




One4All: An Open Source Portal to Validate and Share Microplastics Data and Beyond


Hannah Sherrod ¹, Nicholas Leong ¹, Hannah Hapich ², Fabian Gomez¹, Shelly Moore¹, Ben Maurer³, Scott Coffin⁴, Leah Thornton Hampton⁵, Tony Hale⁶, Richard Nelson⁷, Clare Murphy-Hagan², Oluniyi O. Fadare ², Anna Kukkola⁸, Hsuan-Cheng Lu⁹, Laura Markley¹⁰, and Win Cowger ^{1,2}

1 Moore Institute for Plastic Pollution Research, USA 2 University of California, Riverside, Department of Environmental Sciences, USA 3 National Renewable Energy Laboratory, USA 4 Office of Environmental Health Hazard Assessment, USA 5 Southern California Coastal Water Research Project Authority, USA 6 San Francisco Estuary Institute, USA 7 California State Water Resources Control Board, USA 8 University of Birmingham, Life and Environmental Sciences, UK 9 Australian Rivers Institute, School of Environment and Science, Griffith University, AUS 10 Syracuse University, USA

DOI: [10.21105/joss.06715](https://doi.org/10.21105/joss.06715)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Kanishka B. Narayan 

Reviewers:

- [@hahsan1](#)
- [@stephpenn1](#)

Submitted: 15 April 2024

Published: 16 July 2024

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Microplastics are a diverse suite of contaminants ([Rochman et al., 2019](#)) requiring a variety of data types to capture particle characteristics and study methodologies ([Division of Drinking Water, State Water Resources Control Board, State of California, 2022](#); [Jenkins et al., 2022](#)). Data sharing is critical to advance science and policy ([Coffin, 2023](#)). The One4All portal was created to harmonize and share structured and unstructured microplastic data (and beyond) through validation. Validated data can be automatically uploaded to the following cloud services: Amazon S3, CKAN, and/or MongoDB. One4All is both a graphical user interface (GUI) and an R package. The tools are general-purpose and can be used for applications outside of microplastics. This manuscript provides information about the usage, workflow, and configuration of One4All, emphasizing its adaptability for diverse purposes.

Statement of need

Scientific advancement depends on the ability to reuse data and reproduce results, as highlighted by Peng ([2011](#)). For example, the National Center for Biotechnology Information (NCBI)'s gene databank contains over 3 billion records and is considered a critical component of biological research. Ample funding and attention have been directed toward studying the human genome for decades, while plastic pollution, particularly microplastics, has only recently begun to receive this attention. The challenge in effectively sharing data, lies in the varied structuring and formatting of data across different research areas, especially in emerging fields such as plastic pollution. This inconsistency slows progress and results in inefficient use of resources to make disparate datasets compatible for analysis and decision making. Plastic pollution monitoring uses a variety of data collection methods, environmental matrices, and analytical approaches, leading to a lack of standardized data formats ([Dissanayake et al., 2022](#); [Fox et al., 2024](#); [Hapich et al., 2022](#); [Lu et al., 2021](#)). Although there are efforts towards creating databases, such as Čerkasova et al.'s public database for environmental microplastics ([Čerkasova et al., 2023](#)), widespread adoption of standardized data formats remains elusive.

Several key aspects are required to ensure widespread adoption of a unified data reporting structure: FAIR data principles—making data Findable, Accessible, Interoperable, and Reusable; ease of use; and field-based standards and regulatory guidance. By integrating FAIR principles,

we propose the development of One4All, a user-friendly, open-source portal designed to validate, share, and make microplastics monitoring data accessible to any researcher in a structured format via a web browser. One4All can also be used locally, independent of a web browser, through its R package. This platform aims to support the transparent, reproducible, and flexible exchange of data across scientific communities, thereby fostering collaboration among researchers, regulators, and policymakers (Coffin et al., 2021; Cowger, Gomez, et al., 2023; Jenkins et al., 2022; Simon-Sánchez et al., 2022). Although One4All is the general framework, the [Microplastics Data Portal](#) is a specific implementation designed for microplastics data (Cowger et al., 2020). This data portal was created based on open-source resources, supporting the implementation of the California Senate Bill (SB) No. 1422 (2017-2018), which mandates the sharing of drinking water microplastics data (Coffin, 2023). Additionally, the Microplastics Data Portal has been legislatively established and will be mandated to be used by public water systems and accredited microplastic labs (Division of Drinking Water, State Water Resources Control Board, State of California, 2022). While One4All was initially developed to serve as a resource and regulatory reporting standard for the USA State of California's monitoring campaign for microplastics, it can also be utilized by any researcher, enabling greater harmonization amongst researchers across the globe. One4All represents a pivotal step towards standardizing data in rapidly evolving research areas, and by being responsive to user needs, enabling more effective and collaborative scientific inquiry.

Demo

A full tutorial video is available on [YouTube](#). For additional instructions, access the pkgdown and vignettes here: (<https://moore-institute-4-plastic-pollution-res.github.io/One4All/>).

