Aquatic Food Composition Database, Version 5

2026

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Summary Information

How to cite: Zamborain-Mason, Jessica*; Marwaha, Nisha*; Koehn, J. Zachary*; Viana, Daniel; Hazen, Lucie; Zhang, Angela; Cornell, Christina; Free, Christopher M.; Fairfield, William D.; Vaitla, Bapu; DeSisto, Camille; Kelahan, Heather; Fiorella, Kathryn J.; Kjellevold, Marian; Thilsted, Shakuntala H., Tapera, Tinashe M., Golden, Christopher D. 2026, "Aquatic Food Composition Database", doi:10.7910/DVN/KI0NYM Harvard Dataverse, v5.0, AFCD.CSV. (*indicates joint first co-authorship)

License: CC0 1.0

Temporal coverage: Systematic search from 1990-2020, with various older sources from 1942

Spatial coverage: 111 countries

Nutritional coverage: 718 nutrients

Aquatic food taxa coverage: 3,558 taxa

Disclaimer: The data have been compiled from original publications. The Aquatic Food Composition Database is not the author or owner of the data. The database is provided in good faith and to the best of the authors' knowledge. Please use the provided citation when using the database. This does not preclude citing individual studies when needed. If you have data that you think should be included in the database or notice any potential mistakes, please contact us.

Keywords: nutrition; blue foods; food security; macronutrients; micronutrients; minerals; vitamins; fatty acids; amino acids

Acknowledgments: We acknowledge financial support from the National Science Foundation Human Network Data Science grant (2121239; awarded to CDG), the John and Katie Hansen Family Foundation (awarded to CDG), the Harvard Data Science Initiative with funding support from Amazon Web Services (awarded to CDG), and the Stanford Center for Ocean Solutions supporting JZK through grants from the Moore Foundation, Walton Foundation, and the Stanford Realizing Environmental Innovation Program.

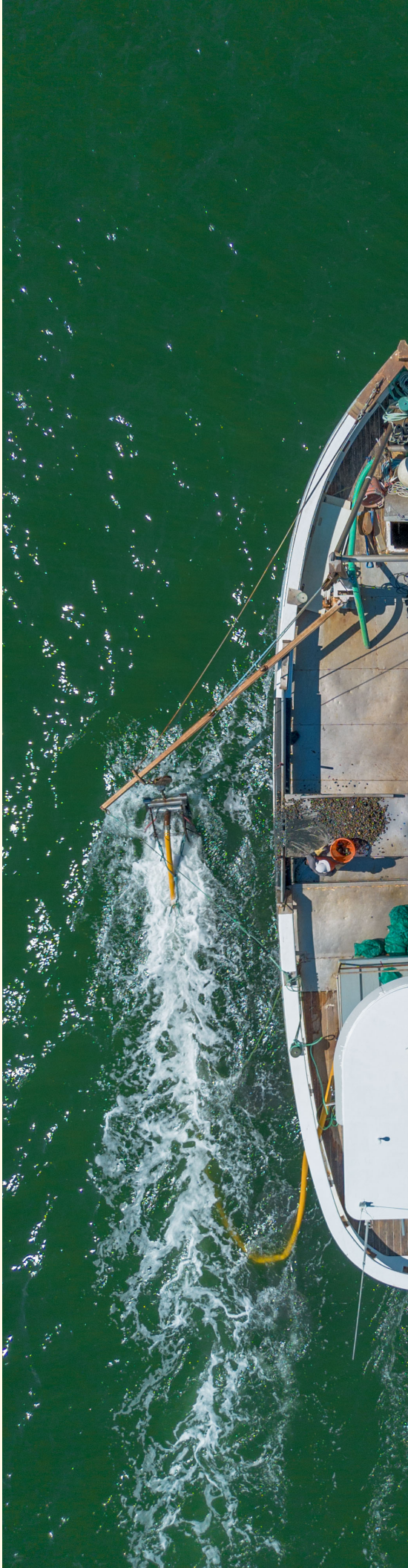
Database Overview

Welcome to the Aquatic Food Composition Database (AFCD)!

