	For dinoflagellate cysts: describe how the cyst and motile taxonomy map onto each other.	Motile and cyst taxonomies can be linked following <a href="https://doi.org/10.1234566789">https://doi.org/10.1234566789</a>	Recommended
Taxonomic details	Links to an external ontology	Include links to an external ontology for taxon names.	113458. See column “ <i>AphiaID</i> ” in example dataset in table 2 for further examples.	Recommended
Sample data	Bottom and top depth	Bottom and top depth of the sample including a unit	4–6 cm. See columns “ <i>Top depth (cm)</i> ” and “ <i>Bottom depth (cm)</i> ” in Table A2.	Essential
Sample data	Age	Age of the sample including a unit	0.2 ka. See column “ <i>Age (ka)</i> ” in Table A2.	Essential
Sample data	Unique sample identifier	A unique identifier of the sample, e.g. an IODP or International Generic Sample Number (when available)	306-U1314C-4H-5,4-6. See column “ <i>Sample UID</i> ” in Table A2.	Essential
Sample data	Sample mass	Dry mass of the sample including a unit (when available)	1.5 g. See column “ <i>Sample mass (g)</i> ” in Table A2.	Recommended–essential

**Table A1.** Continued.

Attribute group	Attribute	Description	Example	Ranking
Sample data	Sedimentary unit	If the samples are collected from sedimentary or lithostratigraphic units, provide the unit the sample belongs to.	I, II. <i>See column “Sedimentary unit” in Table A2.</i>	Recommended–essential
Sample data	Dry bulk density	Dry bulk density of the sample including a unit (when available)	$1.8 \text{ g cm}^{-3}$ . <i>See column “Dry bulk density (<math>\text{g cm}^{-3}</math>)” in Table A2.</i>	Desired–recommended
Abundance data	Counts	Provide microfossil abundance data as counts (integers).	<i>See column “Count” in Table A2.</i>	Essential
Abundance data	Split	Report the fraction of the sample used for counting (if split).	$1/64$ . <i>See column “Split” in Table A2.</i>	Essential
Abundance data	Size fraction	Upper and lower limits of the size fraction including a unit in which the microfossils were counted	$150\text{--}250 \mu\text{m}$ . <i>See column “Size fraction (<math>\mu\text{m}</math>)” in Table A2.</i>	Essential
Abundance data	Unidentified specimens	Include counts of unidentified specimens.	<i>See example in Table A2 (unidentified is included in the species column).</i>	Essential
Abundance data	State of the specimens	Indicate if specimens were stained or contained cytoplasm.	NA, true, false. <i>See column “Stained” in Table A2.</i>	Recommended–essential
Abundance data	Reworked	Indicate if specimens were reworked or not.	True, false. <i>See column “Reworked” in Table A2.</i>	Recommended–essential
Abundance data	Sample preservation	A (quantitative) estimate of the preservation of the microfossils in the sample	0.5. <i>See column “Preservation” in Table A2.</i>	Recommended–essential
Abundance data	Zero abundances	Report taxa not observed in the sample.	<i>See column “Count” in Table A2 where species abundance is indicated as 0.</i>	Recommended

**Table A2.** Hypothetical example of reporting microfossil abundance data. Note that this table only provides an example of how sample and microfossil data may look and that it excludes important data attributes related to the site, methodology, and attribution. The columns “Reworked” and “Preservation” refer to the individual samples. A real-world and complete example of a data file containing all essential data attributes is provided in Alonso-Garcia et al. (2024).

Top depth (cm)	Bottom depth (cm)	Age (ka)	Sample UID	Sample mass (g)	Dry bulk density (g cm <sup>−3</sup> )	Sedi-mentary unit	Split	Species	AphiaID	Size fraction (μm)	Count	Stained	Reworked	Preser-vation
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	<i>Neogloboquadrina pachyderma</i>	113458	150–250	0	NA	False	0.5
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	<i>Neogloboquadrina incompta</i>	113474	150–250	10	True	False	0.5
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	<i>Neogloboquadrina incompta</i>	113474	150–250	3	False	False	0.5
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	<i>Globigerinoides ruber albus</i>	1629739	150–250	11	True	False	0.5
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	<i>Globigerinoides ruber ruber</i>	1629738	150–250	55	True	False	0.5
4	6	0.2	306-U1314C-4H-5,4-6	1.5	1.8	I	1/64	Unidentified	NA	150–250	14	NA	False	0.5
6	8	0.4	306-U1314C-4H-5,6-8	1.6	1.9	II	1/32	<i>Neogloboquadrina pachyderma</i>	113458	150–250	3	False	True	0.45
6	8	0.4	306-U1314C-4H-5,6-8	1.6	1.9	II	1/32	<i>Neogloboquadrina incompta</i>	113474	150–250	15	False	True	0.45
6	8	0.4	306-U1314C-4H-5,6-8	1.6	1.9	II	1/32	<i>Globigerinoides ruber albus</i>	1629739	150–250	2	False	True	0.45
6	8	0.4	306-U1314C-4H-5,6-8	1.6	1.9	II	1/32	<i>Globigerinoides ruber ruber</i>	1629738	150–250	5	False	True	0.45
6	8	0.4	306-U1314C-4H-5,6-8	1.6	1.9	II	1/32	Unidentified	NA	150–250	2	NA	True	0.45

