Microplastics and Trash Cleaning and Harmonization (MaTCH): Semantic Data Ingestion and Harmonization Using Artificial Intelligence

Journal
Environmental Science and Technology, 58(46)

ISSN
0013-936X

Authors
Hapich, Hannah
Cowger, Win
Gray, Andrew B

Publication Date
2024-11-19

DOI
10.1021/acs.est.4c02406

Environmental Science & Technology

Open Access

pubs.acs.org/est

Article

Microplastics and Trash Cleaning and Harmonization (MaTCH): Semantic Data Ingestion and Harmonization Using Artificial Intelligence

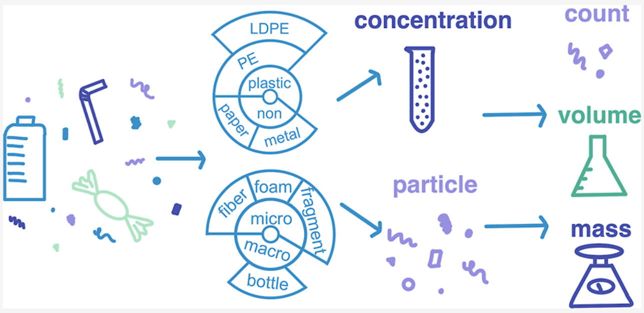
Hannah Hapich,* Win Cowger, and Andrew B. Gray

Cite This: *Environ. Sci. Technol.* 2024, 58, 20502−20512

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: With the rapid expansion of microplastic research and reliance on semantic descriptors, there is an increasing need for plastic pollution data harmonization. Data standards have been developed but are seldom implemented across research sectors, geographic regions, environmental media, or size classes of plastic pollution. Harmonization of existing data is currently hindered by increasingly large datasets using thousands of different categorical variable descriptors, as well as various metrics used to describe particle abundance and differing size ranges studied across groups. For this study, we used manually developed relational databases to build an algorithm utilizing artificial intelligence capable of automatically curating harmonized, more usable datasets describing micro to macro plastic pollution in the environment. The study algorithm MaTCH (microplastics and trash cleaning and harmonization) can harmonize datasets with different formats, nomenclature, methods, and measured particle characteristics with an accuracy of 71−94% when matching semantically. All other non-semantic corrections are reported within a 95% confidence interval and with model uncertainty. All steps of the algorithm are integrated in an open-source software tool for the benefit of the scientific community and ease of integration for all plastic pollution data.



KEYWORDS: *plastic, microplastics, trash, natural language processing, harmonization, data management, artificial intelligence*

1. INTRODUCTION

Studies focused on trash (mismanaged waste > 5 mm in length[1]) and microplastic (plastics 1−5000 μm in length[1]) pollution have increased dramatically in number over recent years.[2] Together, these studies have indicated that the majority of microplastics found in the environment are secondary products of degrading mismanaged plastic waste rather than primary emissions, pointing to a relationship in environmental occurrence between trash and microplastics.[3] Microplastics and trash are diverse environmental pollutants that are difficult to query and quantify, as we generally describe them with incomparable categorical variables, and report environmental concentrations composed of varying reporting metrics and particle size ranges.[4] Microplastics data is not currently standardized and is therefore less easily or reliably comparable between studies,[5,6] leading to many calls for both standardization[6−9] and harmonization.[5,10,11] Nearly all propositions for standardization or harmonization have focused either on nano, micro,[5,9] or macro[8,12] particles, whose size domain thresholds are often arbitrary and inconsistent between groups.[13,14] Given the intrinsic relationship between trash and microplastics in their environmental occurrence and categorical semantics, it follows that data management strategies for microplastics and trash should be harmonious.[15]

Standards have been developed for managing mandated trash assessment data[8] and tabulating microplastics data with respect to specific reporting guidelines.[9] Such strategies serve as hubs for accumulating already standardized data and are often specific to certain geographic regions, study media, or government protocols.[8,10,16] Certain database structures may be better suited for data at the sample level (reported as a concentration) or the particle level (information reported for individual particles). Standardization is limited by the rate at which scientists, government organizations, nongovernmental organizations (NGOs), or industry adapt to such protocols. Additionally, most protocols do not have a strategy to utilize data that does not fit their standardized structures, which alienates potentially useful data. In these cases, users must perform data harmonization manually. It is particularly important in the field of microplastics monitoring to utilize

Received: March 7, 2024
Revised: September 24, 2024
Accepted: October 31, 2024
Published: November 11, 2024

existing databases due to the cost and time prohibitive nature of the field, wherein it commonly requires up to thousands of dollars and tens of hours to process a single sample, making data from each monitoring study highly valuable.

Harmonization posthoc methods, wherein the harmonization of data occurs after a study is complete and no prior consideration for data standards is necessary, allow the inclusion of more data but are limited by the ease at which foreign data can be assimilated to a common structure.[12] Manual harmonization can take weeks for even a single dataset depending on its size, again leading to underutilization of plastics databases.[2] Previous work on harmonizing trash data has focused on manually extracting categorical variables to develop broad encompassing data structures, including databases such as the Trash Taxonomy, wherein terms used to describe trash characteristics are detailed in a relational table database.[12] However, there exists no approach that automatically harmonizes macro debris and microplastics data from nonspecified formats or with unknown categorical descriptors.[2] Best practices for the development of database structures have remained a manual undertaking that should be performed with the input of a wide array of stakeholders, though the addition of new terms is prohibitively effort-intensive. An automated approach to data harmonization would allow for quick ingestion of data from new studies, leading to larger, more valuable databases.

The field of microplastics and trash is not the first to encounter such issues. Many divisions of the environmental and biological sciences have similar problems, which will worsen over time with ever-growing datasets and a focus on curating "big data" to identify knowledge gaps and answer key questions.[17] Previous work has assessed the use of natural language processing (NLP) algorithms as a means for information retrieval to assemble databases and organize their taxonomic structures.[18,19] Until recently, the technology available consisted of different pattern matching and syntactic/semantic parsing, some of which rely on extracting exact matches, and most have a narrow application range tailored to a specific subfield.[19] Results from early exploration of NLP for scientific data curation were discouraging[20] and may have led to underutilization.

NLP technology has vastly improved in accuracy and efficiency just over the past few years, primarily a result of increases in computing power and the development of open-source artificial intelligence (AI) software capable of employing transformers and embeddings.[21] Transformers are a type of neural network structure able to interpret data nonsequentially.[22] The first step in using a transformer is to encode words as vectors, of which embeddings are one of the most efficient vector types to compare and derive meaning. Advanced open-source language models, such as BERT (bidirectional encoder representations from transformers)[21] and GPT (generative pre-training transformer) (provided through OpenAI),[23] have made it possible for scientists to now utilize the power of embeddings and AI to rapidly automate harmonization of categorical data—including those for trash and microplastics.

In addition to semantic harmonization—referring to combining categorical variables based on their semantic meaning in addition to their structural similarity—effective and transferrable microplastics data reporting is challenged by the incongruence of concentration data spanning different particle size ranges and a paucity of the particle-level reporting required to minimize error and ensure accurate representation. Microplastic sample concentration is most often reported on a particle count basis and is dependent on the size range of microplastics characterized, as microplastic occurrence data show particle counts generally increase with decreasing size in the form of an inverse power law relationship.[24−27] Inconsistencies may also arise from differing reported dimensions (i.e., length, mass, projected surface area, and volume).[26] Our framework includes non-semantic harmonization methods that both rescale studied-size ranges and convert between count, volume, and mass given basic particle characteristics. Our framework also attempts to improve the accuracy and transparency of such non-semantic harmonization methods by introducing a large database of polymer densities ($n > 77{,}000$) and reporting 95% confidence intervals and model error.

