Segovia de la Revilla, L, Codd, T., Joy, E.J.M., Mlambo, L., Grande, F., Rittenschober, D., Moltedo, A., Holmes, B.A. and Ander, E.L. 2024. An open science framework and tools to create reproducible food composition data for use in nutrition. *Journal of Food Composition and Analysis - JFCA.* 137, p. 106894. https://doi.org/10.1016/j.jfca.2024.106894

Journal of Food Composition and Analysis 137 (2025) 106894

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

An open science framework and tools to create reproducible food composition data for use in nutrition

Lucia Segovia de la Revilla [a,b,*], Thomas Codd [b], Edward J.M. Joy [a,e], Liberty Mlambo [b], Fernanda Grande [c], Doris Rittenschober [c], Ana Moltedo [d], Bridget A. Holmes [c], E. Louise Ander [b,f]

[a] Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom
[b] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, United Kingdom
[c] Food and Nutrition Division, Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, Rome 00153, Italy
[d] Statistics Division, Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, Rome 00153, Italy
[e] Sustainable Soils and Crops, Rothamsted Research, West Common, Harpenden, AL5 2JQ, United Kingdom
[f] Centre for Environmental Geochemistry, British Geological Survey, Keyworth, Nottingham NG12 5GG, United Kingdom

ARTICLE INFO

Keywords:
Food composition data
INFOODS
Findability
Accessibility
Interoperability
Reproducibility
Fish and aquatic products
Metadata
Micronutrients

ABSTRACT

Food composition tables and databases (FCTs) and Nutrient Conversion Tables (NCTs) are essential for nutrition research. Compiling a new NCT requires multiple FCTs, usually with incompatible formats. FCT cleaning and standardisation is rarely reproducible and requires significant resources. Our aim was to develop a framework and tools for compilation and reporting of reproducible FCTs/NCTs, through expanding the fish and other aquatic products in the global NCT for the Food and Agriculture Organization of the United Nations (FAO) Supply and Utilization Accounts.

FAO/ International Network of Food Data Systems (INFOODS) guidelines, and open science tools were used for processing. New R functions and scripts were developed to: import and standardise 12 FCTs; re-calculate food components; perform quality checks; and format outputs (e.g., spreadsheets).

This resulted in the expansion of the global NCT, providing information on 32 food components for 95 fish and other aquatic products. The workflow takes 160 seconds to run. The scripts are publicly available in GitHub, with a manual, and can be used or adapted.

These open science tools provide a novel resource to create, update and expand FCTs/NCTs in a reproducible, reusable, efficient, and transparent manner, for use in nutrition research.

1. Introduction

Food Composition Tables and Databases (FCTs) contain data on the energy and nutrient composition of food items, and sometimes other information such as recipes or edible portions. FCTs have multiple uses in public health nutrition, including their integration with information on food consumption to estimate intake of energy and nutrients, for example to conduct population nutrition surveillance or to explore the relationship between intake of nutrients and health outcomes (Durazzo and Lucarini, 2022; Traka et al., 2020). Most FCTs are compiled by national authorities or research groups, and they may be tailored to specific contexts or studies. Thus, these datasets are found in various formats, with differences in the list of food items and components (i.e. nutrients) reported, inconsistent use of data conventions (e.g., nutrient definition, analytical methods, mode of expression, units), and with variable quality and completeness of data, metadata and documentation (Clancy et al., 2015; Pennington et al., 2007; Segovia de la Revilla et al., 2023).

National/regional FCTs are often combined to generate study or research specific food composition datasets, for example those used for Household Consumption and Expenditure Surveys: referred to as Nutrient Conversion Tables (NCTs) from this point onwards. Compiling

* Correspondence to: London School of Hygiene and Tropical Medicine, Department of Population Health, Faculty of Epidemiology & Population Health, Keppel St, London WC1E 7HT, United Kingdom.

E-mail addresses: Lucia.Segovia-De-La-Revilla@lshtm.ac.uk (L. Segovia de la Revilla), Thomas.Codd@nottingham.ac.uk (T. Codd), Edward.joy@lshtm.ac.uk (E.J.M. Joy), Liberty.Mlambo@nottingham.ac.uk (L. Mlambo), Fernanda.Grande@fao.org (F. Grande), Doris.Rittenschober@fao.org (D. Rittenschober), Ana.Moltedo@fao.org (A. Moltedo), Bridget.Holmes@fao.org (B.A. Holmes), Louise.Ander@nottingham.ac.uk (E.L. Ander).

https://doi.org/10.1016/j.jfca.2024.106894
Received 5 August 2024; Received in revised form 11 October 2024; Accepted 23 October 2024
Available online 28 October 2024

locally relevant FCTs/NCTs requires multiple data inputs, typically including non-standardised and non-interoperable FCTs resulting in manual data processing, which is rarely reproducible, prone to human error and involves significant human and financial resources (Charrondiere et al., 2023; Durazzo et al., 2022; Zeb et al., 2021). The often inscrutable decisions made (e.g., converting trace to zero, imputing values, etc.) when generating both FCTs and NCTs may decrease the reliability and accuracy of the food composition values reported. Furthermore, many countries do not have a national FCT and/or local analytical food composition data. In such cases data are borrowed from other countries, reducing further the relevance and accuracy of the data (Bruyn et al., 2016; Ene-Obong et al., 2019; Segovia de la Revilla et al., 2023). Consequently, when used along with other datasets (e.g., food consumption) may inadvertently propagate these inconsistencies/inaccuracies into the estimates of nutrient intakes and risk of (in)adequacy (Coates et al., 2017; Joy and Kumssa, 2022; Kapsokefalou et al., 2019).

Increased data comparability and interchangeability can result from the use of standardised methods, including data collation, analysis, food description and formats, to compile FCTs (Durazzo et al., 2022; Ene-Obong et al., 2019; Ispirova et al., 2017; Kapsokefalou et al., 2019). For instance, the use of International Network of Food Data Systems (INFOODS) food component identifiers (also known as tagnames) provides unambiguous component identification facilitating data interchange (FAO/INFOODS, 2012a; Klensin et al., 1989). More recently, the development of the Compositional Dietary Nutrition Ontology (CDNO) serves a similar purpose (Andrés-Hernández et al., 2022). Likewise, data harmonisation is essential for reconciling diverse data sources and to allow for compatibility and comparability among them (Zeb et al., 2021). For example, a standardised and comprehensive food description is needed for accurate food matching when linking two different lists of foods by their descriptions (Moshfegh et al., 2022). A number of standards are currently available for use as food classification and description systems, for instance FoodEx2 classification and description system (European Food Safety Authority, 2015), LanguaL (Møller and Ireland, 2018), or using food ontologies, such as FoodOn (Dooley et al., 2018; Ispirova et al., 2017).

Sector specific classification systems are also available, such as the Aquatic Sciences and Fisheries Information System (ASFIS) (FAO, 2022) which is curated and maintained by FAO and provides consistent classification systems for fishery and aquaculture products, including grouping and identification codes (e.g., International Standard Statistical Classification for Aquatic Animals and Plants (ISSCAAP), taxonomic and 3-alpha group) and taxonomic information (e.g., scientific name, species family, etc.). Harmonised food description and classification facilitate the incorporation of different data sources, including data from different countries, while enabling the aggregation of similar foods, such as the 95 fish and other aquatic products in the FAO Supply Utilisation Accounts (SUAs) (referred in this study as "SUA items") (FAO, 2023, 2021; Grande et al., 2024; Rittenschober et al., 2016). Hence, the use of community standards that align with the Findability, Accessibility, Interoperability and Reusability (FAIR) principles would contribute to increased transparency in the nutrition field (Chan et al., 2021).