**Code availability.** The code used for the analysis and to compile the figures is available at <https://doi.org/10.5281/zenodo.15411453> (Jonkers, 2025).

**Data availability.** Survey questions and anonymised answers are available at <http://doi.org/10.5281/zenodo.12722701> (Jonkers and Strack, 2024).

**Author contributions.** LJ and TS were responsible for the conceptualisation and design of the survey, conducted the survey, and led the analysis of the survey results. LJ was responsible for writing and implementing the code used for survey analysis. LJ and TS prepared the figures and wrote the original draft of the paper. MAG contributed the example dataset, upgraded the original dataset to meet the standards mentioned in this publication, and made major contributions to the text. SD was significantly involved in the design and analysis of the survey and made major contributions to the text. RH and MK acquired the funding and designed the overall project that led to this publication. IHA, CLCJ, BM, RS, LS, and SKV helped in the design of the survey or in the analysis of the results and contributed significantly to the proposed outline of the paper and to the review and editing of the text and figures. All other authors are listed alphabetically and contributed to the paper by editing, commenting on, or reading and approving the text.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

**Acknowledgements.** We are indebted to the researchers who helped design the survey questions and to the people who helped with the dissemination of the survey. Furthermore, we are grateful for the participation of all scientists who took the time to fill out the survey.

**Financial support.** This work has been funded by the German Research Foundation (NFDI4Earth 1st Cohort of Pilots, DFG project no. 460036893, <https://www.nfdi4earth.de/>, last access: 14 May 2025) within the German National Research Data Infrastructure (NFDI; <https://www.nfdi.de/>, last access: 14 May 2025).

The article processing charges for this open-access publication were covered by the University of Bremen.