This study aims to (1) create a data management workflow to merge terminologies for both trash and microplastics in one unified querying system, (2) integrate an automated NLP step into the workflow to account for foreign terminology, (3) create an automated strategy for rescaling heterogeneous microplastic concentrations and particle metrics, (4) validate the performance of the automated data harmonization pipelines relative to manual curation, (5) generate open-source web tools for scientists to rapidly leverage the created algorithm, and (6) create a generalizable approach to perform such analyses so they may be applied in other fields. Through the use of NLP and newly assessed data structures, we aim to ensure that all descriptors of these complex pollutants are retained during harmonization to provide datasets that are comparable between studies without losing any of the rich descriptive information provided to better support plastic pollution investigation and future management.

2. MATERIALS AND METHODS

2.1. Model Development: Semantic Data Harmonization. *2.1.1. Microplastics and Trash Taxonomy.* To assess the current lexicon of microplastics researchers in describing particles, we used Google Scholar to search for studies related to microplastic occurrence in drinking water (75% tap and 25% bottled) and an environmental compartment of interest—rivers—including multiple media types with a focus on surface water (95% surface water, 4% sediment, and 1% river discharged effluent). In total, we reviewed 57 studies totaling 1186 samples (Table S1), recording concentration data and a comprehensive list of all unique terms used to describe particles, all of which included some instrumental verification that particles found were plastic. Data was manually extracted from article text and reported in the format described in SI file "User Guide", Section 3 "Data Structuring". None of the reviewed studies (Table S1) reported individual particle data; therefore, we focused on harmonizing sample level data for our review.

Our goal was to build on the existing Trash Taxonomy to integrate terms describing both microplastics and macro debris into one querying system.[12] We created alias tables that first relate any synonymous terms (e.g., "LDPE" and"low density polyethylene"), as well as hierarchical tables to relate child and parent terms [e.g., "PE" (polyethylene) is the parent term to both "LDPE" (low-density polyethylene) and "HDPE" (high-density polyethylene)] (Figure S1). Through conducting this review, the primary reported categorical characteristics were the morphology (shape), material, and color. All types of categorical variables have their own alias tables, and hierarchical tables were generated for material and morphology. Similarity between terms used to describe morphology and material is calculated via eq 1

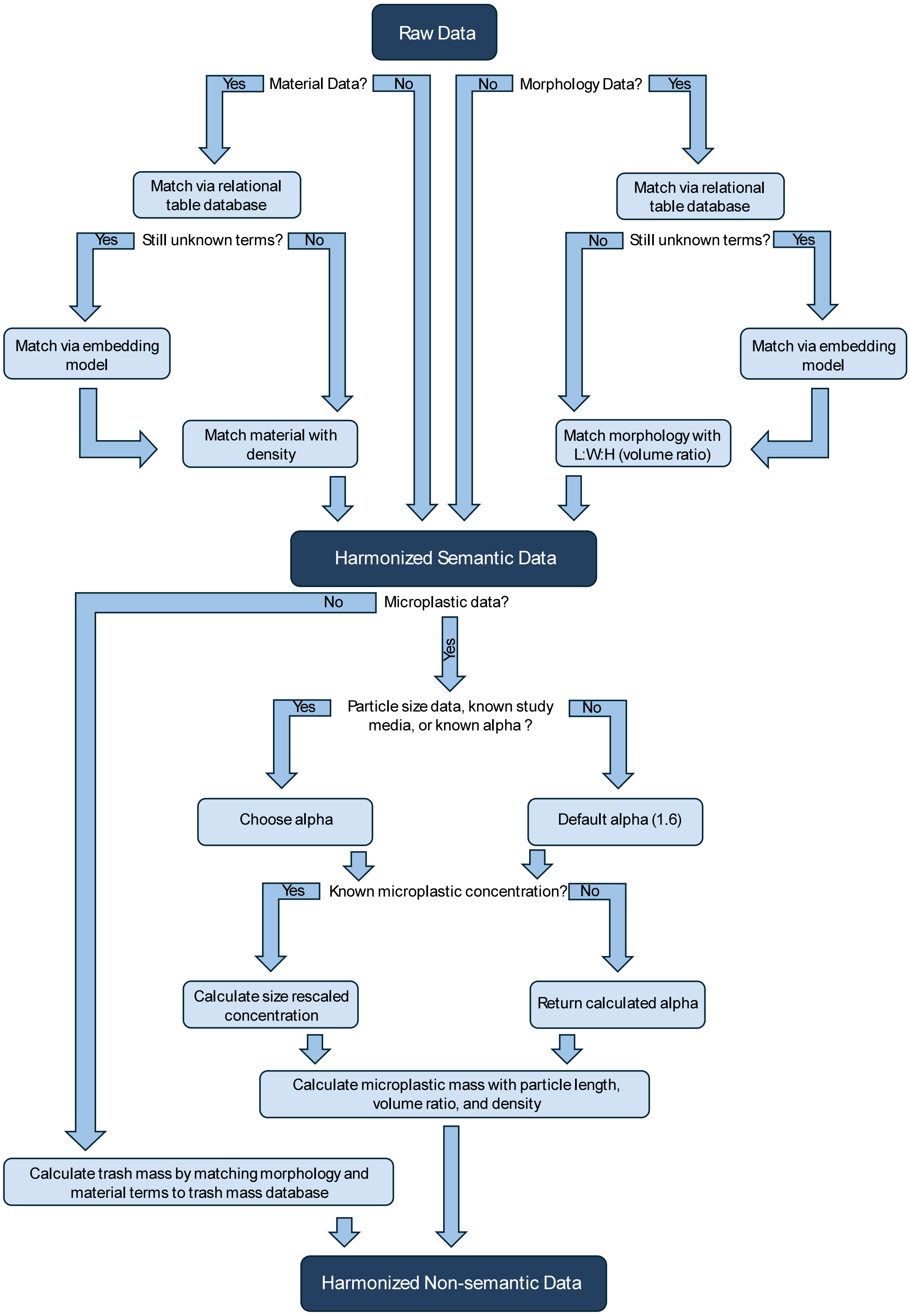


Figure 1. Microplastics and trash cleaning and harmonization (MaTCH) workflow schematic. Dark blue boxes represent data inputs and outputs. Light blue boxes represent model operations. Text outside of boxes and medium blue arrows represent decision trees.

$$\text{Comparability metric}_{X,Y} = \frac{\sum \text{Terms in sheet } X \text{ equivalent with terms in sheet } Y}{\sum \text{All terms in sheet } Y} \tag{1}$$

Equation 1: Comparability metric to be computed for material or morphology between two sheets (sheet X and sheet Y), where the numerator describes the number of terms present in both sheet X and sheet Y, and the denominator describes the total number of terms in sheet Y.

https://doi.org/10.1021/acs.est.4c02406
Environ. Sci. Technol. 2024, 58, 20502−20512

Microplastic morphologies can be easily differentiated from trash, as they all have the parent term "microplastic" in the morphology hierarchy table. To differentiate materials, specific polymer names to describe microplastics are all under the parent term "plastic" in the materials hierarchy table. These merged microplastic and trash taxa allow all plastic pollution data to be studied contemporaneously, regardless of size.

2.1.2. Embeddings and Querying through AI. We used the text-embedding-ada-002 embedding model and API from OpenAI[28] to generate our embeddings—the vector described earlier needed for relating terms via NLP, with each term having a corresponding vector embedding—deployed in R (4.3.0) and RStudio (23.09.0). This model was chosen on the basis of accuracy, speed, and cost. Accuracy was assessed by attempting to reproduce our alias relational tables through matching aliases to their associated prime terms by the similarity of their associated embeddings.