Method

Workflow Overview

Consisting of an R package and R shiny application, this tool was designed to enhance data validation and management processes by employing a set of rules from a 'CSV' or 'Excel' file to validate a dataset. Users can work in the Validator app or the One4All package. The functionality in the application is based on the One4All package (Figure 1).

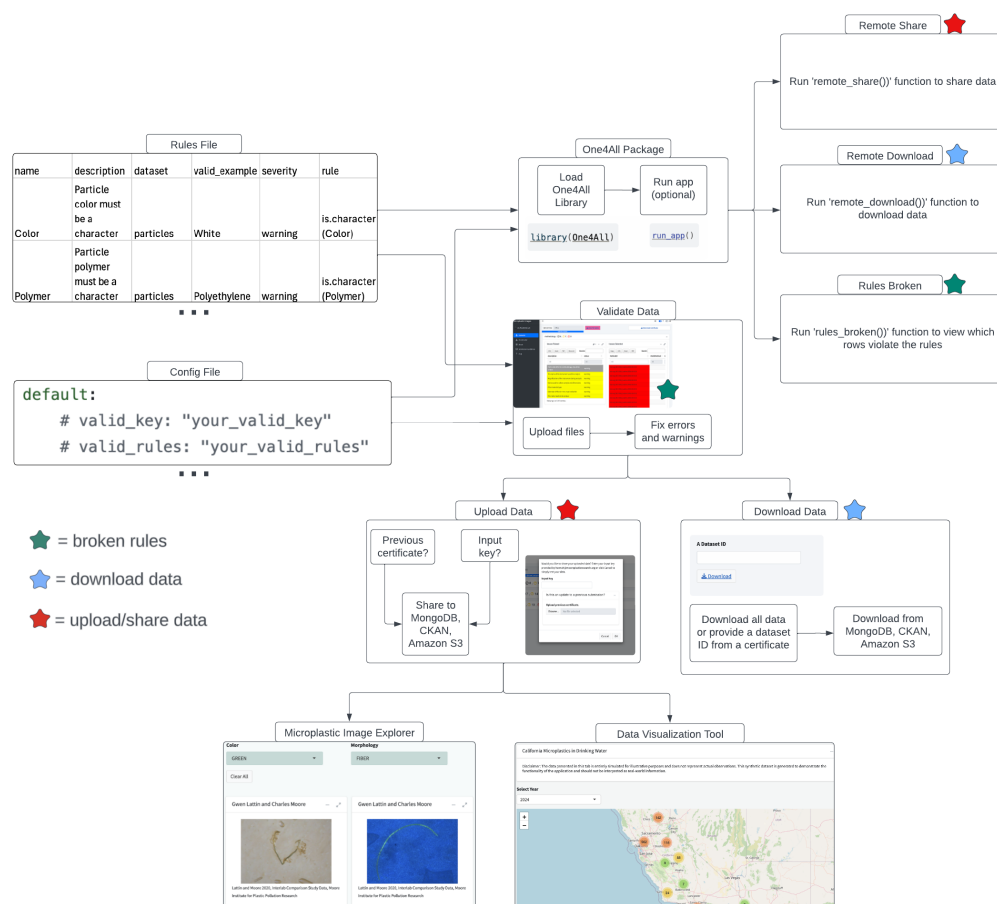


Figure 1: Workflow of the One4All portal. Navigate through the One4All package and the Validator app, based on user preference: (<https://lucid.app/documents/view/f33867e8-5822-4cef-b62a-3a6c2e293b4e>).

The rules sheet and configuration file power the application and package functions to validate, share, and download data. The implemented set of rules comes from a [rules sheet](#), which is a 'CSV' file containing the following fields: name, description, dataset, valid example, severity, and rule. Users can adapt the rules sheet to use the tool for other purposes. Data validation rules are specified as R functions, referring to the column names in the uploaded data and returning TRUE for valid or FALSE invalid. The dataset is an optional field when the data package is separated into multiple sheets (Excel) or CSV files to specify which rule pertains to which sheet. Specify the rules sheet to use in the 'rules_to_use' configuration field, or hash out the field and upload the rules sheet directly within the app (Figure 2).

name	description	dataset	valid_example	severity	rule
Latitude	Latitude of sample in decimal degrees	samples	33.746974	error	<code>in_range(as.numeric(Latitude), min = -90, max = 90)</code>
Longitude	Longitude of sample in decimal degrees	samples	-118.115372	error	<code>in_range(as.numeric(Longitude), min = -180, max = 180)</code>
SampleVolume	Volume of sample (L), must be greater than 0	samples	100	error	<code>SampleVolume > 0</code>

Figure 2: A screenshot showing a subset of the example rules sheet applied in the One4All portal, which comprises a total of 81 rules. The Validator app provides users with sample rules, a valid data example, and an invalid data example, all of which are available for users to download and adapt to suit their own purposes.