**Review statement.** This paper was edited by Sev Kender and reviewed by two anonymous referees.

## References

- Alonso-Garcia, M., Sierro, F. J., and Flores, J. A.: Arctic front shifts in the subpolar North Atlantic during the Mid-Pleistocene (800–400 ka) and their implications for ocean circulation, *Palaeogeogr. Palaeoclimatol.*, 311, 268–280, 2011a.
- Alonso-Garcia, M., Sierro, F. J., and Flores, J.-A.: Mid-Pleistocene (800–400 ka) planktic foraminifer assemblages of IODP Site 306-U1314, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.820283>, 2011b.
- Alonso-Garcia, M., Sierro, F. J., and Flores, J.-A.: Planktonic foraminifera assemblage data spanning the mid Pleistocene (400–800 ka) of IODP Site 306-U1314 from the North Atlantic Ocean, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.968803>, 2024.
- Aurahs, R., Treis, Y., Darling, K., and Kucera, M.: A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence, *Mar. Micropaleontol.*, 79, 1–14, 2011.
- Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., and Honeyman, T.: Introducing the FAIR Principles for research software, *Sci. Data*, 9, 1–6, 2022.
- Broerse, A. T. C., Ziveri, P., and Honjo, S.: Coccolithophore (–CaCO<sub>3</sub>) flux in the Sea of Okhotsk: seasonality, settling and alteration processes, *Mar. Micropaleontol.*, 39, 179–200, 2000.
- Brummer, G.-J. A. and Kučera, M.: Taxonomic review of living planktonic foraminifera, *J. Micropalaeontol.*, 41, 29–74, 2022.
- Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., and Rose, A. K.: A study of the impact of data sharing on article citations using journal policies as a natural experiment, *PLoS One*, 14, e0225883, <https://doi.org/10.1371/journal.pone.0225883>, 2019.
- CLIMAP project members: The Surface of the Ice-Age Earth, *Science*, 191, 1131–1137, 1976.
- Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K., and McGillivray, B.: The citation advantage of linking publications to research data, *PLoS One*, 15, e0230416, <https://doi.org/10.1371/journal.pone.0230416>, 2020.
- Crichton, K. A., Wilson, J. D., Ridgwell, A., Boscolo-Galazzo, F., John, E. H., Wade, B. S., and Pearson, P. N.: What the geological past can tell us about the future of the ocean's twilight zone, *Nat. Commun.*, 14, 2376, <https://doi.org/10.1038/s41467-023-37781-6>, 2023.
- Dietl, G. P., Kidwell, S. M., Brenner, M., Burney, D. A., Flessa, K. W., Jackson, S. T., and Koch, P. L.: Conservation Paleobiology: Leveraging Knowledge of the Past to Inform Conservation and Restoration, *Annu. Rev. Earth Pl. Sc.*, 43, 79–103, 2015.
- Dittert, N., Baumann, K.-H., Bickert, T., Henrich, R., Huber, R., Kinkel, H., and Meggers, H.: Carbonate Dissolution in the Deep-Sea: Methods, Quantification and Paleoceanographic Application, in: *Use of Proxies in Paleoceanography: Examples from the South Atlantic*, edited by: Fischer, G. and Wefer, G., Springer Berlin Heidelberg, Berlin, Heidelberg, 255–284, [https://doi.org/10.1007/978-3-642-58646-0\\_10](https://doi.org/10.1007/978-3-642-58646-0_10), 1999.
- Dunkley Jones, T. and Bown, P. R.: Post-sampling dissolution and the consistency of nannofossil diversity measures: A case study from freshly cored sediments of coastal Tanzania, *Mar. Micropaleontol.*, 62, 254–268, 2007.
- European Commission, Directorate-General for Research and Innovation: Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data, Publications Office of the European Union, <https://data.europa.eu/doi/10.2777/1524> (last access: 14 May 2025), 2018.
- Ezard, T. H. G., Aze, T., Pearson, P. N., and Purvis, A.: Interplay Between Changing Climate and Species' Ecology Drives Macroevolutionary Dynamics, *Science*, 332, 349–351, 2011.
- Felden, J., Möller, L., Schindler, U., Huber, R., Schumacher, S., Koppe, R., Diepenbroek, M., and Glöckner, F. O.: PANGAEA – Data Publisher for Earth & Environmental Science, *Sci. Data*, 10, 347, <https://doi.org/10.1038/s41597-023-02269-x>, 2023.
- Fenton, I. S., Woodhouse, A., Aze, T., Lazarus, D., Renaudie, J., Dunhill, A. M., Young, J. R., and Saupe, E. E.: Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences, *Sci. Data*, 8, 160, <https://doi.org/10.1038/s41597-021-00942-7>, 2021.
- Finkel, Z. V., Katz, M. E., Wright, J. D., Schofield, O. M. E., and Falkowski, P. G.: Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic, *P. Natl. Acad. Sci. USA*, 102, 8927–8932, 2005.
- Finnegan, S., Harnik, P. G., Lockwood, R., Lotze, H. K., McClellan, L., and Kahanamoku, S. S.: Using the Fossil Record to