The aim of this study was to develop an open science framework and tools to compile transparent and reproducible FCTs and NCTs. The adoption of open science approaches including publishing the code would aid other researchers and food/nutrition composition compilers to apply this framework, similar to recent efforts to harmonise food consumption data processing (Luo et al., 2021). The objectives were to combine FCTs from multiple formats by providing standardising and harmonising scripts; reduce costs of updating and generating new FCTs/NCTs; increase reproducibility, re-usability, efficiency and transparency of FCTs/NCTs. Finally, these were applied to: a) the validation of the framework by replicating the compilation of the fish and aquatic products subset of the global NCT for FAO SUAs developed by FAO's Food and Nutrition Division (Grande et al., 2024); b) the extension of the nutrients included for the fish and other aquatic products to showcase its implementation (Fig. 1).

2. Methods

The main framework steps were: identifying and obtaining the food composition data (i.e., FCTs), standardising the food composition data into a common data library, harmonisation of the data (including food matching), checking quality and completeness of the food composition data, and compiling the FCT/NCT and relevant documentation (Fig. 2). The framework was based on the recommendations outlined in the FAO/INFOODS Guidelines (FAO/INFOODS, 2012b, 2012a) and in the Micronutrient Action Policy Support (MAPS) project scripting approaches. The steps outlined here were developed in RStudio version 2023.6.0.421 powered by the R software version 4.4.1 (Posit team, 2023; R Core Team, 2023).

2.1. *Identifying and obtaining the food composition tables and databases*

The selection criteria for FCTs have been fully documented elsewhere (Grande et al., 2024). In brief, FCTs of high quality were selected based on scoring undertaken using the "FAO/INFOODS Evaluation framework to assess the quality of published food composition tables and databases" (Charrondiere et al., 2023). Then, those FCTs that passed the screening were checked for relevancy for the study/context, data availability and missing values (e.g., relevant foods and nutrients are reported), and data quality and reporting (e.g., method of chemical analysis, complete metadata). After the FCTs were reviewed and selected, access to the data was obtained, when possible, in a text or tabular format (.csv, MS Excel, MS Access,.txt), and imported into R/RStudio (Table 1).

2.2. *Importing the data*

FCTs were found in a variety of data structures/formats which influenced the steps and complexity of importing the original FCT files into R(Studio) (see Table 1). Thus, individual R scripts were developed to perform the importing and subsequent FCT-specific steps. In addition, within our framework, a template is provided which includes guidance on several steps (e.g., choosing the import function according to the data format), and operations that are commonly required for cleaning and standardising FCTs. This is designed to facilitate script re-use for future incorporation of new datasets.

2.3. *Data cleaning and standardisation*

After importing the FCTs, the most frequent cleaning and standardisation steps, are described here. These are needed for compilation of the different FCTs into one food composition data library.

2.3.1. *Formatting FCTs into a tabular format*

The first step is getting the FCT into a tabular format, which helps with further processing, as functions can be applied across multiple foods and/or food components. Some of the formatting tasks include: removing empty rows, translocating and relocating columns, and/or merging multiple data tables (e.g., when nutrients were separated in different spreadsheets).

2.3.2. *Renaming variables*

Renaming variables is important for compilation as variables reporting values of the same food component should have exactly the same name. Our framework uses the FAO/INFOODS food component identifiers (tagnames), denoted by "< >" (e.g. <ENERC>) in this document (FAO/INFOODS, 2012b; Klensin et al., 1989) to precisely identify all of the food components, while we propose other common names for the remaining variables (e.g., food identifier (*fcd_id*), food

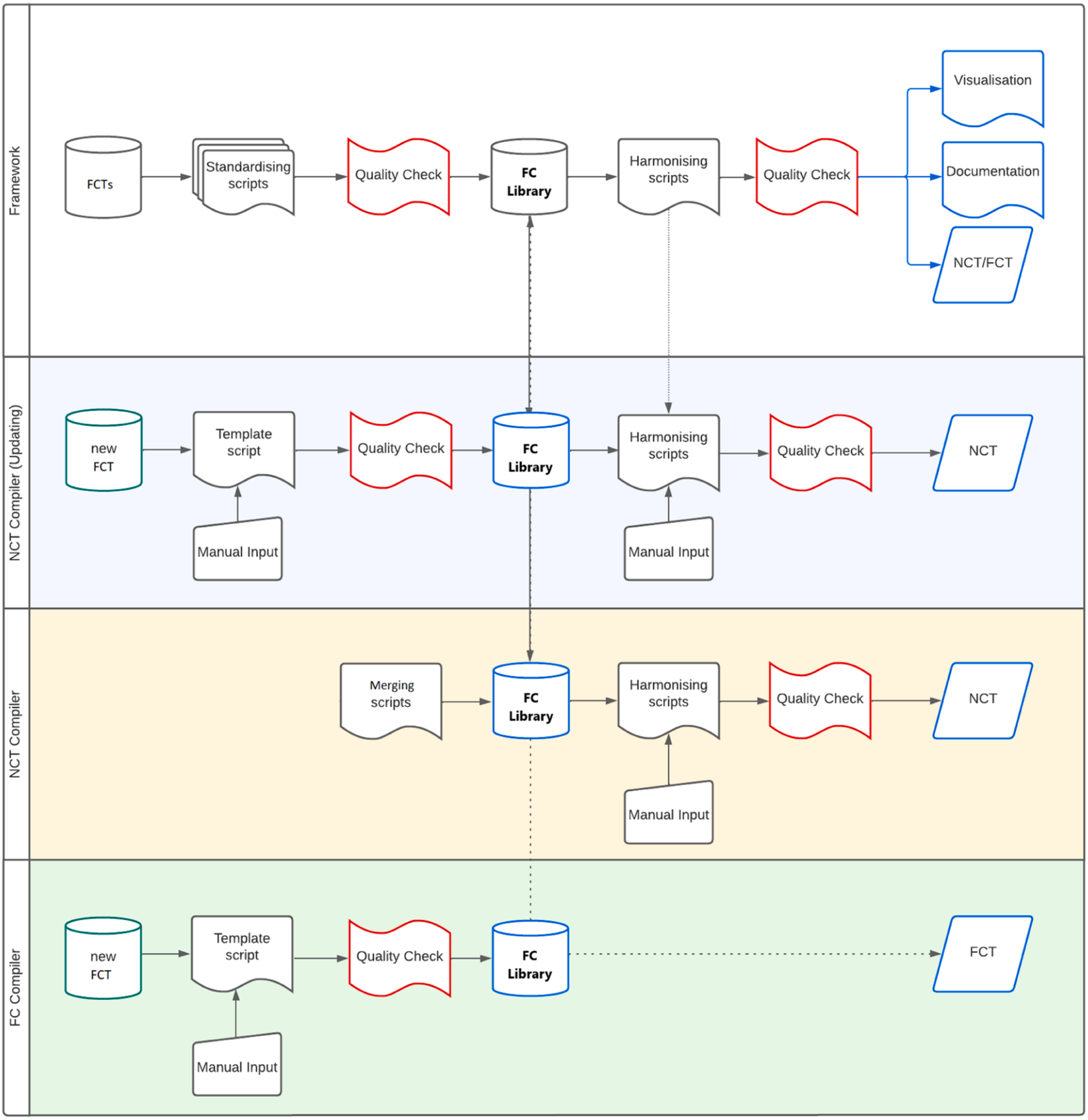

Fig. 1. The Framework (white background) presents the complete data workflow from the extraction of the "raw" Food Composition Tables (FCTs) to the standardisation into a Food Composition (FC) Library and harmonisation to the final generation of Nutrient Conversion Table (NCT), documentation and data visualisation. The NCT Compiler (updating) (blue background) presents the data journey of a user who aims to update any NCT. The NCT Compiler (yellow background) presents the data journey of a user who aims to use the standardised FCTs generated within this framework to produce an NCT. The FC Compiler (green background) presents the journey of user that aims to standardise a new FCT that could be integrated into the overall framework. The arrows across the lanes represent the steps of the workflow where users could benefit from or contribute to the workflow. The Food Composition (FC) data Library refers to one or more standardised Food Composition Tables (FCTs) which are or can be part of the Food Composition (FC) data Library generated in this study.

description (*food_desc*), etc.) (Supplementary Table 1A).