Structure a [configuration file](#) using the template from the One4All portal. Replace the hashed

placeholders with your information to determine fields such as where to share and download the data. An additional layer of security can be implemented by creating a valid key which will require users to provide an input key when sharing or downloading validated data. Users can request an input key from the maintainer listed in the configuration file under 'contact'.

default:

```
# valid_key_share: "your_valid_key_share"
# valid_key_download: "your_valid_key_download"
# valid_rules: "your_valid_rules"
# ckan_key: "your_ckan_key"
# ckan_url: "your_ckan_url"
# ckan_package: "your_ckan_package"
# ckan_url_to_send: "your_ckan_url_to_send"
portal_name: "One4All"
portal_funder_name: !expr shiny::tags$a(href = "https://possibilitylab.berkeley.edu/
portal_funder_link: "https://pbs.twimg.com/profile_images/1540022027288924160/bn2627
# s3_username: "your_s3_username"
# s3_key_id: "your_s3_key_id"
# s3_secret_key: "your_s3_secret_key"
# s3_console_link: "your_s3_console_link"
# s3_region: "your_s3_region"
# s3_bucket: "your_s3_bucket"
# mongo_key: "your_mongo_key"
# mongo_collection: "your_mongo_collection"
# apiKey: "your_mongo_apikey"
rules_to_use: "www/rules_dw_acc.csv"
rules_example: "www/rules_dw_acc.csv"
valid_data_example: !expr c("www/data.xlsx", "www/extra_files.zip")
invalid_data_example: !expr c("www/data_example_errors.xlsx", "www/example_files_err
twitter: "https://twitter.com/Win_OpenData"
github: "https://github.com/Moore-Institute-4-Plastic-Pollution-Res/One4All/tree/ma
license: "https://opensource.org/license/mit"
contact: "win@mooreplasticresearch.org"
dev: TRUE
tutorial: "https://www.youtube.com/embed/LMpf5-K_tYQ"
overview: "https://www.youtube.com/embed/GKsoNega7CY"
google_form_link: "https://forms.gle/kMEjpcuSRREJeYR8"
```

How to use the Validator app

Users will first validate their data by uploading a 'CSV' or 'Excel' file, and a corresponding zip folder consisting of the unstructured data (Figure 3).

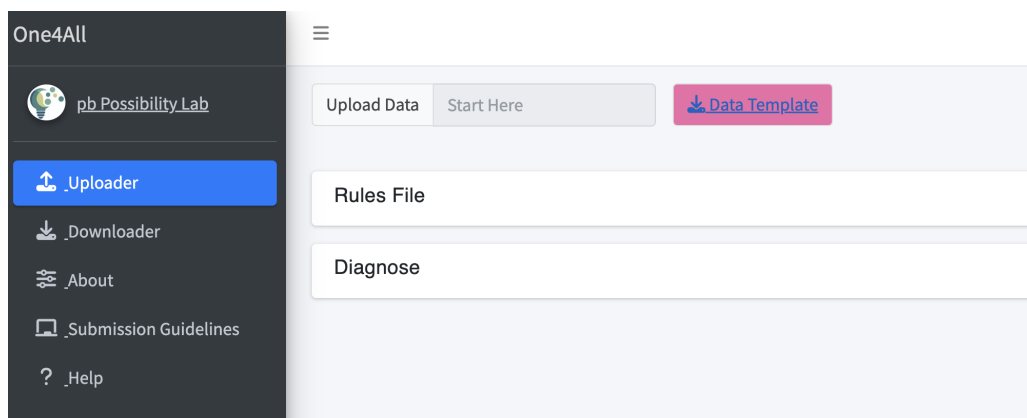


Figure 3: Upload data by selecting a 'CSV' or 'Excel' file and a corresponding zip folder.

If there are any invalid data, errors and warnings will be highlighted on the following screen. The 'Issues Raised' box identifies which fields resulted in an error or warning, along with a description of the error or warning. The 'Issues Selected' box identifies which cells were invalid based on the selection in 'Issues Raised' (Figure 4).