The alias tables consist of one "prime" term that acts as the key linking to other tables in the relational database and other synonymous terms as "aliases". Embeddings were generated for 1326 morphology aliases and 609 material aliases. Of these, 441 morphologies and 406 materials were designated to be "prime" terms, and 885 morphologies and 203 materials were designated as synonymous "alias" terms. Using the "chRoma" package[29] that streamlines vector database management, we queried the top five matches as determined via dot product between each embedding vector (see data availability statement to access open-source code). We compared these results to the actual prime term for that alias in the alias tables—as outlined in Section 2.1.1—and reported those that were correctly matched to their prime term, as well as those that contained the correct prime term in the top 5 matches as ranked by percent similarity. We also used a database not previously integrated from an urban litter study[30] and compared embedding matches to manual matches.

2.2. Model Development: Non-Semantic Data Harmonization. *2.2.1. Particle Count to Mass Conversion.* For the sake of simplicity, when trying to traverse between count to mass-based concentration estimates, some have assumed a constant density and spherical volume, which greatly increases error margins.[31] Kooi and Koelmans (2019)[26] identified the problem of applying subjective morphological and polymeric descriptors to microplastic particles and rather proposed reporting as continuous distributions. In their study, ranges were established for the *L:W:H* (length to width to height) ratios of five common morphology types (fiber, fragment, film, foam, and sphere).[26] We have integrated these proposed *L:W:H* ratios into our algorithm to convert morphology and length information into particle volume. Error was calculated by deriving a 95% confidence interval for each axis given the reported possible ranges, assuming a normal distribution. A measurement error of ±5% for particle length was also included to better characterize all possible sources of uncertainty.

To obtain mass, we then multiply the volume estimates, as outlined above, by material density. To obtain density, we curated a database including over 77,000 polymer density measurements (provided by Dale Kipp of MatWeb),[32] for which we computed median values,[33] adding to existing databases[26,34] of $n = \sim 1000$ polymer densities. If only a range of densities were provided, a Gaussian distribution was assumed, and a 95% confidence interval was derived. If individual density measurements were available, as in the case of our internally curated dataset, confidence intervals were derived using actual values. If no range or individual measurements were provided, then the average coefficient of variation of polymers with known error was applied. Our updated database contains 310 unique densities that are specific constituents of 43 parent polymer classes. By using densities derived from a larger number of measurements, we hope to better encompass the possible ranges of actual plastic particle densities while better representing uncertainty (Figure 1). In order to estimate mass for macro debris, we have used the method outlined in Cowger et al. (2022), which includes a literature review of studies that report litter masses to obtain average values for 205 distinct trash morphologies (Figure 1).[30]

If only sample level data is provided, we simulate particle data according to the morphological/material proportions provided and then follow the workflow mentioned above. Additionally, although we expect a majority of user input terms to be known within our database, the use of embedding models to describe polymer or morphology could lead to compounding error when calculating mass if improper matching was to arise for morphology or material.

2.2.2. Particle Size Rescaling. Investigations of continuous distributions of microplastic sizes have found that smaller microplastics contribute a much larger proportion to total concentrations by count.[26,34] Kooi et al. (2021) derived several inverse power law models[26] to describe microplastic size distributions for various environmental media.[34] In their approach, a meta-analysis was performed on studies that report microplastic size and relative abundance to develop alpha values (power law exponents) for probability density functions describing size. In the current study, we applied this method to any dataset with more than 5 size bins (the recommended minimum n by Nor et al., 2021[35]) to derive an alpha value (eq S1) (Figure 1).[26] This value is used to calculate a correction factor (eq S2),[27] which is then multiplied by the given concentration to describe the corrected range, which may be larger or smaller than the original value depending on if you want to expand or collapse your size range (Figure 1). Additionally, we integrated the standard error from the inverse power law model into our total error budget. Note, particle size rescaling is contingent only upon sample size information and not on any particle characteristic distributions, meaning there will be no compounding error derived from embedding match uncertainty. Additionally, note that rescaling is not applicable to macro debris (Figure 1). Degradation pathways suspected to be the cause for observed microplastic particle size distributions (PSDs) would not apply to macro debris that is typically a result of primary production, not secondary degradation.

2.3. Model Testing: Drinking Water vs Riverine Microplastic Occurrence Meta-Analysis Harmonization. To assess the applicability of our harmonization model to real world plastic pollution data, we merged curated datasets of microplastic occurrence in drinking water ($n = 600$) and in rivers across the United States ($n = 586$) (Table S1). These sources were chosen due to their suspected incompatibility in terms of particle size and concentration, with drinking water studies assumed to have lower concentrations and smaller particles observed. Our goals were to assess differences in semantics and our ability to make cross-study comparisons more concise as well as rescaling concentrations to analyze subsequent transformations in both occurrence and semantics.

To rescale the concentrations, correction factors were obtained for each sample. Values from Kooi et al. (2021) were used for studies in freshwater surface, freshwater sediment, and effluent discharged into streams ($\alpha = 2.64 \pm 0.01$, 3.25 ± 0.19, and 2.54 ± 0.04, respectively) (Figure 1). An alpha value for drinking water was developed ($\alpha = 1.64 \pm 0.55$), in which 11 samples were isolated that met the criteria of having ≥5 reported-size bins. The method described in Section 2.2.2 above was used to obtain a corrected concentration that describes the full 1−5000 μm range.[34]

When semantic data was provided, morphology and polymer were converted to one of the prime terms in the database presented in this study (Figure 1). Correcting various samples with differing size ranges, as we have done in our meta-analysis, will also have implications on the final proportions of categorical variables, as each study has a unique correction factor. Results displaying the reported and corrected concentrations, morphologies, and materials are discussed below. While this case study utilizes our model's non-semantic capability to correct for differing size ranges, see the Supporting Information section titled "Model Testing: Micro vs. Macro Roadway Debris -Meta-Analysis Harmonization" for a similar case study of count to mass conversions for different sized roadway debris.

2.4. Open-Source Web-Tools. We developed a simple user interface—the Microplastics and Trash Cleaning and Harmonization (MaTCH) app—that will perform all possible harmonization with a single upload is available at https://hannahhapich.shinyapps.io/match/. MaTCH will analyze the column headings in the uploaded data table to determine the format (particle or concentration) and which data cleaning operations can be performed (semantic matching, count to mass conversion, and size rescaling). Particle and concentration test data are available for download to illustrate MaTCH's features. This app was developed with the shiny,[36] tidyr,[37] skimr,[38] chRoma,[29] tibble,[39] dplyr,[40] data.table,[41] data.tree,[42] plotly,[43] bs4 Dash,[44] classInt,[45] aws.s3,[46] digest,[47] DT,[48] shinyTree,[49] shinyhelper,[50] and shinyWidgets[51] packages in R (4.3.0) and RStudio (23.09.0).

2.4.1. Semantic Merging Interface. If material and morphology are provided, a match through our relational table system is first performed (Section 2.1.1), and if the term is unknown, an embedding match is performed (Section 2.1.2). The algorithm defaults to the top embedding match, but also provides a dropdown list with the top five matches that allows the user to manually override this selection. With user approval, reported particle characteristic frequencies are saved to an Amazon S3 Cloud Database to help inform future algorithm development by indicating most commonly found particle types. Additionally, the tool provides rapid data visualization in the form of sunburst plots (hierarchical pie charts) for materials and morphologies and is available for download.

2.4.2. Non-Semantic Merging Interface. The count to mass conversion tool allows users to upload data that must include at least length, morphology, and material. Additional variables can be included as outlined in the data template tab on the tool; however, they are not required and therefore do not need to be a complete column, as in the case of the "Sample Test Data" available on the homepage. Data may be uploaded in the form of particle data with actual values for each or as concentration data with particle size and proportions for morphology and material. Using the model and error calculations detailed above (Section 2.2.1), the output consists of volume, density, mass, and confidence intervals for each.