2.3.3. Standardisation of values

Firstly, special characters (e.g., "*", "[]", "-") and character strings (e.g., "trace", "LOD", "N") which are often used within FCTs need to be converted into numeric values to allow mathematical operations. For instance, values displayed alongside special characters (such as, "[45.6]") usually denote "low quality values" and/or a different method of analyses/ tagname (e.g., "Total fat using a mixed solvent extraction vs Soxhlet method"). The special characters are removed and, if a suitable tagname is available, the values are reported in a new column with the appropriate tagname. Furthermore, documentation is generated and added as metadata indicating, for instance, "low quality value" for a given food entry and component. This is stored in a new variable (called "*comments*") to retain full transparency in the data processing decision steps and for informing end-users. Similarly, for special characters and/

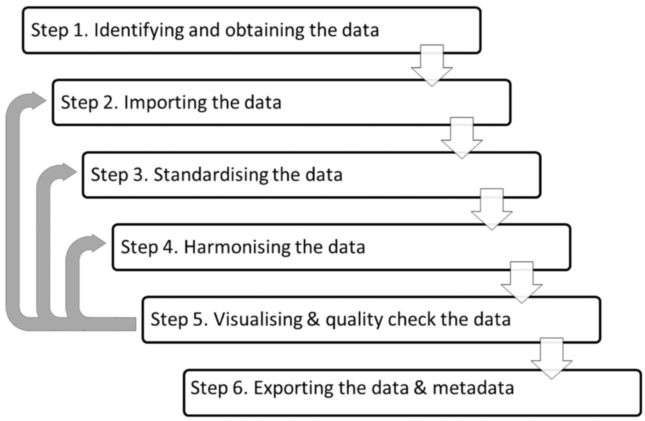

Fig. 2. The six main steps framework: 1) identifying and obtaining the data (e.g., "raw" Food Composition Tables (FCTs)), 2) importing the "raw" FCTs into R, 3) cleaning and standardisation of the FCTs into a common food composition library, 4) harmonisation of the data (including food matching), 5) checking quality and completeness of the food composition data, and 6) compiling the Nutrient Conversion Table (NCT) and relevant documentation and exporting in a standard format (e.g. Microsoft Excel, Word, etc.). Grey arrows indicate iterative feedback.

or strings used to indicate missing values and trace or below limits of detection, information is added to the "*comments*" variable, and characters are transformed to "*NA*" and zero, respectively.

2.3.4. *Units of measurements*

Finally, food components are occasionally expressed using different units (e.g., g of calcium per 100 g of fresh weight, edible portion, or mg of calcium per 100 g of fresh weight, edible portion) or denominators (e.g. per 100 g of fresh weight or per 100 g of fatty acids) between FCTs. Both units and denominators were standardised following the "FAO/INFOODS Guidelines for Converting Units, Denominators and Expressions" (FAO/INFOODS, 2012c).

2.4. *Data compilation and harmonisation*

After the FCTs are standardised, they can be compiled into a single food composition data library because they share the same variable names, units, and structure. The next sections outline the steps proposed for harmonisation and evaluation of FCT/NCTs.

2.4.1. *Food description classification/harmonisation and food matching*

One critical and time-consuming step in generating NCTs is food matching, which is the process of linking a food item (or group of food items) with the relevant foods described in the FCT. In our case study,

Table 1
List of the Food Composition Tables and Databases (FCTs) standardised and compile in the food composition data library used to validate and expand the energy and nutrient information for 95 fish and other aquatic products, including number of food items, the coverage, and data provenance.

FCT id.	FCT name	Reference	Food items (N)	Coverage	Access	Format (version)	Link to the original data
US19	USDA National Nutrient Database for Standard Reference, Legacy Release	United States Department of Agriculture USDA, 2019 (1)	7793	The United States of America	Publicly available	MS Access (legacy release)	https://agdatacommons.nal.usda.gov/articles/dataset/USDA_National_Nutrient_Database_for_Standard_Reference_Legacy_Release/24661818
AU19	Australian Food Composition Database	FSANZ Food Standards Australia New Zealand, 2019(2)	1534	Australia	Restricted-use licence[(1)]	MS Excel (version 1)	https://www.foodstandards.gov.au/science-data/monitoringnutrients/afcd/australian-food-composition-database-download-excel-files
NZ18	New Zealand Food Composition Database	New Zealand Institute for Plant and Food Research Limited & Ministry of Health, 2019 (3)	2767	New Zealand	Publicly available	MS Excel (version 01)	http://www.foodcomposition.co.nz/foodfiles
DK19	Frida: Food Database	DTU Technical University of Denmark Food Institute, 2019 (4)	1186	Denmark	Publicly available	MS Excel (version 4)	https://frida.fooddata.dk
WA19	FAO/INFOODS Food Composition Table for Western Africa	(Vincent et al., 2020) (5)	1028	Western Africa region	Publicly available	MS Excel (-)	https://www.fao.org/fileadmin/user_upload/faoweb/2020/WAFCT_2019.xlsx
KE18	Kenya Food Composition Tables	FAO & Government of Kenya, 2018 (6)	658	Kenya	Publicly available	MS Excel (-)	https://nutritionhealth.or.ke/programmes/healthy-diets-physical/food-composition-tables/
IN17	Indian Food Composition Tables	Longvah et al., 2017 (7)	528	India	Restricted-use licence[(2)]	MS Excel (-)	Not available
JA15	Standard Tables of Food Composition in Japan	MEXT, 2015 (8)	2191	Japan	Publicly available	MS Excel (version 7)	http://www.mext.go.jp/a_menu/syokuhinseibun/1365451.htm/
BA13	Food Composition Table for Bangladesh	Shaheen et al., 2013 (9)	381	Bangladesh	Publicly available	MS Excel (-)	https://www.fao.org/fileadmin/templates/food_composition/documents/FCDB_7_4_14.xlsx
BR11	Brazilian FCT (TACO)	NEPA-UNICAMP Núcleo de Estudos e Pesquisas em Alimentação – Universidade Estadual de Campinas, 2011(10)	597	Brazil	Restricted-use licence	MS Excel (version 4)	http://www.nepa.unicamp.br/arquivo/uploads/taco-4a-edicao/taco-4a-edicao-2/
UF16	FAO/INFOODS Global Food Composition Database for Fish and Shellfish (uFiSh)	FAO, 2016 (11)	515	Global	Publicly available	MS Excel (version 1.0)	https://www.fao.org/fileadmin/templates/food_composition/documents/uFiSh1.0.xlsx /
NO21	The Norwegian Food Composition Table	Norwegian Food Safety Authority, 2021(12)	2070	Norway	Publicly available	MS Excel (-)	https://www.matportalen.no/verktoy/the_norwegian_food_composition_table/
Footnote	[(1)]The file is restricted because the version 1 has been replaced with the new release (v.2). [(2)]Only publicly available in pdf which is currently not available.						

the original fish and other aquatic products in all selected FCTs, except for one, the Norwegian FCT (2021), were previously matched with their corresponding SUA item by experts from FAO's Food and Nutrition Division, as part of the compilation of the Global NCT for SUA. A detailed description of the methodology and principles applied for the food matching are documented elsewhere (Grande et al., 2024). In brief, the highest quality (i.e., the highest similarity between the SUA item reported and the FCT item description) in raw form of the food was matched, unless specified as "prepared" in the SUA item description. For fisheries and aquatic products, the scientific names were used to classify and identify the food items using the ISSCAAP code and 3-alpha codes which together with the food name description aided the food matching process.