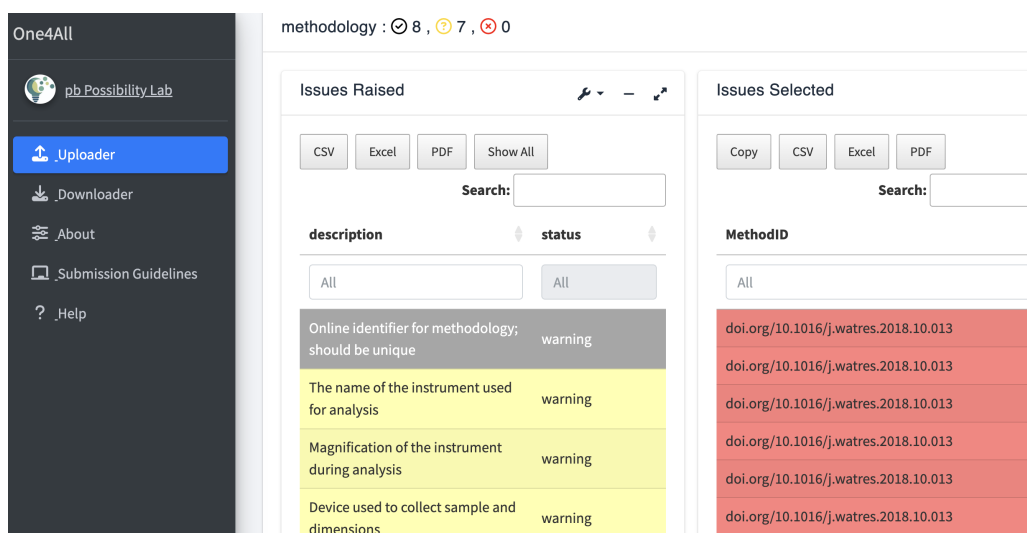


Figure 4: View any errors and warnings identified here. Errors must be corrected to validate the data, but warnings are allowed. Reupload the files after fixing the errors.

Users can upload their validated data to the previously mentioned cloud services, based on the credentials provided in the configuration file. To share data, provide an input key from the configuration file, if required. Data will not be allowed to be shared unless no errors are present so if there are errors, data must be corrected and reuploaded. Additionally, a certificate from a previous submission can be uploaded if updating a previous submission (Figure 5).

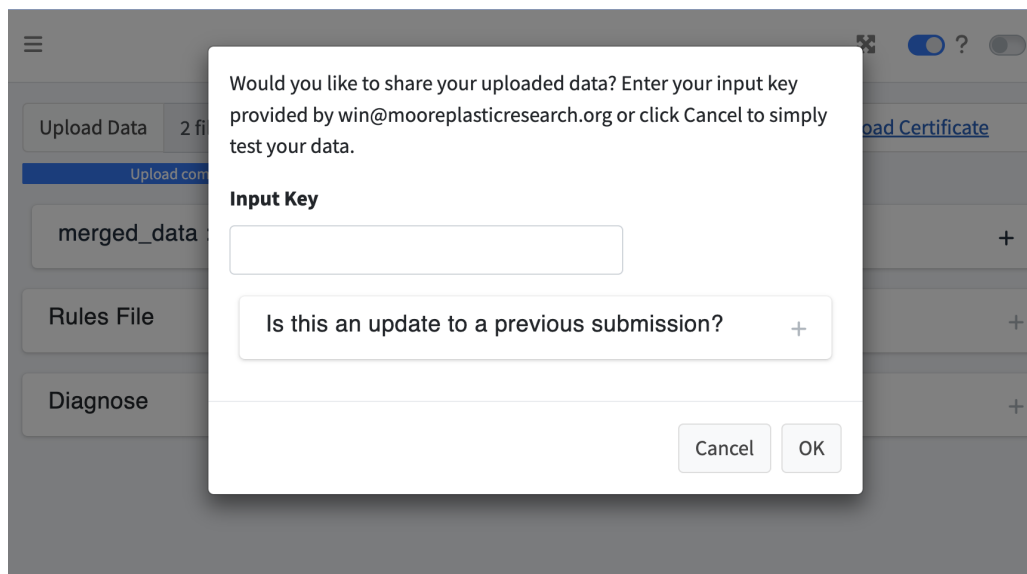


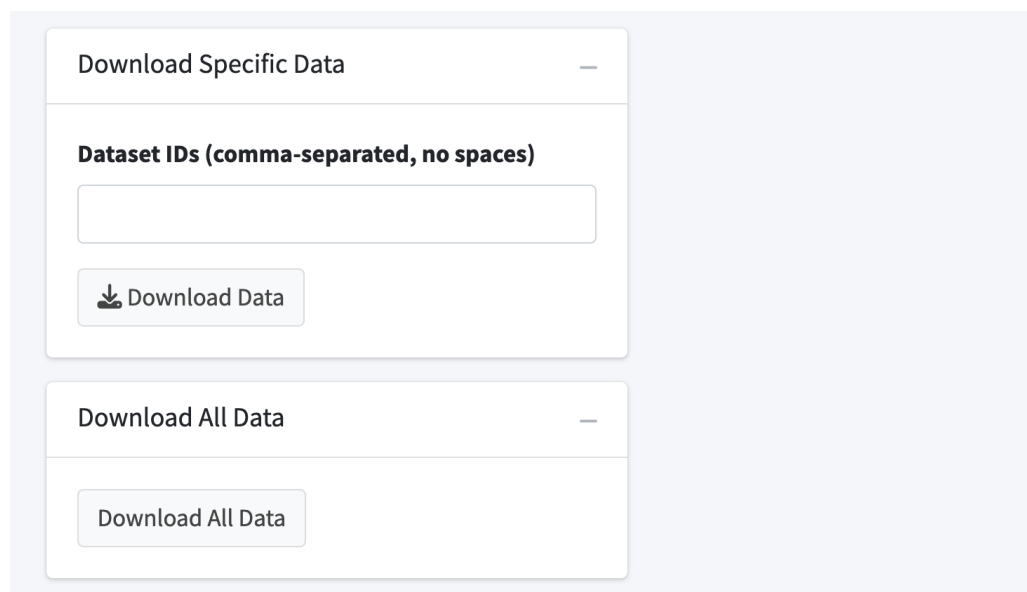
Figure 5: To share the validated data, provide an input key, if required, and click 'OK'. If this is an update to a previous submission, upload a previous certificate (see Figure 6).