We also developed a tool for concentration rescaling that allows users to upload microplastic concentrations of their sample(s), studied size range, and desired extrapolated size range (within the 1−5000 μm microplastics size limit) to obtain rescaled concentrations. Users have the option to upload concentration data or particle level data. If binned concentration data with ≥5 bins or particle data are available, an alpha value will be generated for the sample calculated in accordance with methods detailed above (Section 2.2.2). If no study media or <5 size bins are provided, a default of $\alpha = 1.6 \pm 0.5$ will be used.[26] These alpha values are used to derive a correction factor and corrected concentration for accompanying sample data.

3. RESULTS AND DISCUSSION

3.1. Model Development: Semantic Data Harmonization. Of the 57 microplastic studies reviewed, only 13 reported their findings for categorical descriptors of microplastics. While several studies reported microplastic colors present in their samples, only three studies in our analysis reported proportion values, and color was therefore omitted from our comparability analysis. In total, this allowed for 156 comparability metrics to be derived analyzing the comparability of microplastic morphology and material (Figure S2).

Using the comparability metric (eq 1), we found mean comparability between unharmonized studies was 34.0% for morphologies and 6.4% for materials. The number of 100% comparable observations was 31 for morphologies and 0 for materials. Alternatively, the number of 0% comparable studies was 80 for morphologies and 136 for materials. Our findings illustrate both a lack in reporting characteristic distributions among microplastic studies and little harmonization between groups that do report.

Through the validation of the MaTCH workflow, we found that 70.62% of morphologies and 76.14% of materials were a top match to their correct aliases (Figure S2). Additionally, 87.91% of morphologies and 94.32% of materials had a correct match in the top 5 matches (Figure S2). We also assessed terms outside of our database obtained from a litter accumulation study on urban roadways in Southern California[30] and a microdebris roadway study[52] to further verify accuracy. Of the total 204 descriptors used between these studies, 119 terms were unknown to our database, of which 93 (78%) were correctly matched to a prime term as compared to the manually harmonized dataset. Note that the results here reflect those obtained from running the analysis at the time of publication. The OpenAI embedding model used here is subject to change over time, whether it be due to actual changes in the model or intentionally injected randomness, so exact percentages may vary slightly over time. Also, note that the accuracy percentages reported above assume all terms must be matched via embeddings. Terms are first matched via relational table, and it is our expectation that this database—with thousands of terms—will be able to match user input descriptors a majority of the time, resulting in 100% accuracy for those terms; the use of the embedding model acts more as a secondary failsafe method.

However, there are still some limitations to embedding-based semantic matching strategies. When attempting to match full polymer names to common abbreviations (e.g., "polyethylene" and "PE"), our matches were much lower (38.56−47.32%). In embedding space, unabbreviated polymer names were more closely related to each other than to their corresponding abbreviations. For example, "polyethylene" would be closer to "poly(ethylene glycol)" than to its associated abbreviation "PE". Conversely, "PET" (polyethylene terephthalate) would be more closely related to "PE" than to "polyethylene terephthalate". To account for this technical limitation, known polymer abbrevia-

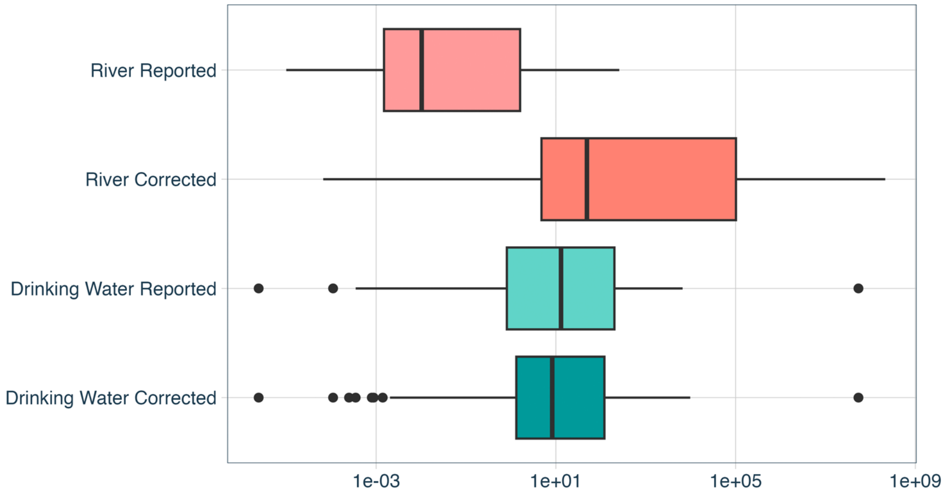


Figure 2. Microplastic concentrations (particle/L) before and after particle size rescaling. Data sourced from meta-analyses of the present study (Table S1) including data from microplastics monitoring of riverine systems (n = 586) and drinking water (n = 600).

tions were excluded from our embedding database, though they are still included in the relational tables. This means incoming unknown terms will be matched to an unabbreviated alias first (e.g., "poly(ethylene)" would match to "polyethylene") and then matches to common abbreviations are done via the relational tables (e.g., to "PE").

We have developed a data validation routine that allows users in any field to assess the accuracy of embedding matches for their chosen field, available via our GitHub page. If user desired accuracy is achieved, our algorithm can be applied to any use case by swapping out relational tables. Also, note that our model has not undergone any fine-tuning (model training of user data) to better characterize our use case. With proper funding and sufficient training data, the accuracy of embedding matches will improve through fine-tuning.

Supported by the high accuracy found in our study, integration of embedding-based NLP and other AI technologies to match synonymous semantic terminology appears promising. These results illustrate the power of NLP to increase the automation of categorical data harmonization. Automated curation of big datasets will increase the usability of existing data and may expedite future research studies that rely on the synthesis of microplastics data. With low-to-no-cost models becoming more accessible, we believe future studies should consider these methods of data integration in the field of plastic pollution research and beyond.

3.2. Model Testing: Drinking Water vs Riverine Microplastic Occurrence Meta-Analysis Harmonization. Median concentrations of all drinking water studies curated were found to be 9.0 particles/L, which when run through MaTCH and corrected to the full microplastic particle size range (1−5000 μm) decreased to 6.03 ± 2.19 particles/L (Figure 2). Some drinking water studies included size ranges beneath 1 μm (filters with 0.2 μm pore size and no reported lower limit), resulting in a slight decrease in concentration when fitting to the 1−5000 μm distribution. Median river concentrations were 0.004 particles/L, increasing 4 orders of magnitude to 23.13 ± 1.53 particles/L when run through MaTCH (Figure 2). In contrast to drinking water, riverine studies focused on larger size ranges (median studied range of 355−5000 μm), likely due to analytical hurdles when dealing with environmental media. This resulted in a dramatic increase in concentration when size ranges were rescaled to the expanded 1−5000 μm range. This shift in riverine concentrations reversed the plotted interpretation of which media have a higher concentration. Though differences in analyzed size ranges were simply a result of differing objectives and analytical limitations across studies, when trying to draw a broad conclusion about particle abundance in different environmental compartments, uncertainties and errors can hide between microplastics datasets that are misaligned with one another.

For studies reporting on morphology or polymer composition of their samples, categorical terms were run through MaTCH. Terms used to describe morphology were reduced from 15 to 6, and polymer terms were reduced from 39 to 25. Parent hierarchical terms were plotted in sunburst plots whether or not parent terms were explicitly used in sample data to help visualize all scales of classification simultaneously (Figure 3a,b). Using hierarchical structuring maximizes characteristic data comparability between studies while still retaining more detailed information from studies that use more specific terminology to better inform non-semantic harmonization techniques (Figures 1 and 3).