This database assembles existing nutrient composition data for aquatic food species, comprising hundreds of nutrients including minerals, vitamins, and fatty acids from 3,558 aquatic food taxa. These data originate from national food composition tables (FCT) and international datasets from FAO that were in machine-readable formats, as well as from peer reviewed journal articles extracted from a systematic literature review of Web of Science in 2020. A second review focused on species in the Western and Central Pacific in 2023. Sources were excluded from AFCD if the scientific name of the organism was not mentioned, they only assessed fish oil or processed seafood products containing other ingredients, or they were part of an experimental trial. The full list of included sources can be found in supplementary file afcd_references.xlsx. A Dataverse deposit — comprised of the most up-to-date dataset, documentation, and a verified code snapshot — is available on the Harvard DataVerse. New and updated data are periodically added in this repository. Therefore, summary and regional values may vary, and species-nutrient observations may differ in the future. However, a code release for this version is available in the DataVerse deposit.

Version 1-4 of AFCD can be found at Golden, C.D., Koehn, J.Z., Vaitla, B., DeSisto, C., Kelahan, H., Manning, K., Fiorella, K.J., Kjellevold, M., Thilsted, S.H., 2021. Aquatic Food Composition Database. Harvard Dataverse, v4. https://doi.org/10.7910/DVN/KI0NYM.

Version 5 of AFCD (AFCD v5) focused on streamlining the data entry workflow, expanding sample metadata, general data quality control, including additional data, and improving documentation for transparency.



Database Update

Streamlining data entry workflow: In previous AFCD versions, an R script joined six spreadsheets with nutritional data grouped by nutrient category by the metadata associated with each peer-reviewed study. The script standardized nutrient units and variable names from the merge and converted relative nutrient values to absolute values (e.g., the value of a fatty acid relative to total fat content in the sample). In AFCD v4, the spreadsheets have been compiled so all study metadata and nutritional information are in one file. Relevant scripts can be found in the DataVerse deposit, as well as on GitHub (see DataVerse for GitHub Link). It is important to note that the separate spreadsheets of previous versions were compiled by matching all available sample metadata. However, if samples with identical metadata were entered in different orders between spreadsheets (e.g., samples not reported in a consistent order in source, differing number of samples reported per nutrient category in source), data between nutrient variables may not correspond to the same analytical sample.

Sample metadata: Certain existing variables have been modified and new variables added to better capture spatial, temporal and food attributes of each sample in AFCD which are included in the list below.

- Sample_id: All samples from journal articles now have an ID that differentiates them from other samples with the same scientific name with the same Study ID. A unique ID for each sample in AFCD can be created with the Study ID, Sample ID, and scientific name.
- Country_origin_study: The alpha-3 ISO country code(s) of the study origin, i.e. country associated with the corresponding author.
- Country_origin_sample: The alpha-3 ISO country code(s) of the sample origin, i.e. harvest location. This may differ from where the sample was procured.
- Country_origin_sample-detail: Further detail about the sample origin, where available.
- Sample_year: The year(s) of sample collection. If the sample was collected during multiple years, the years are separated by an underscore.
- Sample_month: The month(s) of sample collection. If the sample was collected during multiple months, the months are separated by an underscore.
- Food_processing: All food processing, i.e. mechanical or chemical processes for product transformation or value addition.
- Other_ingredients: Names of other ingredients in the sample, separated by commas.
- Sample_preparation: Sample preparation before nutrient analyses. This variable previously held both food processing and sample preparation data.
- Relative_weight: Basis used for data reporting, i.e. wet weight or dry weight.

Data quality control: A significant amount of effort was focused on improving data quality.