- Understand Extinction Risk and Inform Marine Conservation in a Changing World, *Ann. Rev. Mar. Sci.*, 16, 307–333, 2024.
- Georgescu, M. D.: *Microfossils through Time: An Introduction*, Schweizerbart Science Publishers, Stuttgart, Germany, ISBN 9783510654130, 2018.
- Hess, S., Alve, E., Andersen, T. J., and Joranger, T.: Defining ecological reference conditions in naturally stressed environments – How difficult is it?, *Mar. Environ. Res.*, 156, 104885, <https://doi.org/10.1016/j.marenvres.2020.104885>, 2020.
- Imbrie, J. and Kipp, N. G.: A new micropaleontological method for quantitative paleoclimatology: application to a late Pleistocene Caribbean core, in: *The late Cenozoic glacial ages*, edited by: Turekian, K. K., Yale University Press, New Haven, 71–181, 1971.
- Jonkers, L.: `lukasjonkers/mipaguidelines: Code for guidelines (Version v1)`, Zenodo [code], <https://doi.org/10.5281/zenodo.15411454>, 2025.
- Jonkers, L. and Strack, T.: Marine microfossil data guidelines survey results, NFDI4Earth Community on Zenodo [data set], <https://doi.org/10.5281/zenodo.12722701>, 2024.
- Jonkers, L., Hillebrand, H., and Kučera, M.: Global change drives modern plankton communities away from the pre-industrial state, *Nature*, 570, 372–375, 2019.
- Jonkers, L., Mix, A., Voelker, A., Risebrobakken, B., Smart, C. W., Ivanova, E., Arellano-Torres, E., Eynaud, F., Naoufel, H., Max, L., Rossignol, L., Simon, M. H., Martins, M. V. A., Petró, S., Caley, T., Dokken, T., Howard, W., and Kucera, M.: ForCenS-LGM: a dataset of planktonic foraminifera species assemblage composition for the Last Glacial Maximum, *Sci. Data*, 11, 361, <https://doi.org/10.1038/s41597-024-03166-7>, 2024.
- Khider, D., Emile-Geay, J., McKay, N. P., Gil, Y., Garijo, D., Ratnakar, V., Alonso-Garcia, M., Bertrand, S., Bothe, O., Brewer, P., Bunn, A., Chevalier, M., Comas-Bru, L., Csank, A., Dassié, E., DeLong, K., Felis, T., Francus, P., Frappier, A., Gray, W., Goring, S., Jonkers, L., Kahle, M., Kaufman, D., Kehrwald, N. M., Martrat, B., McGregor, H., Richey, J., Schmittner, A., Scroxton, N., Sutherland, E., Thirumalai, K., Allen, K., Arnaud, F., Axford, Y., Barrows, T. T., Bazin, L., Pilaar Birch, S. E., Bradley, E., Bregy, J., Capron, E., Cartapanis, O., Chiang, H. W., Cobb, K., Debret, M., Dommain, R., Du, J., Dyez, K., Emerick, S., Erb, M. P., Falster, G., Finsinger, W., Fortier, D., Gauthier, N., George, S., Grimm, E., Hertzberg, J., Hibbert, F., Hillman, A., Hobbs, W., Huber, M., Hughes, A. L. C., Jaccard, S., Ruan, J., Kienast, M., Konecky, B., Le Roux, G., Lyubchich, V., Novello, V. F., Olaka, L., Partin, J. W., Pearce, C., Phipps, S. J., Pignol, C., Piotrowska, N., Poli, M. S., Prokopenko, A., Schwanck, F., Stepanek, C., Swann, G. E. A., Telford, R., Thomas, E., Thomas, Z., Truebe, S., von Gunten, L., Waite, A., Weitzel, N., Wilhelm, B., Williams, J., Williams, J. J., Winstrop, M., Zhao, N., and Zhou, Y.: PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data, Paleooceanography and Paleoclimatology, 34, 1570–1596, 2019.
- Lazarus, D.: Neptune: A marine micropaleontology database, *Math. Geol.*, 26, 817–832, 1994.
- Lazarus, D., Barron, J., Renaudie, J., Diver, P., and Türke, A.: Cenozoic planktonic marine diatom diversity and correlation to climate change, *PLoS One*, 9, e84857, <https://doi.org/10.1371/journal.pone.0084857>, 2014.
- Lazarus, D. B., Renaudie, J., Lenz, D., Diver, P., and Klump, J.: Raritas: a program for counting high diversity categorical data with highly unequal abundances, *PeerJ*, 6, e5453, <https://doi.org/10.7717/peerj.5453>, 2018.
- Lowery, C. M., Bown, P. R., Fraass, A. J., and Hull, P. M.: Ecological Response of Plankton to Environmental Change: Thresholds for Extinction, *Annu. Rev. Earth Pl. Sc.*, 48, 403–429, 2020.
- MARGO project members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nat. Geosci.*, 2, 127–132, 2009.
- McKay, N. P. and Emile-Geay, J.: Technical note: The Linked Paleo Data framework – a common tongue for paleoclimatology, *Clim. Past*, 12, 1093–1100, <https://doi.org/10.5194/cp-12-1093-2016>, 2016.
- Millard, A. R.: Conventions for Reporting Radiocarbon Determinations, *Radiocarbon*, 56, 555–559, 2014.
- Patterson, R. T. and Fishbein, E.: Re-examination of the statistical methods used to determine the number of point counts needed for micropaleontological quantitative research, *J. Paleontol.*, 63, 245–248, 1989.
- Peeters, F., Ivanova, E., Conan, S., Brummer, G.-J., Ganssen, G., Troelstra, S., and van Hinte, J.: A size analysis of planktic foraminifera from the Arabian Sea, *Mar. Micropaleontol.*, 36, 31–63, 1999.
- Petersen, M., Glöckler, F., Kiessling, W., Döring, M., Fichtmüller, D., Laphakorn, L., Baltruschat, B., and Hoffmann, J.: History and development of ABCDEFG: a data standard for geosciences, *Mitt. Mus. Nat. Berl. Foss. Rec.*, 21, 47–53, 2018.
- Schiebel, R., Smart, S. M., Jentzen, A., Jonkers, L., Morard, R., Meilland, J., Michel, E., Coxall, H. K., Hull, P. M., de Garidel-Thoron, T., Aze, T., Quillévéré, F., Ren, H., Sigman, D. M., Vonhof, H. B., Martínez-García, A., Kučera, M., Bijma, J., Spero, H. J., and Haug, G. H.: Advances in planktonic foraminifer research: New perspectives for paleoceanography, *Revue de Micropaléontologie*, 61, 113–138, 2018.
- Schlagintweit, F. and Simmons, M.: Developing best practice in micropalaeontology: Examples from the mid-Cretaceous of the Zagros Mountains, *Actapalrom*, 63–84, 2022.
- Schmidt, D.: Determining climate change impacts on ecosystems: The role of palaeontology, *Palaeontology*, 61, 1–12, 2018.
- Schönfeld, J., Alve, E., Geslin, E., Jorissen, F., Korsun, S., and Spezzaferri, S.: The FOBIMO (FORaminiferal BIO-MONitoring) initiative – Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies, *Mar. Micropaleontol.*, 94–95, 1–13, 2012.
- Schott, W.: Die Foraminiferen in dem äquatorialen Teil des Atlantischen Ozeans, in: *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff Meteor 1925–1927*, edited by: Correns and Schott, vol. 3, 43–134, De Gruyter GmbH, 1937.
- Self-Trail, J. M. and Seefelt, E. L.: Rapid dissolution of calcareous nannofossils: a case study from freshly cored sediments of the south-eastern Atlantic Coastal Plain, *Journal of Nannoplankton Research*, 27, 149–158, 2005.
- Smith, J. A., Rillo, M. C., Kocsis, Á. T., Dornelas, M., Fastovich, D., Huang, H.-H. M., Jonkers, L., Kiessling, W., Li, Q., Liow, L. H., Margulis-Ohnuma, M., Meyers, S., Na, L., Penny, A. M., Pippenger, K., Renaudie, J., Saupe, E. E., Steinbauer, M. J., Sugawara, M., Tomašovych, A., Williams, J.