The Norwegian FCT (2021) was used to expand the nutrients, and as a case study to develop and test a semi-automated food matching process for fish and other aquatic products. The semi-automated food matching used the ISSCAAP groups (i.e. 50 groups in which commercial species are grouped based on their taxonomic, ecological and economic characteristics) and the harmonised food description which was based on the previous work of Grande and colleagues (2024), and coded in R.

The first step was the food entry classification/identification using the ASFIS list (FAO, 2022) which contains information of the scientific names and common names of 13,420 species for fisheries statistics and their corresponding taxonomic, ISSCAAP group and the 3-alpha code (i.e. a unique three letter code for each species allowing for easier inter-agency data exchange). When possible, food entries (i.e., fish species) were directly linked (i.e., joining two datasets together) using the scientific name. Where this did not successfully identify matches, an approach was developed to facilitate the semi-automatic matching of names, using approximate matching (e.g. fuzzy matching) based on the scientific name or the common names (in English).

The second step was the harmonisation of the food description based on the one-to-seven SUA item descriptor for fish and other aquatic products: "Fresh" (1), "Frozen Whole" (2), "Fillets" (3), "Frozen Fillets" (4), "Cured" (5), "Canned" (6) and "Preparations" (7). Whereby using string identification of terms such as "raw", "whole", "fillet" or "dried" each food entry was assigned to one of the seven food description groups.

Finally, food matching between the food entries in the Norwegian FCT (2021) and the SUA items for fish and other aquatic products were performed using the ISSCAAP group and the food description codes. For example, the food entry "Whiting, raw" was matched to SUA item demersal, fresh, whole (1514) according to the ISSCAAP group (32) and the food description code (1).

Manual identification of the remaining unmatched food entries and checks for the coherence of all the matches were performed. Food match quality criteria adapted from the "FAO/INFOODS Guidelines for food matching" (FAO/INFOODS, 2012b) was assigned to all matches, based on the similarity of the food description between the foods matched (Grande et al., 2024).

2.4.2. Dealing with missing values

When compiling an NCT, missing values should be avoided in the food components of interest and in those that are needed to (re-)calculate other nutrients, for example, retinol and provitamin A carotenoids to calculate vitamin A equivalents (Moltedo et al., 2021). Within this framework, scripts were developed to identify any missing values for individual food components, and to perform conventional and alternative approaches to address and reduce the number of missing values. These operations, which are described in the following sections, may affect the data quality and the derived results, hence, all of them are performed by independent functions and/or scripts that can be omitted if new and/or more accurate data become available. In addition, metadata were added (to the "*comments*" variable) for their identification, and for performing sensitivity analyses.

2.4.2.1. Food component imputation. Data imputation is used for food components in the following situations: 1) "borrowing" value(s) from similar food(s) in case of a missing value; 2) using a value reported in the original FCT when it should be calculated for harmonisation purposes (see Section 2.4.3) (for instance, beta-carotene equivalents should be calculated using values of provitamin A carotenoids; however, in cases when the individual components are not provided the beta-carotene equivalent value is imputed from the original FCT for the same food item); and 3) assuming zero, for example when a value is calculated and yielded a negative result, e.g., values for carbohydrates calculated by difference from the other proximate values, and when components are not naturally present in a food, including alcohol assumed to be zero in all foods except alcoholic drinks and some fermented products, and fibre in animal-source foods with exception of insects and food products and preparations (FAO/INFOODS, 2012a).

2.4.2.2. Food component combination. Certain food components are expressed in FCTs using different tagnames according to the fraction analysed (e.g., vitamin D3, vitamin D2, resulting in tagnames <ERGCAL> and <CHOCAL>, respectively), or the method of analysis (e.g., total vitamin B6 analysed with microbiological assay or HPLC, resulting in tagnames <VITB6A> and <VITB6C>, respectively). Here, a function combines and stores them into a new variable, where appropriate, which is named with the respective tagname and the word "*compiled*" (e.g., <VITB6-compiled >) and information about the original tagname is stored as metadata.

2.4.2.3. Food component back-calculation. There are cases where back-calculation (i.e., inferring or calculating one nutrient from other(s)) is needed, for example when calculating edible portion factor from refuse factor (Suppl. Mat. Eq.1a-b). Additionally, there are special instances where this is used for reducing the total number of missing values. For example, for retinol, or beta-carotene equivalent, when it could not be re-calculated (i.e., individual carotenoids values were missing) or imputed from beta-carotene equivalent, then in some specific foods beta-carotene equivalent or retinol can be back-calculated (Suppl. Mat. Eq. 3 and 4a-b) using an iteration of the equation to calculate vitamin A expressed as Retinol Activity Equivalent (RAE) and/or vitamin A as Retinol Equivalent (RE) (Suppl. Mat. Eq.2b-c).

2.4.3. Food component re-calculation

According to the FAO/INFOODS guidelines, some food components should be re-calculated from other food components even when they are reported in the original FCT. For example, sum of proximate (g), energy (kcal, kJ), carbohydrates available, by difference (g) among others (Suppl. Mat. Eq. 2a-g) (FAO/INFOODS, 2012a; Grande et al., 2024; Moltedo et al., 2021). Functions to perform those calculations were developed and combined into the R package: *NutritionTools* (Codd, Segovia de la Revilla, 2023).

2.5. Quality checks and visualisation

Iterative quality checking and processing script updating was used to ensure that any inconsistencies missed in the processing scripts were identified and rectified at the appropriate location in the processing steps. Identifying data inconsistencies that can lead to missing values is one of the main tasks when compiling and quality checking an FCT/NCT. Here, we used the *naniar package* (Tierney and Cook, 2023) for missing value visualisation and analysis. This step was essential to identify any typos or issues in previous steps, e.g., renaming the food components, unit of measurement transformations, and to evaluate the potential food components for inclusion/ exclusion for the final NCT depending on the number of missing values. Similarly, histograms, density plots, and boxplots were employed to assess the data availability, variability and identify potential implausible values. For instance, a script is available to generate a histogram of the sum of proximate components that allow for the identification of values within and outside the acceptable (95–105 g per 100 g of fresh weight, edible portion) and preferred (97–103 g per 100 g of fresh weight, edible portion) range or to identify items with unrealistic values (e.g., above 30 g of protein in fresh fish) as recommended by the FAO/ INFOODS guidelines (FAO/INFOODS, 2012a).

2.6. Data compilation and documentation

The final dataset corresponding to the NCT contained information on food components for fish and aquatic products as part of the Global NCT for SUA (Grande et al., 2024). Of which, energy and 15 nutrients were previously compiled by experts from FAO's Food and Nutrition Division, serving as a validation of the framework. To facilitate the comparison a script with a specific formatting structure was developed to generate and export the NCT into commonly used and understood formats, such as a Microsoft Excel workbook. Additionally, nine nutrients (total saturated, monounsaturated, polyunsaturated fatty acids, docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), vitamin B6, B12, copper, and selenium) were compiled and expanded for the fish and aquatic products to show-case the implementation of the framework. The complete NCT could be used within R or exported to other software, which can then be visually inspected, peer-reviewed and/or used in multiple software commonly available to researchers working with FCT/NCT datasets.

3. Results

Twelve FCTs of national, regional or global coverage from 2011–2021 were standardised, which included renaming food components and assigning tagnames, standardising units and denominators, and compiling food components and formatting into a unique structure to build the food composition data library (n=24,429 food entries). The original FCTs were in multiple data structures which conditioned the length and complexity of the scripts developed and used. For instance, well-formatted tabular FCTs, such as the FAO/INFOODS FCT for Western Africa (2019) or the Kenya FCT (2018), only needed one or two lines of code to import the dataset, while a Microsoft Access relational database, like US Department of Agriculture (USDA) (2019), may need to load the index of files, then identify the tables that are related to food composition data and then import those files. Table 1 presents the FCTs included and related information, including number of food items, and data provenance.