As mentioned previously, correcting concentrations by rescaling particle size ranges also has implications for shifts in categorical variable proportions when comparing multiple studies. Some of the notable changes in morphological proportions after being run through MaTCH are an apparent 11% decrease in fibers and an 18% increase in nurdles for drinking water studies, as illustrated below (Figure 4a). Even greater changes were found for riverine studies, with an apparent 86% decrease in the number of fibers and a 34% increase in the number of fragments. Lower changes in drinking water morphological proportions with size rescaling are likely due to the tendency for drinking water studies to focus on a similar size range to the full distribution (1−5000 μm), meaning that their correction factors will be smaller than those only analyzing larger microplastics. Corrected concentrations also shifted material proportions (Figure 4b). Notably, corrected drinking water concentrations saw an apparent decrease in PET (polyethylene terephthalate) by 34% and an increase in PEST (polyester) by 164%. For riverine studies, we saw an apparent 61% increase in PS (polystyrene) and a 55% decrease in PE (polyethylene).

Examining differences in morphological makeup before and after being run through MaTCH may help to provide insight into some methodological limitations. Though values obtained

https://doi.org/10.1021/acs.est.4c02406
Environ. Sci. Technol. 2024, 58, 20502−20512

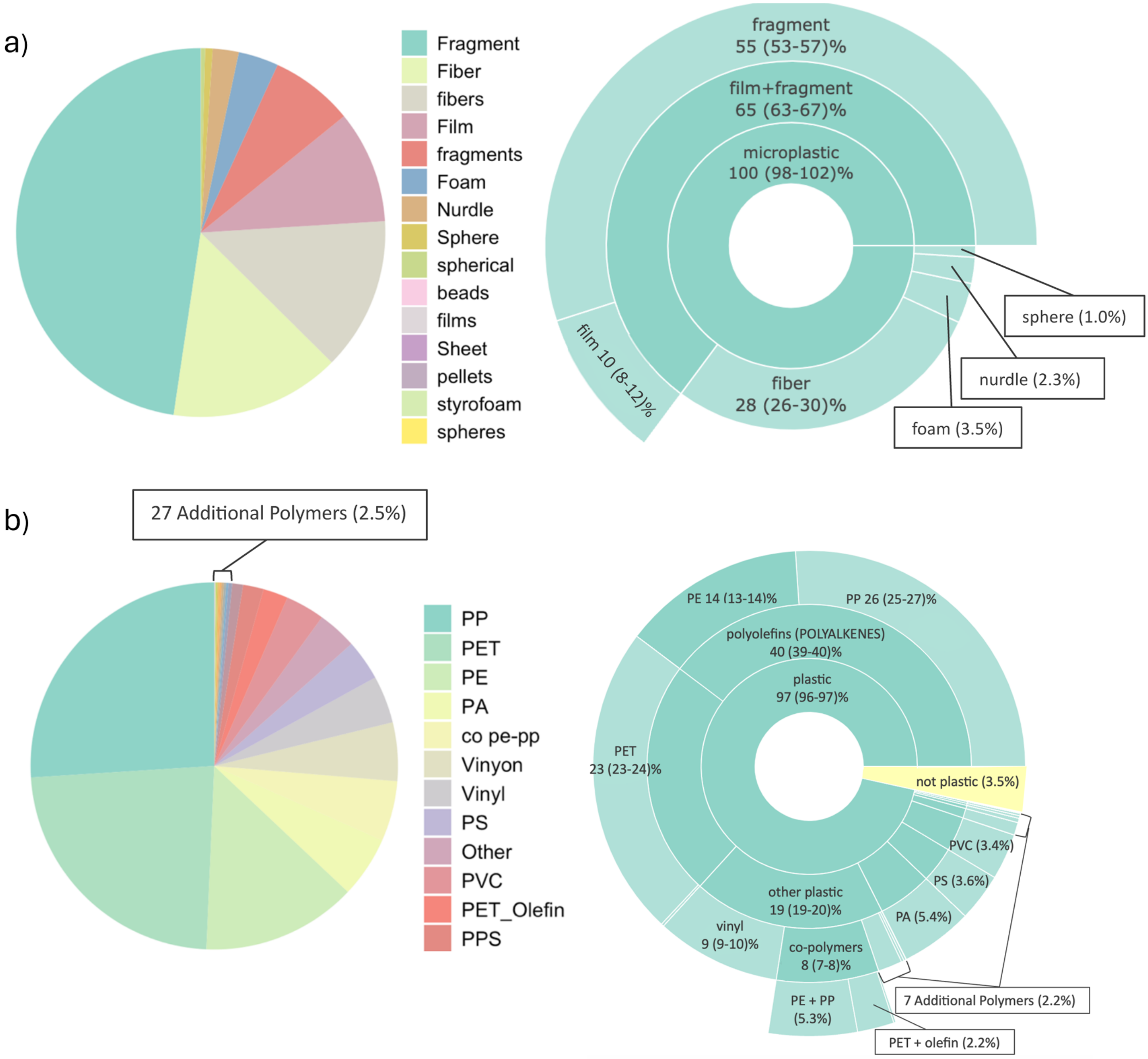


Figure 3. Change in morphological (a) and polymeric (b) categorical variables before and after semantic alignment. Illustration of hierarchical lumping via sunburst plot with 95% confidence intervals. (PP: polypropylene, PET: polyethylene terephthalate, PE: polyethylene, PA: polyamide, PS: polystyrene, PVC: polyvinyl chloride, and PPS: polyphenylene sulfide).

through MaTCH do contain a high degree of uncertainty, we believe that the analytical output is representative of the mean tendency. As drinking water studies tend to analyze smaller particles, it is possible that they are underreporting fibers at a higher proportion than environmental monitoring studies. Multiple reviews have found fibers to be the most abundant microplastic morphology in the environment,[53−55] meaning our findings may highlight decreasing recovery of selective morphologies as the analyzed size range decreases, assuming a constant PSD across morphologies. Given that most studies define microplastic size by the longest axis, it follows that fibers and their very high length to width ratio have a lower projected surface area as compared to other morphologies. Therefore, fibers are more likely to be beneath the limit of diffraction for common spectral analysis methods such as Fourier transform infrared spectroscopy (FTIR) than their morphological counterparts within the same size class when classes are defined on the basis of the longest principal axes. We also see a much higher proportion of fibers reported in riverine data that target larger particle sizes than we see in drinking water studies or in size range-adjusted concentrations of riverine samples.

Conversely, there may also exist variable recovery rates between morphologies based on the sampling method, regardless of the studied size range. Net-based sampling and defining size classes by mesh size but defining particle size by the longest axis can propose a problem for morphologies such as fibers. Fibers are unique due to their flexibility and previously mentioned high length to width ratio, giving them the potential to deform and slip through nets, despite the length of the fiber being larger than the mesh size. This same issue of variable recovery rates applies to any sample that undergoes sieving, meaning this would lead to the opposite effect as mentioned in

https://doi.org/10.1021/acs.est.4c02406
Environ. Sci. Technol. 2024, 58, 20502−20512