Once successfully shared, a confirmation screen will be generated and the data will be uploaded to the appropriate cloud services. Users can download a certificate at the top right corner to view the credentials of their submission (Figure 6).

time	data	rules	package_version	web_hash
6/27/24 17:11	3af803128a0	534d5adff27	1.1.3	01146a2249db8

Figure 6: A screenshot displaying an example of a downloaded certificate, including the downloaded time, the data ID necessary to download this dataset (see Figure 7), the rules ID identifying the set of rules used for this dataset, the package version, and the web hash. The data ID and rules ID are also represented as hashes.

Users also have the option to download any previously uploaded data from the cloud services. Users can download one dataset at a time, providing the dataset ID from a downloaded certificate, or choose to download all data (Figure 7).



The screenshot shows a web interface with two main sections. The top section is titled 'Download Specific Data' and contains a text input field labeled 'Dataset IDs (comma-separated, no spaces)' and a button with a download icon and the text 'Download Data'. The bottom section is titled 'Download All Data' and contains a button with the text 'Download All Data'.

Figure 7: To download a dataset, provide the dataset ID from a downloaded certificate or download all data.

Other Tools

Additionally, the One4All portal includes two tools, the [Data Visualization Tool](#) and the [Microplastic Image Explorer](#), both sourced from the [Microplastics Data Portal](#) to increase discoverability and adhere to FAIR data principles.

Data Visualization Tool

The Data Visualization Tool allows users to analyze microplastics data through visual representations. The data presented in the first three tabs are real data sourced from the [California Open Data Portal](#), displaying the concentration of microplastics, globally, through maps, plots, and tables. Users can explore the relationships between various microplastic characteristics, including morphology, color, and polymer type. The fourth tab uses synthetic data to demonstrate the potential functionality of the tool in a world with abundant and standardized drinking water microplastic data, displaying the yearly average microplastic concentrations, treatment levels, polymer and width distributions (Figure 8).

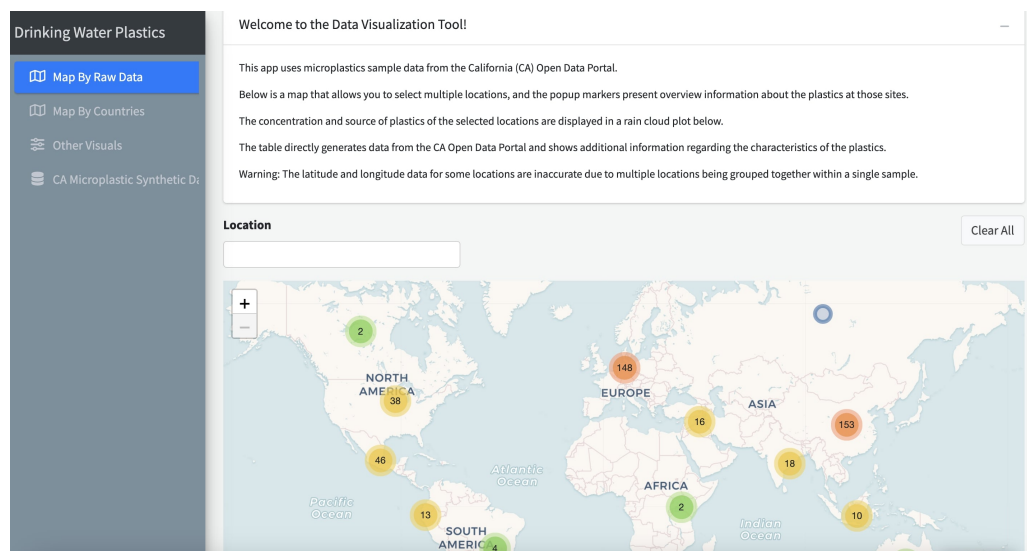


Figure 8: Analyze microplastics data through various visual representations.

Microplastic Image Explorer

The Microplastic Image Explorer enables users to query microplastic images from our database of approximately 8,000 images. This tool was designed to improve the use of visual microscopy in microplastic identification where users can apply filters, including citation, color, morphology, and polymer type to query specific images (Figure 9).