The classification of all fish and other aquatic products available in the selected FCTs (n=1846) was harmonised using the ASFIS list, i.e. identified according to their ISSCAAP group and assigned 3-alpha codes, and then were matched to one or more of the 95 SUA items for fish and other aquatic products. This resulted in a total of 4855 matched foods from the food composition library meaning that in many cases the same food from the FCTs was matched to more than one SUA item. One important source of duplication was the use of the same food items from FCTs for both "fresh" and "frozen" SUA items, since this description was rarely included in FCTs. Suppl. Table 3 shows the number of unique food items included per FCT whereas the Fig. 3 shows the number of effective food matched to each SUA item. Out of the total matched foods from the food composition library, 233 items were matched from the Norwegian FCT (2021), and 97 % (n= 225) were successfully matched using the semi-automatic matching developed here. When accounting for duplicates, on average, each SUA item was matched to 51 (range 1–290) foods from compiled FCTs, with the largest proportion extracted from The Standard Tables of Food Composition in Japan (2015) (19 %, n=947), followed by FAO/INFOODS Global FCT for Fish and Shellfish (uFiSh) (2016) (17 %, n=819) and US Department of Agriculture (USDA) (2019) (17 %, n=816), while the FCTs with fewest matches (<3 % of total) were the Brazilian FCT (2011) and the Kenya FCT (2018) (n=110 and n=118 respectively; Fig. 3).

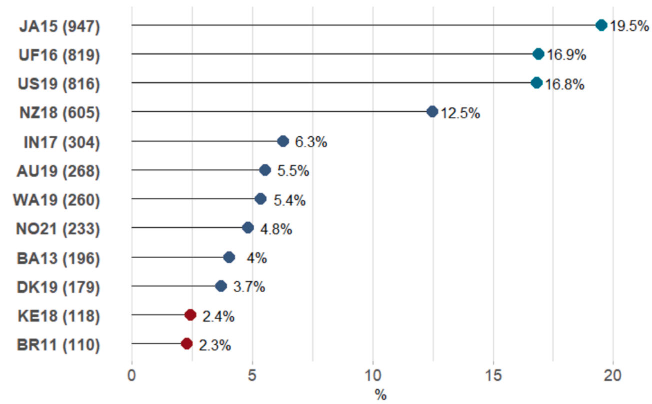

Fig. 3. The percentage of fish and other aquatic products contribution to the Global Nutrient Conversion Table from each Food Composition Table and Database, and in parenthesis is the number of food entries (including duplicated food entries being matched to multiple SUA items (see results section)) that were included. In alphabetical order: Australian Food Composition Database (AU19); Food Composition Table for Bangladesh (BA13); Brazilian FCT (TACO) (BR11); Frida: Food Database (DK19); Indian Food Composition Tables (IN17); Standard Tables of Food Composition in Japan (JA15); Kenya Food Composition Tables (KE18); The Norwegian Food Composition Table (NO21); New Zealand Food Composition Database (NZ18); uFiSh: FAO/INFOODS User Database for Fish and Shellfish (UF16); USDA National Nutrient Database for Standard Reference, Legacy Release (US19); FAO/INFOODS Food Composition Table for Western Africa (WA19).

Missing values were evaluated in the matched foods only (Fig. 4a) for each food component, and specific approaches were taken to resolve them, for instance, data for different tagnames were combined for certain components, using the function *nutri_combiner()* (Codd and Segovia de la Revilla, 2023). As a result, missing values for thiamin, fat and vitamin B6 were reduced by 99 %, 91 % and 88 %, respectively (Fig. 4b). Similarly, while most of the beta-carotene equivalent values were recalculated from the pro-vitamin A carotenoids, 6 % (n=289) were imputed from beta-carotene equivalent as presented in the original FCTs using the function *CARTBEQ_standardised()* (Codd and Segovia de la Revilla, 2023). After missing values were resolved, seven food components were re-calculated for matched foods using functions, for instance, sum of proximate, and energy (kcal)/(kJ), carbohydrates available, by difference, or vitamin A (RE)/(RAE) amongst others presented in Section 2.4.

General quality checks were then performed, first for the matched foods and then, for the average values corresponding to each of the 95 SUA items. For instance, 12 food entries were outside the acceptable range of the sum of proximate components (95—105 g), all of which were due to overestimation of the proximate values (Fig. 5a). When aggregated at SUA item level, the sum of proximate values, which were calculated based on the averaged/ re-calculated values, were all within acceptable range (FAO/INFOODS, 2012b; Greenfield and Southgate, 2003). In addition, protein was below 30 g per 100 g of fresh weight, edible portion for all the food entries considered "fresh fish" (e.g., "fresh", "frozen whole", "fillets", etc.) (Fig. 5b).

Finally, from the initial 89 tagnames collected from FCTs, data for 32 food components were compiled and reported for the 95 SUA items for fish and aquatic products (Supplementary Table 1A). From the food components compiled, energy, edible portion and other 21 food components, which were previously compiled by experts from FAO's Food and Nutrition Division, showed comparable results serving this compilation as a validation of the process, additionally nine nutrients (total saturated, monounsaturated, polyunsaturated fatty acids, docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), vitamin B6, B12, copper, and selenium) were compiled only for the fish and aquatic products and expanded as part of the Global NCT for SUA (Grande et al., 2024).

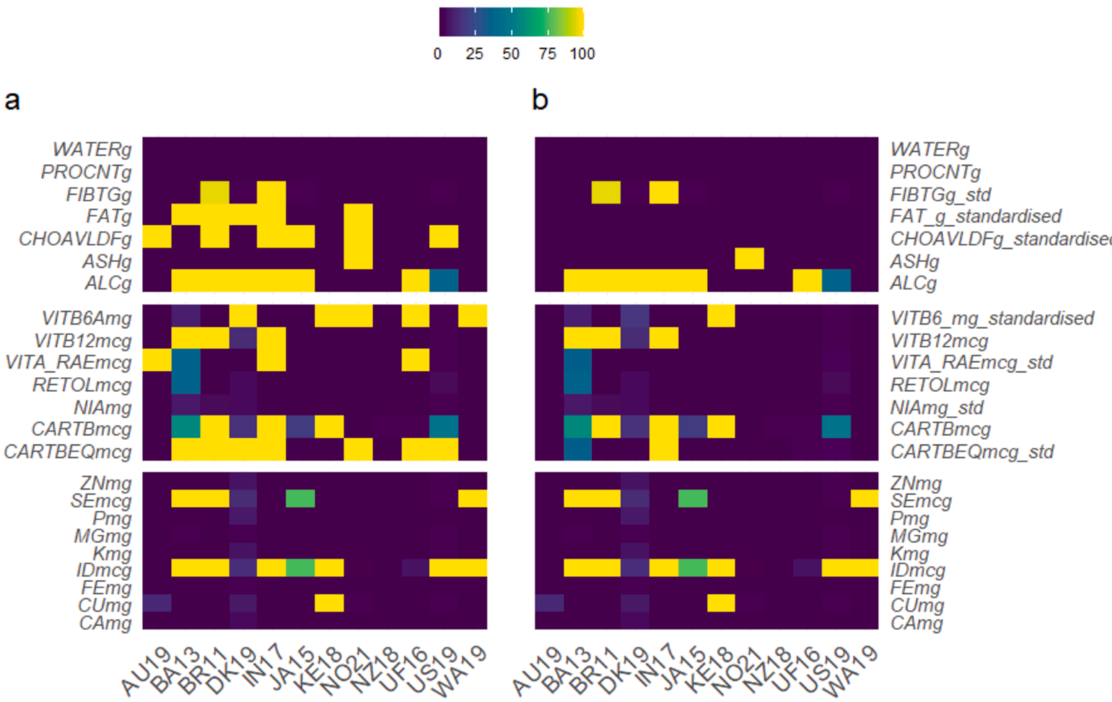

Fig. 4. Heatmap representing the missing values (% from 0 (dark blue) to 100 (yellow)) in proximate, vitamin and mineral values in the fish and other aquatic products after the standardisation and before dealing with the missing values (a), and after dealing with missing values (b) as part of the harmonisation steps.