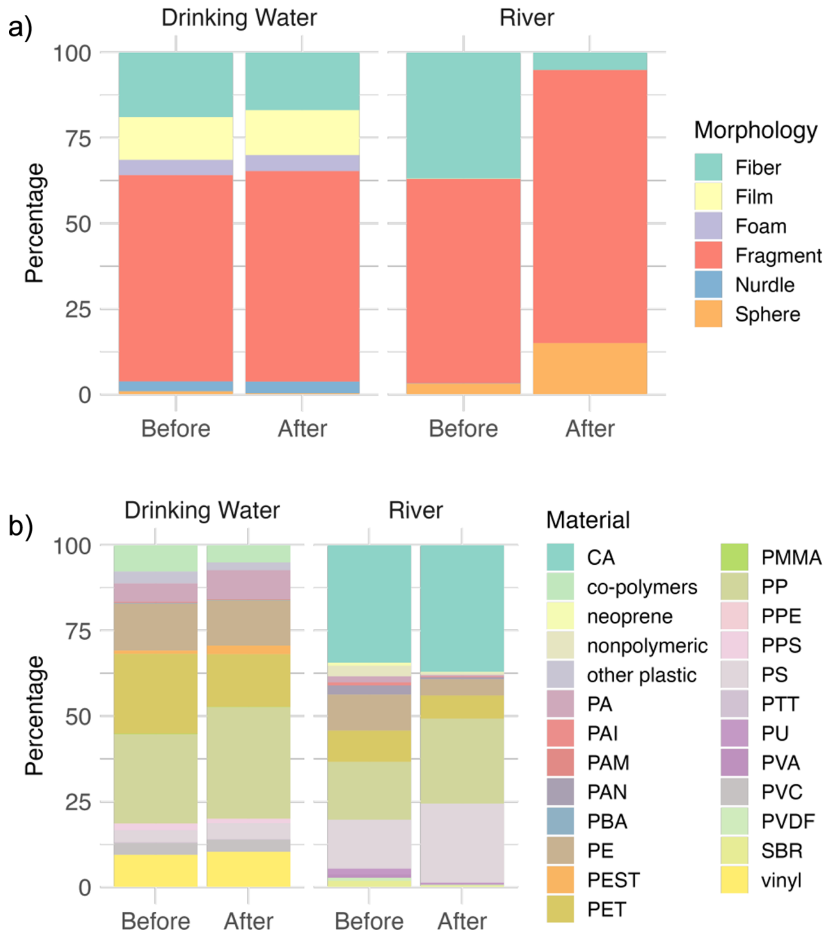


Figure 4. Changes in morphological (a) and polymeric (b) composition of microplastics by count in rivers and drinking water before and after size rescaling. Co-polymers, nonpolymeric, and other plastics have been lumped via the taxonomy for visualization. (CA: cellulose acetate, PA: polyamide, PAI: polyamide-imides, PAM: polyacrylamide, PAN: polyacrylonitrile, PBA: polybutylene terephthalate, PE: polyethylene , PEST: polyester, PET: polyethylene terephthalate, PMMA: poly(methyl methacrylate), PP: polypropylene, PPE: polyphenylene ether, PPS: polyphenylene sulfide, PS: polystyrene, PTT: polytrimethylene terephthalate, PU: polyurethane, PVA: polyvinyl acetate, PVC: polyvinyl chloride, PVDF: polyvinylidene fluoride, and SBR: styrene−butadiene rubber).

the prior paragraph and would, in fact lead us to believe riverine studies that utilize net-based sampling and undergo sieving would underreport fibers. Future work should investigate dominant sources of preferential recovery due to methodology based on morphology and fibers in particular.

Semantic proportions changing at different rates across size distributions in these examples indicate an interesting proposition: different PSDs may need to be developed for different morphologies or materials to traverse between semantic and quantitative data structures more accurately as well as to better characterize methodological limitations such as those mentioned above. Biases in modes of preferential environmental transport, limits of existing sampling and analytical methods, or reflections of differences in actual total abundances of various microplastic particle types across size distributions (possibly due to differing degradation pathways between materials or morphologies) may lead to a wide variety of PSDs when developed empirically rather than theoretically or with limited datasets. Additionally, some degree of particle level data reporting is recommended moving forward to promote transparency and enable direct comparisons between studies by allowing for more robust concentration rescaling. Though most groups do not record particle level data for an entire sample, typically, some subset of particles is taken to be characterized via spectroscopy, and even the sharing of morphological or size information on more robustly characterized subsets would strengthen existing empirical relationships. This type of reporting allows for the association of different particle characteristics (e.g., size distributions of individual morphologies can be analyzed). We believe this is an important topic to be investigated in future studies and that the use of MaTCH will help with identifying and pursuing these insights.

3.3. Methodological Limitations. The application of data-transforming operations posthoc is not without its limitations. While the application of different alpha values that reflect varying PSDs across environmental compartments is an improvement over assuming constant PSDs universally, these compartment-specific values are limited by the quality, amount, and specificity of the data with which empirical models were fit. More broadly, empirical models, such as those included in this study, are limited by the volume and quality of data used to develop them. Further work will be required to evaluate whether we observe these trends in systems not used in the development of the empirical models themselves and preferentially fit models to data from individual systems. Additionally, this model currently does not consider methodological variability, including from sampling apparatus or sample processing steps. Such variability should be explored to investigate if we observe significantly different trends as a function of study design that can then be used to further improve interstudy harmonization. One final limitation worth noting is the extrapolation of particle size ranges beyond what is studied. Though these empirically developed distributions are useful in improving harmonization across existing studies, it remains ideal for future studies to investigate the largest possible particle size range to avoid introducing unnecessary bias.

This study attempts to maximize transparency and capture the full range of uncertainties associated with such empirical models. However, we recognize the large range of uncertainty derived from data transformations such as those utilized here and encourage future studies to build off this framework to further reduce uncertainty. This tool is a step toward increased utilization of existing and new empirical harmonization methods to make improvements toward (1) creating more comparable results from misaligned monitoring studies and (2) informing future study designs from existing data on similar systems.

3.4. Implications and Next Steps. We believe our findings highlight a general need for more careful consideration of dataset characteristics when comparing between studies in the future. Moving forward, we propose building off existing non-semantic data harmonization frameworks with larger quantities of empirical data to better establish these fundamental properties of plastic characterization. This includes the relationship between morphology and *L:W:H*, as well as polymer–density relationships. The development of PSDs for nanoplastics could also be incorporated into the proposed framework to extend the reach of the potential particle size rescaling. Refining these distributions with respect to more detailed and higher volumes of data to describe more accurately what we observe in the environment is an important area of research moving forward. We hope this goal can be achieved by utilizing open-source, user-supported databases such as the one described in this study.

We believe these findings illustrate not just the need for data alignment in future studies and meta-analyses but also the interdependency of data harmonization techniques. Each step in the proposed algorithm is necessary to harmonize the quantity and characteristic makeup of microplastic samples. While a universal framework for microplastics data reporting does not currently exist, we will continue to develop and refine data harmonization algorithms to leverage valuable existing data with proper consideration of possible biases.

■ ASSOCIATED CONTENT

Data Availability Statement

Data and code used to generate figures, algorithm, and the Shiny App are openly available on Zenodo at: https://zenodo.org/records/13901058, GitHub at: https://github.com/hannahhapich/MaTCH/releases/tag/Publication.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c02406.

Additional equations, database diagram, additional case study, and meta-analysis study information (PDF)

User guide for accompanying R Shiny web tool (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Hannah Hapich – *University of California, Riverside, Riverside, California 92521, United States*; orcid.org/0000-0003-0000-6632; Email: hannahhapich@gmail.com

Authors

Win Cowger – *University of California, Riverside, Riverside, California 92521, United States; Moore Institute for Plastic Pollution Research, Long Beach, California 90803, United States*; orcid.org/0000-0001-9226-3104

Andrew B. Gray – *University of California, Riverside, Riverside, California 92521, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c02406

Funding

This article was developed based upon funding from the Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the U.S. Department of Energy, supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE-2236413, and in part by the Possibility Lab, the USDA National Institute of Food and Agriculture, Hatch

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the contributions of Samiksha Singh and Clare Murphy-Hagan during testing and development, as well as Haig Minasian, Anja Oca, and Holden Ford during data acquisition. This publication would not have been possible without their efforts.

■ REFERENCES

(1) Brander, S.; Hoh, E.; Unice, K.; Cook, A. M.; Holleman, R.; Rochman, C.; Thayer, J. *Microplastic Pollution in California: A Precautionary Framework and Scientific Guidance to Assess and Address Risk to the Marine Environment*; California Ocean Science Trust: Sacramento, 2021; .

(2) Jenkins, T.; Persaud, B. D.; Cowger, W.; Szigeti, K.; Roche, D. G.; Clary, E.; Slowinski, S.; Lei, B.; Abeynayaka, A.; Nyadjro, E. S.; Maes, T.; Thornton Hampton, L.; Bergmann, M.; Aherne, J.; Mason, S. A.; Honek, J. F.; Rezanezhad, F.; Lusher, A. L.; Booth, A. M.; Smith, R. D. L.; Van Cappellen, P. Current State of Microplastic Pollution Research Data: Trends in Availability and Sources of Open Data. *Front. Environ. Sci.* **2022**, *10*, 1–10.