- W., Yasuhara, M., Finnegan, S., and Hull, P. M.: BioDeep-Time: A database of biodiversity time series for modern and fossil assemblages, *Glob. Ecol. Biogeogr.*, 32, 1680–1689, <https://doi.org/10.1111/geb.13735>, 2023a.
- Smith, J. A., Raja, N. B., Clements, T., Dimitrijević, D., Dowding, E. M., Dunne, E. M., Gee, B. M., Godoy, P. L., Lombardi, E. M., Mulvey, L. P. A., Nätscher, P. S., Reddin, C. J., Shirley, B., Warnock, R. C. M., and Kocsis, Á. T.: Increasing the equitability of data citation in paleontology: capacity building for the big data future, *Paleobiology*, 1–12, 2023b.
- Spezzaferri, S., Olsson, R. K., Hemleben, C., Wade, B. S., and Coxall, H. K.: Taxonomy, biostratigraphy, and phylogeny of Oligocene and lower Miocene Globoturbotalita, in: *Atlas of Oligocene Planktonic Foraminifera*, vol. 46, edited by: Wade, B. S., Olsson, R. K., Pearson, P. N., Huber, B. T., and Berggren, W. A., Cushman Foundation of Foraminiferal Research, 231–268, ISBN 9781970168419, 2018.
- Strack, T., Jonkers, L., Huber, R., and Kucera, M.: Reusability of data with complex semantic structure, *Zenodo*, <https://doi.org/10.5281/zenodo.8124211>, 2023.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglais, D.: Darwin Core: an evolving community-developed biodiversity data standard, *PLoS One*, 7, e29715, <https://doi.org/10.1371/journal.pone.0029715>, 2012.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>, 2016.
- Yasuhara, M., Hunt, G., Breitburg, D., Tsujimoto, A., and Katsuki, K.: Human-induced marine ecological degradation: micropaleontological perspectives, *Ecol. Evol.*, 2, 3242–3268, 2012.
- Yasuhara, M., Tittensor, D. P., Hillebrand, H., and Worm, B.: Combining marine macroecology and palaeoecology in understanding biodiversity: microfossils as a model, *Biol. Rev. Camb. Philos. Soc.*, 92, 199–215, 2017.
- Yasuhara, M., Huang, H.-H., Hull, P., Rillo, M., Condamine, F., Tittensor, D., Kučera, M., Costello, M., Finnegan, S., O'Dea, A., Hong, Y., Bonebrake, T., McKenzie, R., Doi, H., Wei, C.-L., Kubota, Y., and Saupe, E.: Time Machine Biology: Cross-Timescale Integration of Ecology, Evolution, and Oceanography, *Oceanography*, 33, 16–28, 2020.