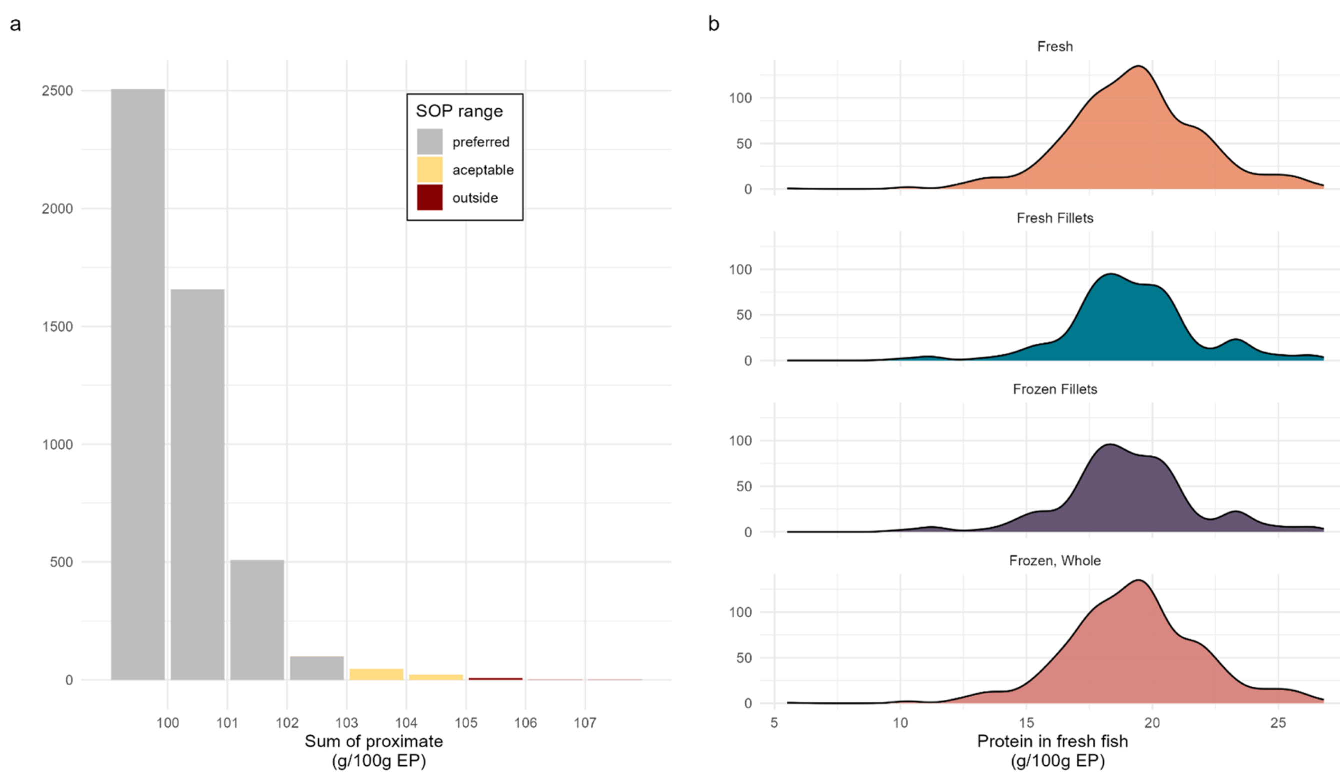

Fig. 5. Example visualisation for quality check of food composition or nutrient conversion tables: a) the histogram of the Sum of Proximate (SOP) with the grey bins representing the preferred range (97–103 g/ 100 g EP), the yellow bins representing outside preferred range and within the acceptable range (93–95 and 103–105 g/ 100 g EP) and the red bins representing the outside acceptable range, and b) the density plots of the protein (g/ 100 g EP) content in fresh fish entries grouped by their state (e.g., raw, frozen, fillet etc).

This global NCT was applied to the SUA items and statistics on energy and nutrient availability at national level and are presented for 186 countries as part of the FAOSTAT Food and Diet Domain (FAO, 2024a), a web-hosted portal dedicated to the dissemination of statistics on different types of dietary data.

The R scripts developed in this study are presented in the public repository published in GitHub (https://github.com/LuciaSegovia/FAO-fisheries) which provides commands for cleaning and standardisation of the 12 FCTs, harmonisation and compilation of the NCT, including visualisation and quality checks, functions to re-calculate energy and other food components, food matching aid and FCT/NCT output formatting. All functions are published as an R package: *NutritionTools* (Codd and Segovia de la Revilla, 2023). Furthermore, a food composition data library with the nine open and freely accessible FCTs can be compiled by running the scripts in the repository. For the other FCTs of interest, scripts could be re-used if data access/licences permit, or scripts can be adapted to current publicly available version of the data (e.g. Food Standards Australia New Zealand, release 1 (2019), to script to Food Standards Australia New Zealand, release 2 (2023), see Table 1). Moreover, a template for processing new FCTs is included together with a manual that describes every step and provides guidance on standardisation decisions.

4. Discussion

4.1. Data standardisation and harmonisation framework for food composition data

A framework and the R tools to ingest, process and standardise FCTs and to compile and report FCTs/NCTs was developed. Our objective was to increase efficiency, reproducibility, and transparency in the processing of food composition data for nutrition.

Several projects have undertaken standardisation/harmonisation of FCTs for Europe, such as the European Prospective Investigation into Cancer and Nutrition (EPIC) Nutrient Database (ENDB) (Slimani et al., 2007), European Food Information Resource (EuroFIR) (Finglas et al., 2014) and, more recently Stance4Health (Hinojosa-Nogueira et al., 2021). All of them, including our project, have faced analogous data challenges, i.e., lack of food component and food description standardisation, diverse measurement of units, missing and/or implausible values, etc. Despite the amount of thought and effort invested in the cleaning and standardisation process from numerous experts and projects, most of the steps and/or clean datasets are not openly available. This results in a lack of transparency and reproducibility of the methods, concerns about the reusability of the cleaned FCTs/NCTs, and ultimately, researchers repeating the process (Clancy et al., 2015). To address these issues, we developed an open science workflow, generating and publishing all the data processing steps as scripts in a format compatible with an open and freely available software. Thus, our processing can be readily reproduced by anyone able to access the original FCTs, to avoid continued duplication of effort. Moreover, the use of scripted approaches allows for full audit of the data manipulation/decisions performed when generating a food composition library and NCT (Coates et al., 2017; Micha et al., 2018).

4.2. Framework re-usability: the users' journey

There are numerous ways in which the scripts and functions can be used as provided, or further enhanced and adapted, to support the principles of FAIR data in food composition science and nutritional assessments using NCTs. Four example users of the open science workflow are provided to demonstrate relevance for the nutrition research community (Fig. 1). Three of the user examples are described in detail below while the fourth, represented in the green band in Fig. 1, and which is interested in standardising a new FCT only, would just need to adapt the standardisation scripts, as described in the Example 2.

In all cases, the first step will be to visit the GitHub repository for the scripts and functions used in this study and follow the instructions to obtain the data and tools as needed. Data provenance are reported in Table 1 and within the repository.