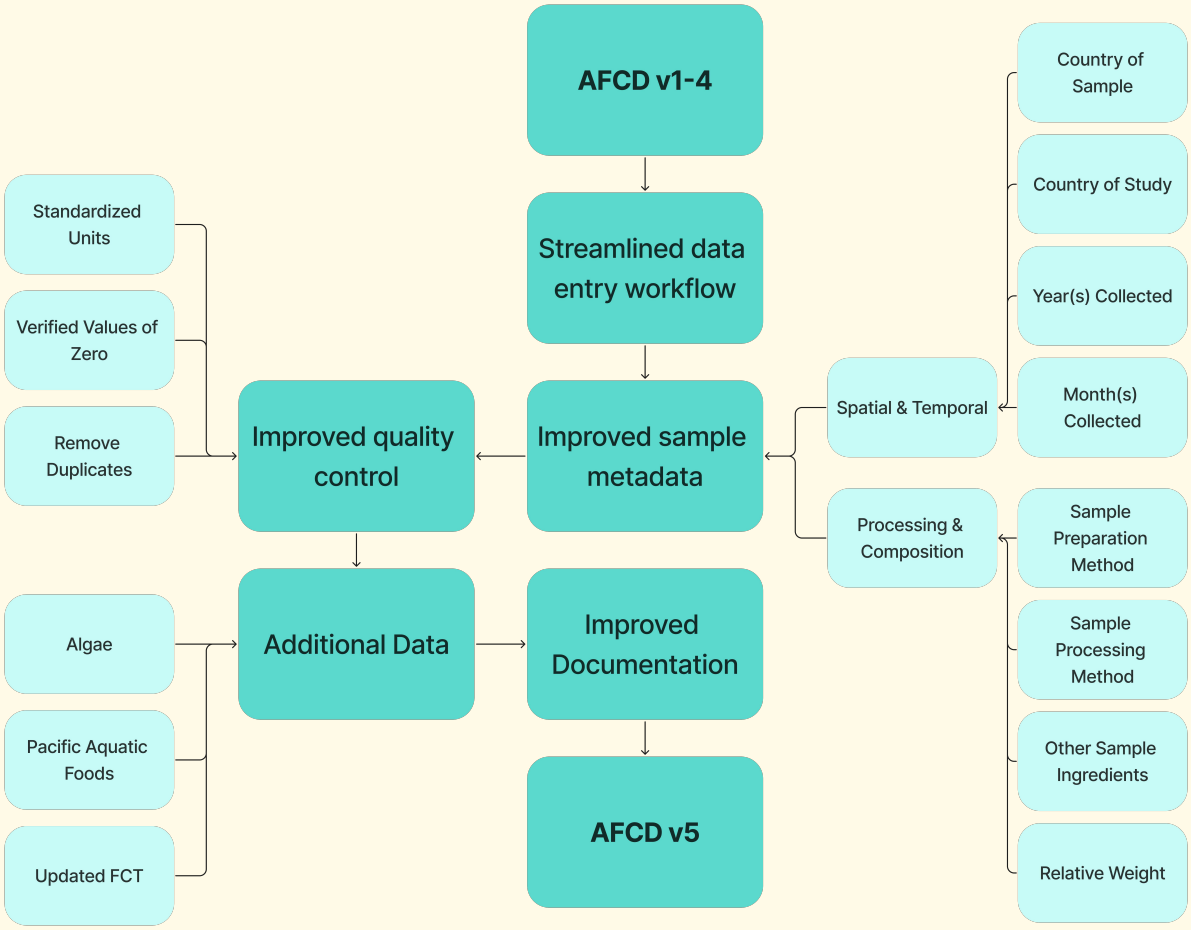
- Unit inconsistencies between FCTs, FAO datasets, and journal articles were corrected.
- Zeros were checked for accuracy – it is important to note that zeros can represent true zeros or measurements below the detectable limit.
- Replicated FCTs that were already compiled as part of a previous project (i.e., GENuS-Fish) and included in previous AFCD versions were removed.

Database Update

Additional data: Specific projects developed throughout the course of this update. As a consequence, the data below was added or updated following the same workflow (i.e., adding the relevant sample metadata and performing the same data quality control).

- Algae data: A targeted effort to add nutrient data for algae was performed on Web of Science in 2023. We include algae data from the Japanese 2015 FCT (v7) and 26 additional journal article studies identified in a targeted search. The specific search terms are detailed in the "Search Terms" document.
- Pacific aquatic foods data: A targeted effort to add nutrient data for Pacific reef species from journal articles was completed in 2023, following the same protocol used to collect other peer reviewed data. The review paralleled the 2020 Web of Science search terms and exclusion criteria, but also included geographic and species-specific terms to target Pacific-specific nutritional observations. These search terms, along with the search terms for the original peer review data, can be found in the "Search Terms" document.
- Updated FCTs: Eight of the 19 currently included FCTs and one of the two FAO datasets have updated versions available (AUS, BGD, CAN, NOR, NZL, KOR, THA, USA; FAO/INFOODS BioFoodComp). Given time constraints, only the Korean FCT was updated with the latest available version from 2024. Updates to other FCTs will be added to a subsequent version of AFCD, along with new FCTs from countries not currently included.

Improving documentation: General database documentation was clarified and expanded to increase transparency around data selection, data sources, and data processing.



Additional Documentation

The following documents are also available for this database

- Variable Key
- References
- Search Terms

Photo Credit:
Title Page - Bijay Chaurasia Photography, Wikimedia Commons, CC BY-SA 4.0
Page 1 - Tomascatelazo, Wikimedia Commons, CC BY-SA 3.0. Image has been cropped.
Page 2 - Hayden Soloviev, Wikimedia Commons, CC BY-SA 4.0. Image has been cropped.
Page 5 - W. Carter, Wikimedia Commons, CC BY-SA 1.0. Image has been cropped.