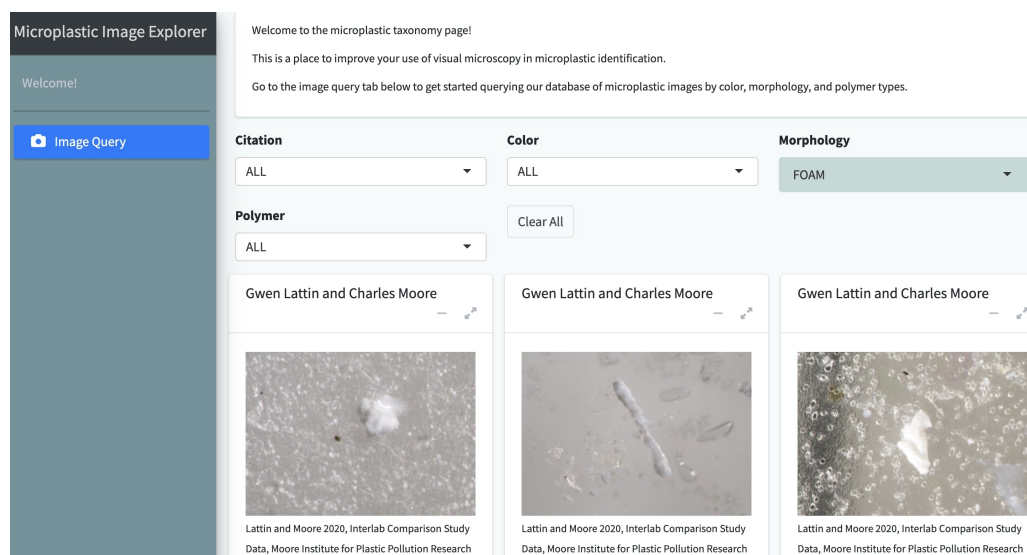


Figure 9: Query microplastic images from the database by applying filters, including citation, color, morphology, and polymer.

How to use the One4All package

The features in the applications are driven using the One4All package functions. The overarching goal of the R package is to allow functions to be reused in contexts beyond the imagined scope of One4All. In most cases, the applications will be more user-friendly than the code.

First, load the One4All library, then optionally run the app using the 'run_app()' function.


```
library(One4All)
run_app()
```

Validate data using the 'validate_data' function. Replace 'files_data', 'data_names', and 'file_rules' with your values.

```
result_valid <- validate_data(files_data = valid_example,
                              data_names = c("methodology", "particles", "samples"),
                              file_rules = test_rules)
```

Identify errors and warnings using the 'rules_broken' function. Replace the 'results' and 'show_decision' parameters with your values. This function filters the validation results to show only the broken rules, optionally including successful decisions as well.

```
broken_rules <- rules_broken(results, show_decision)
```

Share your validated data to the cloud services using the 'remote_share' function. Ensure that your data are validated before sharing your data. Replace the pertaining placeholders with your information.

```
shared_data <- remote_share(validation = result_valid,
                             data_formatted = result_valid$data_formatted,
                             files = test_file,
                             verified = "your_verified_key",
                             valid_key_share = "your_valid_key_share",
                             valid_rules = digest::digest(test_rules),
                             ckan_url = "https://example.com",
                             ckan_key = "your_ckan_key",
                             ckan_package = "your_ckan_package",
                             url_to_send = "https://your-url-to-send.com",
                             rules = test_rules,
                             results = valid_example$results,
                             s3_key_id = "your_s3_key_id",
                             s3_secret_key = "your_s3_secret_key",
                             s3_region = "your_s3_region",
                             s3_bucket = "your_s3_bucket",
                             mongo_key = "your_mongo_key",
                             mongo_collection = "your_mongo_collection",
                             old_cert = NULL)
```

Download data from the cloud services using the 'remote_download' function. Replace the pertaining placeholders with your information.

```
downloaded_data <- remote_download(hashled_data = "example_hash",
                                    ckan_url = "https://example.com",
                                    ckan_key = "your_ckan_key",
                                    ckan_package = "your_ckan_package",
                                    s3_key_id = "your_s3_key_id",
                                    s3_secret_key = "your_s3_secret_key",
                                    s3_region = "your_s3_region",
                                    s3_bucket = "your_s3_bucket",
                                    mongo_key = "mongo_key",
                                    mongo_collection = "mongo_collection")
```

Limitations

The main limitation is finding a balance between expanding the variety of rules to broaden the tool's usability and maintaining efficient collaboration among researchers. The current rules implemented in the One4All portal encompass a wide variety of microplastics data, but those

rules are limited to microplastics. To accept a wider variety of data, or to use the tool for an entirely different purpose, additional rules sheets will need to be created with variables about the other fields. However, the more variables added to a rules sheet, the more complex the overall QA/QC compliance pathway, increasing the chances of data entry errors.

Availability and Documentation

One4All is published on CRAN and was created in R(4.2.1) (R Core Team, 2022) using the following libraries: shiny (Chang et al., 2023), dplyr (Wickham, François, et al., 2023), validate (van der Loo & de Jonge, 2021), digest (Antoine Lucas et al., 2023), data.table (Barrett et al., 2023), ckanr (Chamberlain et al., 2023), openxlsx (Schauberger & Walker, 2023), lexicon (Rinker, 2018), readr (Wickham, Hester, et al., 2023), readxl (Wickham & Bryan, 2023), tibble (Müller & Wickham, 2023), aws.s3 (Leeper, 2020), rlang (Henry & Wickham, 2023), jsonlite (Ooms, 2014), mongolite (Ooms, 2014), and httr (Wickham, 2023). The microplastic version of the Validator app is hosted on the web at openanalysis.org/one4all/. The source code is available on GitHub with an MIT License. The Moore Institute for Plastic Pollution Research maintains the package. The source code can be run offline on any machine that can install the One4All package.