4.2.1. Example user 1: reproducing the global NCT for SUA for the fish and other aquatic products

The first example represented in the white band in the Fig. 1 (as "Framework") presents a user aiming to reproduce the steps undertaken in this project, to replicate the energy and nutrient values of the 95 SUA items for fish and other aquatic products compiled as part of the global NCT (Grande et al., 2024), and for which statistics are presented in the Food and Diet Domain (FAO, 2024a). The annotated scripts can be followed to review decisions on data cleaning and standardisation. With the scripts, which contain the necessary functions, and FCTs all in place, the user could run the workflow and obtain the NCT with 95 SUA items for fish and other aquatic products in less than five minutes. The majority of the scripts will work even if one or more FCT is not included. However, we would recommend that our documented decisions are reviewed and consciously adopted, or adapted, by the user, according to their needs. Nonetheless, the workflow will save considerable time compared to the effort required to recreate all the steps involved; it will also generate a traceable record of decisions made within the NCT preparation, which is often lacking (Clancy et al., 2015; Coates et al., 2017; Pennington et al., 2007; Segovia de la Revilla et al., 2023).

4.2.2. Example user 2: standardising an additional FCT and expanding the current NCT

The second example presents the situation of a food composition data compiler who wishes to add an additional FCT to the current list, for instance, to update or expand the energy and/or nutrient for the 95 SUA items for fish and other aquatic products. These steps are presented in the blue band in Fig. 1 (as "NCT Compiler (Updating)") and require more time and technical expertise (i.e., R programming skills) from the user than required in the first example user, however by copying and adapting existing FCT import scripts, this activity can be accomplished with considerably reduced effort, and increased transparency compared to manual inclusion of the additional FCT.

Firstly, to standardise a new FCT, the user is advised to check the quality of the FCT by using the FAO/INFOODS evaluation framework (Charrondiere et al., 2023). Following this, a template script, which can be accessed here (https://github.com/LuciaSegovia/FAO-fisheries/tree/main/00_template) for import and standardisation is provided which is designed to help the user navigate the scripting of the steps 1 to 3 of the framework (Fig. 2). The template script provides guidance and options covering the most common tasks for an array of FCT formats, such as those detailed in Sections 2.2 and 2.3 of the methods. After completing the standardisation steps the user can either export the standardised FCT in a tabular format (e.g., Microsoft Excel, text-delimited), or include it in the food composition data library. Once an FCT is standardised no extra user inputs are needed for merging it with the food composition data library available in the repository (which can be obtained by running the *merging_all.R* script). The R scripts that perform the harmonisation steps (Fig. 2) (e.g., harmonisation of food names and description, dealing with missing values, etc.) should be reviewed and updated and the R scripts re-run. For instance, if the aim is to update energy and/or nutrient information for the 95 SUA items for fish and other aquatic products with this additional FCT, the user needs to adapt the scripts of the semi-automatic standardising of the food description and food matching between the food entries in the additional FCT and the 95 SUA items: this can be adapted from the *NO21_harmonising.R* script. Then, the other harmonisation, quality checks, visualisation and formatting steps can be performed as described in the method section using existing scripts, without further adaption, which produce the traceable updated/extended NCT and metadata.

4.2.3. Example user 3: The Nutrient Conversion Table compiler: Re-usability of the food composition data library

The third example represents a user wishing to obtain a survey specific NCT, e.g. for the list of foods reported as consumed and/or acquired in a household consumption and expenditure survey (as the "NCT Compiler" in the yellow band in Fig. 1). This user needs the highest R/data literacy of the examples provided whilst it requires similar, or less time and effort than in the second example. This is because all the import and standardisation tasks for the entirety of the FCTs, comprising steps 1 – 3 of Fig. 2, are already scripted and documented. The product, after obtaining the original FCTs and running the scripts, is a standardised food composition data library which contains food items (n=24,429) from 12 high quality FCTs covering different regions of the world. From the library the user can extract information from 32 food components and/or benefit from the harmonisation and formatting scripts available to compile a new NCT. This, in turn, would reduce considerably the time and effort needed while increasing the transparency and reproducibility of the output. Nevertheless, the user would need to implement and adapt some of the scripts/steps, such as, the food matching between the food composition data library and the food consumption dataset, setting rules and priorities as appropriate (e.g. using a country/region specific FCT as the main source for performing the food matching). An additional function was developed within this framework, the *Fuzzy_Matcher(),* that provides an aid to this time consuming step. The output, a user-led matched dataset, can be integrated with the subsequent functions and scripts to perform the rest of the harmonisation, quality check and formatting steps. Particular attention should be given to the food matches which should be carefully checked for coherence and context relevancy. An example of the use of food composition data library and the *Fuzzy_Matcher()* for generating household survey NCTs can be found in the MAPS project repository here: [https://github.com/micronutrientsupport].

4.3. Strengths and limitations

4.3.1. Strengths

This study generated a novel open science workflow that can increase the findability, interoperability and reproducibility of FCTs and derived NCTs. The workflow will reduce the time and effort needed by other food composition data compilers and users, as exemplified by three use-cases. By providing a framework and the R tools (i.e., repository) in a freely available and open software, an array of different users can adopt and adapt these steps into their own workflow.

Furthermore, the scripted approach outlined here ensures that each data processing decision and assumption that may influence outputs derived from FCTs, is recorded and annotated. The documentation (reporting and metadata) proposed within this framework (e.g., reusable scripts and functions, reporting and exporting structure, etc.) provides a solution to the insufficient reporting of the food composition data found in most of FCTs/NCTs (Bruyn et al., 2016; Ispirova et al., 2020; Segovia de la Revilla et al., 2023). Additionally, the detailed documentation and visualisation for quality checks can aid data processing decisions and allow for revisiting assumptions. This improves transparency and reproducibility, and reduces the uncertainty around data and derived outputs.

4.3.2. Limitations

One of the main limitations of the study is that, although all the scripts, functions and decisions made are recorded and publicly available, the raw data used (i.e., FCTs) are not all publicly available, and some have been updated since the publication of the global NCT. For instance, the Food Standards Australia New Zealand (FSANZ Food Standards Australia New Zealand, 2019) has been replaced with a newer version, and the Indian Food Composition Tables (2019) is only available as pdf. Furthermore, the impact of the data processing decisions (including imputation, calculation, etc.) on the energy and nutrient supplies was not evaluated. For instance, the influence of the combination of different tagnames reporting different methods of analysis (i.e., used for vitamin B6) or transforming trace and below detection limit values to zero had on mean nutrient values for each SUA item.

Another limitation related to the quality checks is that despite the sum of proximate components frequently being used an indicator of data quality, as suggested by most of the food composition compilation guidelines (FAO/INFOODS, 2012a; Greenfield and Southgate, 2003), it is only a reliable measurement of quality for analytical values. Here, we used carbohydrates available, by difference (i.e., calculated from the other proximate values) which is not a reliable component for its calculation given that the same proximate components are used in both the calculation of carbohydrates available, by difference and sum of proximate. However, this decision was made as carbohydrates available, by weight were not available in all FCTs used in the present work and for most of the SUA items included in this case study this component would be zero or assumed zero (FAO/INFOODS, 2012a). Other quality checks that could be used instead are 95 % confidence intervals and/or implausible values detection per food group. Nevertheless, these checks may not be effective for all food components as there is high variability in the concentration of certain food components due to multiple factors, such as: broad spectrum of food entries within each SUA item (e.g., sea urchins and turtles are considered under the same SUA item), often compounded by limited number of values available for certain food components and, low or uncertain quality of the values. This uncertainty around the food component values makes the establishment of quality checks difficult and increases the need for expert assessment and inputs (FAO, 2024a).

Finally, we acknowledge that resources are needed to regularly update the R scripts and packages to maintain its functionality and compatibility with newer R versions and packages, ensuring its relevance for the community. Further development of the framework and the tools could also be undertaken by the network of food composition experts and data users who will benefit from this repository.