(3) Microplastics: sources, effects and solutions https://www.europarl.europa.eu/news/en/headlines/society/20181116STO19217/microplastics-sources-effects-and-solutions (accessed Sep 10, 2023).

(4) Hartmann, N. B.; Hüffer, T.; Thompson, R. C.; Hassellöv, M.; Verschoor, A.; Daugaard, A. E.; Rist, S.; Karlsson, T.; Brennholt, N.; Cole, M.; Herrling, M. P.; Hess, M. C.; Ivleva, N. P.; Lusher, A. L.; Wagner, M. Are We Speaking the Same Language? Recommendations for a Definition and Categorization Framework for Plastic Debris. *Environ. Sci. Technol.* **2019**, *53* (3), 1039–1047.

(5) Cowger, W.; Booth, A. M.; Hamilton, B. M.; Thaysen, C.; Primpke, S.; Munno, K.; Lusher, A. L.; Dehaut, A.; Vaz, V. P.; Liboiron, M.; Devriese, L. I.; Hermabessiere, L.; Rochman, C.; Athey, S. N.; Lynch, J. M.; De Frond, H.; Gray, A.; Jones, O. A. H.; Brander, S.; Steele, C.; Moore, S.; Sanchez, A.; Nel, H. Reporting Guidelines to Increase the Reproducibility and Comparability of Research on Microplastics. *Appl. Spectrosc.* **2020**, *74* (9), 1066–1077.

(6) Metz, T.; Koch, M.; Lenz, P. Quantification of Microplastics: Which Parameters Are Essential for a Reliable Inter-Study Comparison? *Mar. Pollut. Bull.* **2020**, *157*, 111330.
(7) Yu, J. T.; Diamond, M. L.; Helm, P. A. A Fit-for-Purpose Categorization Scheme for Microplastic Morphologies. *Integr. Environ. Assess. Manag.* **2023**, *19* (2), 422−435.
(8) Molina Jack, M. E.; Chaves Montero, M. d. M.; Galgani, F.; Giorgetti, A.; Vinci, M.; Le Moigne, M.; Brosich, A. EMODnet marine litter data management at pan-European scale. *Ocean Coast. Manag.* **2019**, *181*, 104930.
(9) Čerkasova, N.; Enders, K.; Lenz, R.; Oberbeckmann, S.; Brandt, J.; Fischer, D.; Fischer, F.; Labrenz, M.; Schernewski, G. A Public Database for Microplastics in the Environment. *Microplastics* **2023**, *2* (1), 132−146.
(10) Bessa, F. *Harmonized Protocol for Monitoring Microplastics in Biota*, 2019..
(11) Miller, E.; Sedlak, M.; Lin, D.; Box, C.; Holleman, C.; Rochman, C. M.; Sutton, R. Recommended Best Practices for Collecting, Analyzing, and Reporting Microplastics in Environmental Media: Lessons Learned from Comprehensive Monitoring of San Francisco Bay. *J. Hazard. Mater.* **2021**, *409*, 124770.
(12) Hapich, H.; Cowger, W.; Gray, A.; Tangri, N.; Hale, T.; Magdy, A.; Vermilye, A.; Yu, W.; Ayres, D.; Moore, C.; Vermilye, J.; Singh, S.; Haiman, A. N. K.; Youngblood, K.; Kang, Y.; McCauley, M.; Lok, T.; Moore, S.; Baggs, E.; Lippiatt, S.; Kohler, P.; Conley, G.; Taing, J.; Mock, J. Trash Taxonomy Tool: Harmonizing Classification Systems Used to Describe Trash in Environments. *Microplastics and Nanoplastics* **2022**, *2* (1), 15.
(13) Waldschläger, K.; Brückner, M. Z.; Carney Almroth, B.; Hackney, C. R.; Adyel, T. M.; Alimi, O. S.; Belontz, S. L.; Cowger, W.; Doyle, D.; Gray, A.; Kane, I.; Kooi, M.; Kramer, M.; Lechthaler, S.; Michie, L.; Nordam, T.; Pohl, F.; Russell, C.; Thit, A.; Umar, W.; Valero, D.; Varrani, A.; Warrier, A. K.; Woodall, L. C.; Wu, N. Learning from Natural Sediments to Tackle Microplastics Challenges: A Multidisciplinary Perspective. *Earth-Science Rev.* **2022**, *228*, 104021.
(14) Bermúdez, J. R.; Swarzenski, P. W. A Microplastic Size Classification Scheme Aligned with Universal Plankton Survey Methods. *MethodsX* **2021**, *8*, 10−15.
(15) van Emmerik, T. Macroplastic Research in an Era of Microplastic. *Microplastics and Nanoplastics* **2021**, *1* (1), 4−5.
(16) Walker, T. R.; McGuinty, E.; Hickman, D. Marine Debris Database Development Using International Best Practices: A Case Study in Vietnam. *Mar. Pollut. Bull.* **2021**, *173*, 112948.
(17) Demchenko, Y.; Grosso, P.; De Laat, C.; Membrey, P. Addressing Big Data Issues in Scientific Data Infrastructure. In *2013 International Conference on Collaboration Technologies and Systems*; IEEE 2013..
(18) Olivetti, E. A.; Cole, J. M.; Kim, E.; Kononova, O.; Ceder, G.; Han, T. Y. J.; Hiszpanski, A. M. Data-Driven Materials Research Enabled by Natural Language Processing and Information Extraction. *Appl. Phys. Rev.* **2020**, *7* (4), 041317.
(19) Thessen, A. E.; Cui, H.; Mozzherin, D. Applications of Natural Language Processing in Biodiversity Science. *Adv. Bioinformatics* **2012**, *2012*, 1−17.
(20) Brants, T. *Application of Natural Language Processing for Information Retrieval*; CLIN, 2003; .
(21) Devlin, J.; Chang, M. W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL HLT 2019*; North American Chapter of the Association for Computational Linguistics, 2019.
(22) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *31st Conference on Neural Information Processing Systems*; NIPS, 2017.
(23) Qu, Y.; Liu, P.; Song, W.; Liu, L.; Cheng, M. A Text Generation and Prediction System: Pre-Training on New Corpora Using BERT and GPT-2. In *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*; IEEE, 2020..
(24) Mattsson, K.; Björkroth, F.; Karlsson, T.; Hassellöv, M. Nanofragmentation of Expanded Polystyrene Under Simulated Environmental Weathering (Thermooxidative Degradation and Hydrodynamic Turbulence). *Front. Mar. Sci.* **2021**, *7*, 1−9.
(25) Kaandorp, M. L. A.; Dijkstra, H. A.; van Sebille, E. Modelling Size Distributions of Marine Plastics under the Influence of Continuous Cascading Fragmentation. *Environ. Res. Lett.* **2021**, *16*, 054075.
(26) Kooi, M.; Koelmans, A. A. Simplifying Microplastic via Continuous Probability Distributions for Size, Shape,and Density. *Environ. Sci. Technol. Lett.* **2019**, *6* (9), 551−557.
(27) Koelmans, A. A.; Redondo-Hasselerharm, P. E.; Mohamed Nor, N. H.; Kooi, M. Solving the Nonalignment of Methods and Approaches Used in Microplastic Research to Consistently Characterize Risk. *Environ. Sci. Technol.* **2020**, *54* (19), 12307−12315.
(28) Greene, R.; Sanders, T.; Weng, L.; Neelakantan, A. OpenAI text-embedding-ada-002 https://openai.com/blog/new-and-improved-embedding-model.
(29) Cowger, W. chRoma R Package https://github.com/wincowgerDEV/chRoma.
(30) Cowger, W.; Gray, A.; Hapich, H.; Osei-Enin, J.; Olguin, S.; Huynh, B.; Nogi, H.; Singh, S.; Brownlee, S.; Fong, J.; Lok, T.; Singer, G.; Ajami, H. Litter Origins, Accumulation Rates, and Hierarchical Composition on Urban Roadsides of the Inland Empire, California. *Environ. Res. Lett.* **2022**, *17* (1), 015007.
(31) Besseling, E.; Redondo-Hasselerharm, P.; Foekema, E. M.; Koelmans, A. A. Quantifying Ecological Risks of Aquatic Micro- and Nanoplastic. *Crit. Rev. Environ. Sci. Technol.* **2019**, *49* (1), 32−80.
(32) *Mat Web*, https://www.matweb.com/.
(33) *PolymerDatabase*, https://polymerdatabase.com/.
(34) Kooi, M.; Primpke, S.; Mintenig, S. M.; Lorenz, C.; Gerdts, G.; Koelmans, A. A. Characterizing the Multidimensionality of Microplastics across Environmental Compartments. *Water Res.* **2021**, *202* (April), 117429.
(35) Mohamed Nor, N. H.; Kooi, M.; Diepens, N. J.; Koelmans, A. A. Lifetime Accumulation of Microplastic in Children and Adults. *Environ. Sci. Technol.* **2021**, *55* (8), 5084−5096.
(36) Chang, W.; Cheng, J.; Allaire, J. J.; Xie, Y.; McPherson, J. *Shiny: Web Application Framework for R*, 2023.
(37) Wickham, H.; Vaughan, D.; Girlich, M.; Ushey, K. tidyr: Tidy Messy Data https://cran.r-project.org/web/packages/tidyr/index.html.
(38) Waring, E.; Quinn, M.; McNamara, A.; Arino de la Rubia, E.; Zhu, H.; Lowndes, J.; Ellis, S.; McLeod, H.; Wickham, H.; Müller, K.; Kirkpatrick, C.; Brenstuhl, S.; Schratz, P.; Ibusett; Korpela, M.; Thompson, J.; McGehee, H.; Roepke, M.; Kennedy, P.; Possenriede, D.; Zimmermann, D.; Butts, K.; Torges, B.; Saporta, R.; Stewart, H. M. skimr.: Compact and Flexible Summaries of Data https://cran.r-project.org/web/packages/skimr/index.html.
(39) Müller, K.; Wickham, H.; François, R.; Bryan, J. *Tibble: Simple Data Frames*, 2023.
(40) Wickham, H.; François, R.; Henry, L.; Müller, K. *Dplyr: A Grammar of Data Manipulation*, 2023.
(41) Barrett, T.; Dowle, M.; Srinivasan, A.; Gorecki, J.; Chirico, M.; Hocking, T.; Stetsenko, P.; Short, T.; Lianoglou, S.; Antonyan, E.; Bonsch, M.; Parsonage, H.; Ritchie, S.; Ren, K.; Tan, X.; Saporta, R.; Seiskari, O.; Dong, X.; Lang, M.; Iwasaki, W.; Wenchel, S.; Broman, K.; Schmidt, T.; Arenburg, D.; Smith, E.; Cocquemas, F.; Gomez, M.; Chataignon, P.; Blaser, N.; Selivanov, D.; Riabushenko, A.; Lee, C.; Groves, D.; Possenriede, D.; Parages, F.; Toth, D.; Yaramaz-David, M.; Perumal, A.; Sams, J.; Morgan, M.; Quinn, M.; Storey, R.; Saraswat, M.; Jacob, M.; Schubmehl, M.; Vaughan, D.; Silvestri, L.; Hester, J.; Damico, A.; Freundt, S.; Simons, D.; Sales de Andrade, E.; Miller, C.; Meldgaard, J. P.; Tlapak, V.; Ushey, K.; Eddelbuettel, D.; Schwendinger, B.; Fischetti, T.; Shilon, O.; Khotilovich, V.; Wickham, H.; Becker, B.; Haynes, K.; Kamgang, B. C.; Delmarcell, O.; O'Brien, J.; Mezquita, D. de.; Czekanski, M. *Data.Table: Extension of data.Framè*. 2024.
(42) Hyde, R.; Hammill, C.; Munoz, F.; Wamser, M.; Formont, P.; Russel, K.; Ross, N.; Garmonsway, D.; Glur, C. data.tree: General Purpose Hierarchical Data Structure. https://cran.r-project.org/web/packages/data.tree/index.html.