Future Goals

The One4All portal is intended to be useful for users to validate and share their data. Our goal is to promote open-source resources for the public and update the software with other resources that have yet to be implemented. We plan to integrate Open Specy, developed by Cowger et al. (2021), for users to validate their spectra submissions before submitting them to the Open Specy reference library. Additionally, once the Validator tool receives user data, we plan to incorporate this data into the Data Visualization Tool and the Microplastic Image Explorer. For the Data Visualization Tool, we intend to utilize an API key generated from the cloud services within One4All to integrate shared data near real-time, enabling the tool to visually represent the shared data through maps, plots, and tables. Our objective for the Microplastic Image Explorer is to validate datasets containing microplastic images before integrating them into the tool's database. Similarly, we can collect data for Trash AI, a GUI that identifies individual items of trash within images, by validating datasets containing images of trash before uploading them (Cowger, Hollingsworth, et al., 2023). Furthermore, we aim to expand our current set of rules by incorporating AI strategies such as ChatGPT (OpenAI, 2024) to automate the generation of rule templates for macroplastics or other pollutants.

Acknowledgments

We acknowledge and greatly appreciate the financial support from the National Renewable Energy Laboratory and the Possibility Lab. The Moore Institute for Plastic Pollution Research led the development of the software tool and drafting of this manuscript. The One4All portal was inspired by the data.validator tool created by Dubel et al. (2024). We acknowledge the work and the input of the San Francisco Estuary Institute and the California State Water Resources Control Board. We would like to thank Scott Coffin, Tony Hale, Diana Lin, Gemma Shusterman, Rebecca Sutton, Adam Wong, Richard Nelson, Leah Thornton Hampton, Libby Heeren, and Gabriel Daïess. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any California State agency.

References

- Antoine Lucas, D. E. with contributions by, Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M., Lewis, B., Stokely, M., Muehleisen, H., Murdoch, D., Hester, J., Wu, W., Kou, Q., Onkelinx, T., Lang, M., Simko, V., Hornik, K., Neal, R., Bell, K., de Queljoe, M., ... Attali., D. (2023). *Digest: Create Compact Hash Digests of R Objects*. <https://CRAN.R-project.org/package=digest>
- Barrett, T., Dowle, M., & Srinivasan, A. (2023). *Data.table: Extension of 'data.frame'*. <https://CRAN.R-project.org/package=data.table>
- Čerkasova, N., Enders, K., Lenz, R., Oberbeckmann, S., Brandt, J., Fischer, D., Fischer, F., Labrenz, M., & Schernewski, G. (2023). A Public Database for Microplastics in the Environment. *Microplastics*, 2(1), 132–146. <https://doi.org/10.3390/microplastics2010010>
- Chamberlain, S., Costigan, I., Wu, W., Mayer, F., Gelfand, S., & Alves, F. (2023). *Ckanr: Client for the Comprehensive Knowledge Archive Network ('CKAN') API*. <https://CRAN.R-project.org/package=ckanr>
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). *Shiny: Web Application Framework for R*. <https://CRAN.R-project.org/package=shiny>
- Coffin, S. (2023). The emergence of microplastics: charting the path from research to regulations. *Environmental Science: Advances*, 2(3), 356–367. <https://doi.org/10.1039/D2VA00275B>
- Coffin, S., Wyer, H., & Leapman, J. C. (2021). Addressing the environmental and health impacts of microplastics requires open collaboration between diverse sectors. *PLOS Biology*, 19(3). <https://doi.org/10.1371/journal.pbio.3000932>
- Cowger, W., Booth, A. M., Hamilton, B. M., Thaysen, C., Primpke, S., Munno, K., Lusher, A. L., Dehaut, A., Vaz, V. P., Liboiron, M., I., D. L., Hermabessiere, L., Rochman, C., Athey, S. N., Lynch, J. M., De Frond, H., Gray, A., Jones, O. A. H., Brander, S., ... Nel, H. (2020). Reporting Guidelines to Increase the Reproducibility and Comparability of Research on Microplastics. *Applied Spectroscopy*, 74(9), 1066–1077. <https://doi.org/10.1177/0003702820930292>
- Cowger, W., Gomez, I., Martinez-Rubin, N., Moriarity, A., Harwell, T., & Anich, L. (2023). Community Science-Informed Local Policy: A Case Study in Pinole Creek Litter Assessment. *Community Science*, 2(1). <https://doi.org/10.1029/2022CSJ000017>
- Cowger, W., Hollingsworth, S., Fey, D., Norris, M. C., Walter, Y., Kerge, K., Haamer, K., Durante, G., & Hernandez, B. (2023). Trash AI: A Web GUI for Serverless Computer Vision Analysis of Images of Trash. *Journal of Open Source Software*, 8(89), 5136. <https://doi.org/10.21105/joss.05136>
- Cowger, W., Steinmetz, Z., Gray, A., Munno, K., Lynch, J., Hapich, H., Primpke, S., DeFrond, H., Rochman, C., & Herodotou, O. (2021). Microplastic Spectral Classification Needs an Open Source Community: Open Specy to the Rescue! *Analytical Chemistry*, 93, 7543–7548. <https://doi.org/10.1021/acs.analchem.1c00123>
- Dissanayake, P. D., Kim, S., Sarkar, B., Oleszczuk, P., Sang, M. K., Haque, M. N., Ahn, J. H., Bank, M. S., & Ok, Y. S. (2022). Effects of microplastics on the terrestrial environment: A critical review. *Environmental Research*, 209. <https://doi.org/10.1016/j.envres.2022.112734>
- Division of Drinking Water, State Water Resources Control Board, State of California. (2022). *Policy Handbook Establishing a Standard Method of Testing and Reporting of Microplastics in Drinking Water*.

- Dubel, M., Przytuła, P., Nowicki, J., & Igras, K. (2024). *Data.validator: Automatic Data Validation and Reporting*. <https://appilon.github.io/data.validator/>
- Fox, S., Stefánsson, H., Peternell, M., Zlotskiy, E., Jón Ásbjörnsson, E., Sturkell, E., Wanner, P., & Konrad-Schmolke, M. (2024). Physical characteristics of microplastic particles and potential for global atmospheric transport: A meta-analysis. *Environmental Pollution*, 342. <https://doi.org/10.1016/j.envpol.2023.122938>
- 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., ... Mock, J. (2022). Trash Taxonomy Tool: harmonizing classification systems used to describe trash in environments. *Microplastics and Nanoplastics*, 2, 15. <https://doi.org/10.1186/s43591-022-00035-1>
- Henry, L., & Wickham, H. (2023). *Rlang: Functions for Base Types and Core R and 'Tidyverse' Features*. <https://CRAN.R-project.org/package=rang>
- Jenkins, T., Persaud, B. D., Cowger, W., Szigeti, K., Roche, D. G., Clary, E., Slowinski, S., Lei, B., Abeynayaka, A., Nyadjro, E., Maes, T., Thornton Hampton, L., Bergmann, M., Aherne, J., Mason, S. A., Honek, J. F., Rezanezhad, F., Lusher, A. L., Booth, A. M., ... Van Cappellen, P. (2022). Current State of Microplastic Pollution Research Data: Trends in Availability and Sources of Open Data. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.912107>
- Leeper, T. J. (2020). *Aws.s3: AWS S3 Client Package*.
- Lu, H., Ziajahromi, S., Neale, P. A., & Leusch, F. D. L. (2021). A systematic review of freshwater microplastics in water and sediments: Recommendations for harmonisation to enhance future study comparisons. *Science of The Total Environment*, 781. <https://doi.org/10.1016/j.scitotenv.2021.146693>
- Müller, K., & Wickham, H. (2023). *Tibble: Simple Data Frames*. <https://CRAN.R-project.org/package=tibble>
- Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *arXiv:1403.2805 [Stat.CO]*. <https://arxiv.org/abs/1403.2805>
- OpenAI. (2024). *ChatGPT: Large language model*. <https://chat.openai.com/chat>
- Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rinker, T. W. (2018). *lexicon: Lexicon Data*. <http://github.com/trinker/lexicon>
- Rochman, C. M., Brookson, C., Bikker, J., Djuric, N., Earn, A., Bucci, K., Athey, S., Huntington, A., McIlwraith, H., Munno, K., De Frond, H., Kolomijeca, A., Erdle, L., Grbic, J., Bayoumi, M., Borrelle, S. B., Wu, T., Santoro, S., Werbowski, L. M., ... Hung, C. (2019). Rethinking microplastics as a diverse contaminant suite. *Environmental Toxicology and Chemistry*, 38(4), 703–711. <https://doi.org/10.1002/etc.4371>
- Schauberger, P., & Walker, A. (2023). *Openxlsx: Read, Write and Edit xlsx Files*. <https://CRAN.R-project.org/package=openxlsx>
- Simon-Sánchez, L., Grelaud, M., Franci, M., & Ziveri, P. (2022). Are research methods shaping our understanding of microplastic pollution? A literature review on the seawater and sediment bodies of the Mediterranean Sea. *Environmental Pollution*, 292. <https://doi.org/10.1016/j.envpol.2021.118275>
- van der Loo, M. P. J., & de Jonge, E. (2021). Data Validation Infrastructure for R. *Journal of Statistical Software*, 97(10), 1–31. <https://doi.org/10.18637/jss.v097.i10>

- Wickham, H. (2023). *Httr: Tools for Working with URLs and HTTP*. <https://CRAN.R-project.org/package=httr>
- Wickham, H., & Bryan, J. (2023). *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>