(43) Sievert, C.; Parmer, C.; Hocking, T.; Chamberlain, S.; Ram, K.; Corvellec, M.; Despouy, P.; Bruggermann, S. *Plotly: Create Interactive Web Graphics via "Plotly.js."*, 2024.
(44) Granjon, D.; Studio, A.; Chang, W. *Bs4Dash: A "Bootstrap 4" Version of "Shinydashboard*, 2024.
(45) Bivand, R.; Denney, B.; Dunlap, R.; Hernangomez, D.; Ono, H.; Parry, J.; Stigler, M. *ClassInt.: Choose Univariate Class Intervals*, 2023.
(46) Leeper, T. J.; Carl, B.; Martin, A.; Thompson, M.; Hunt, T.; Akins, S.; Nguyen, B.; Onkelinx, T.; Degtiarov, A.; Aggarwal, D.; Columbus, A.; Urbanek, S. *aws.s3: "AWS S3" Client Package*.
(47) Eddelbuettel, D. *Digest: Create Compact Hash Digests of R Objects*.
(48) Xie, Y.; Cheng, J.; Tan, X.; Allaire, J.; Girlich, M.; Ellis, G. F.; Rauh, J.; Reavis, B.; Gersen, L.; Szopka, B.; Pickering, A.; Holmes, W.; Marttila, M.; Quintero, A.; Laurent, S. D. T.: A Wrapper of the JavaScript Library "DataTables" https://cran.r-project.org/web/packages/DT/index.html.
(49) Allen, J.; Bozhanov, I.; Schaffer, M.; Danker, T.; Bell, M.; Gatscha, S.; Thaler, T. *ShinyTree: JsTree Bindings for Shiny*, 2023.
(50) Mason-Thom, C. *Shinyhelper: Easily Add Markdown Help Files to "Shiny" App Elements*, 2019.
(51) Perrier, V.; Meyer, F.; Granjon, D.; Fellows, I.; Davis, W.; Matthews, S. shinyWidgets: Custom Inputs Widgets for Shiny https://cran.r-project.org/web/packages/shinyWidgets/index.html.
(52) Järlskog, I.; Jaramillo-Vogel, D.; Rausch, J.; Gustafsson, M.; Strömvall, A. M.; Andersson-Sköld, Y. Concentrations of Tire Wear Microplastics and Other Traffic-Derived Non-Exhaust Particles in the Road Environment. *Environ. Int.* **2022**, *170*, 107618.
(53) Feng, Q.; An, C.; Chen, Z.; Lee, K.; Wang, Z. Identification of the Driving Factors of Microplastic Load and Morphology in Estuaries for Improving Monitoring and Management Strategies: A Global Meta-Analysis. *Environ. Pollut.* **2023**, *333*, 122014.
(54) Finnegan, A. M. D.; Süsserott, R.; Gabbott, S. E.; Gouramanis, C. Man-Made Natural and Regenerated Cellulosic Fibres Greatly Outnumber Microplastic Fibres in the Atmosphere. *Environ. Pollut.* **2022**, *310*, 119808.
(55) Kukkola, A.; Krause, S.; Lynch, I.; Sambrook Smith, G. H.; Nel, H. Nano and Microplastic Interactions with Freshwater Biota – Current Knowledge, Challenges and Future Solutions. *Environ. Int.* **2021**, *152*, 106504.

https://doi.org/10.1021/acs.est.4c02406
Environ. Sci. Technol. 2024, 58, 20502−20512