5. Conclusion and recommendations

To the best of our knowledge, this is the first example of a comprehensive method to develop NCTs suitable for global application in a reproducible, reusable, efficient, and transparent manner. The NCT output has been used in the published statistics based on SUA data for 186 countries on the FAOSTAT Food and Diet Domain (FAO, 2024a; 2024b). Statistics comprise energy and nutrient availability for 26 nutrients including the nine nutrients expanded by the present work for fish and other aquatic products.

Open science offers opportunities to greatly reduce the resources required to compile food composition data for nutrition research. Here, scripts and functions are provided aimed at making the data processing more reproducible, reusable, efficient, and transparent. To further support open research, the food composition community should agree and implement standardised practices for data management and documentation (i.e., minimum information standards, metadata, ontologies) that improve reproducibility, transparency, efficiency and (re-) usability of future FCTs and NCTs.

Disclaimer

The views expressed in this publication are those of the author(s) and do not necessarily reflect the views or policies of the Food and Agriculture Organization of the United Nations.

CRediT authorship contribution statement

E. Louise Ander: Writing – review & editing, Writing – original draft, Project administration, Investigation, Funding acquisition, Conceptualization. **Lucia Segovia de la Revilla:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Codd:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Edward J.M.**

Joy: Writing – review & editing, Supervision, Project administration, Methodology. **Liberty Mlambo:** Writing – review & editing, Validation, Software. **Fernanda Grande:** Writing – review & editing, Supervision, Resources, Methodology. **Doris Rittenschober:** Writing – review & editing, Resources, Methodology. **Ana Moltedo:** Writing – review & editing, Supervision, Resources, Methodology. **Bridget A. Holmes:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Food and Agriculture Organization of the United Nations (FAO) (PO Number 352277). In addition, this work was supported in part by the Bill & Melinda Gates Foundation [INV-002588; Micronutrient Action Policy Support (MAPS) project]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The funder had no role in the study design, data analysis, decision to publish, or preparation of the manuscript. We also would like to thank Claire Dooley and Elaine L. Ferguson for their help on the earliest versions of the manuscript

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2024.106894.

Data Availability

The authors do not have permission to share data.

References

Andrés-Hernández, L., Blumberg, K., Walls, R.L., Dooley, D., Mauleon, R., Lange, M., Weber, M., Chan, L., Malik, A., Møller, A., Ireland, J., Segovia, L., Zhang, X., Burton-Freeman, B., Magelli, P., Schriever, A., Forester, S.M., Liu, L., King, G.J., 2022. Establishing a common nutritional vocabulary - from food production to diet. Front. Nutr. 9.

Bruyn, J., de, Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., de Bruyn, J., Ferguson, E., Allman-Farinelli, M., Darnton-Hill, I., Maulaga, W., Msuya, J., Alders, R., 2016. Food composition tables in resource-poor settings: exploring current limitations and opportunities, with a focus on animal-source foods in sub-Saharan Africa. Br. J. Nutr. 116, 1709–1719. https://doi.org/10.1017/S0007114516003706.

Chan, L., Vasilevsky, N., Thessen, A., McMurry, J., Haendel, M., 2021. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. Database 2021. https://doi.org/10.1093/database/baab003.

Charrondiere, U.R., Stadlmayr, B., Grande, F., Vincent, A., Oseredczuk, M., Sivakumaran, S., Puwastien, P., Judprasong, K., Haytowitz, D., Gnagnarella, P., 2023. FAO/INFOODS Evaluation Framework to Assess the Quality of Published Food Composition Tables and Databases: User Guide. FAO, Rome, Italy. https://doi.org/10.4060/cc5371en.

Clancy, A.K., Woods, K., McMahon, A., Probst, Y., 2015. Food composition database format and structure: a user focused approach. PLOS ONE 10, e0142137. https://doi.org/10.1371/journal.pone.0142137.

Coates, J.C., Colaiezzi, B.A., Bell, W., Charrondiere, U.R., Leclercq, C., 2017. overcoming dietary assessment challenges in low-income countries: technological solutions proposed by the international dietary data expansion (INDDEX) project. Nutrients 9.

Codd, T., Segovia de la Revilla, 2023. NutritionTools: Tools for the Organisation, Matching, Calculation, and Summarisation of Nutrition Data.

Dooley, D.M., Griffiths, E.J., Gosal, G.S., Buttigieg, P.L., Hoehndorf, R., Lange, M.C., Schriml, L.M., Brinkman, F.S.L., Hsiao, W.W.L., 2018. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. Npj Sci. Food 2, 23. https://doi.org/10.1038/s41538-018-0032-6.

DTU (Technical University of Denmark) Food Institute, 2019. Frida Food Databank. Version 4. Copenhagen, DTU, National Food Institute. [Cited 29 August 2023]. https://frida.fooddata.dk.

Durazzo, A., Astley, S., Kapsokefalou, M., Costa, H.S., Mantur-Vierendeel, A., Pijls, L., Bucchini, L., Glibetić, M., Presser, K., Finglas, P., 2022. Food composition data and tools online and their use in research and policy: EuroFIR AISBL contribution in 2022. Nutrients 14. https://doi.org/10.3390/nu14224788.

Durazzo, A., Lucarini, M., 2022. Editorial: databases and nutrition. Front. Nutr. 9. https://doi.org/10.3389/fnut.2022.853600.

Ene-Obong, H., Schönfeldt, H.C., Campaore, E., Kimani, A., Mwaisaka, R., Vincent, A., Ati, J.E., Kouebou, P., Presser, K., Finglas, P., Charrondiere, U.R., 2019. Importance and use of reliable food composition data generation by nutrition/dietetic professionals towards solving Africa's nutrition problem: constraints and the role of FAO/INFOODS/AFROFOODS and other stakeholders in future initiatives. Proc. Nutr. Soc. 78, 496–505. https://doi.org/10.1017/S0029665118002926.

European Food Safety Authority, 2015. The Food Classification and Description System FoodEx 2 (revision 2) (No. 804), EFSA supporting publication.

FAO, 2016. FAO/INFOODS Global Food Composition Database for Fish and Shellfish. Version 1.0 – uFiSh1.0. Rome.

FAO, 2021. FAO Yearbook. Fishery and Aquaculture Statistics 2019/FAO annuaire. Statistiques des pêches et de l'aquaculture 2019/FAO anuario. Estadísticas de pesca y acuicultura 2019. FAO. https://doi.org/10.4060/cb7874t.

FAO, 2022. ASFIS List of Species for Fishery Statistics Purposes. Fisheries and Aquaculture Division [WWW Document]. URL. https://www.fao.org/fishery/en/collection/asfis/en (accessed 2.21.23).

FAO, 2024a. Food and Diet. Statistics on Dietary Data (No. 82), FAOSTAT Analytical Briefs. Rome. https://openknowledge.fao.org/handle/20.500.14283/cc9454en.

FAO, 2024b. Food and Diet / Availability (Supply utilization accounts). FAOSTAT. https://www.fao.org/faostat/en/#data/SUA.

FAO & Government of Kenya, 2018. Kenya Food Composition Tables [online]. Nairobi. https://www.fao.org/infoods/infoods/standards-guidelines/food-component-identifiers-tagnames/en/.

FAO/INFOODS, 2012a. FAO/INFOODS Guidelines for Checking Food Composition Data prior to the Publication of a User Table/Database-Version 1.0. FAO, Rome.

FAO/INFOODS, 2012b. FAO / INFOODS Guidelines for Food Matching Version 1.2.

FAO/INFOODS, 2012c. FAO/ INFOODS Guidelines for Converting Units, Denominators and Expressions Version 1.0. FAO/ INFOODS.

Finglas, P.M., Berry, R., Astley, S., 2014. Assessing and improving the quality of food composition databases for nutrition and health applications in Europe: the contribution of EuroFIR123. Adv. Nutr. 5, 608S–614S. https://doi.org/10.3945/an.113.005470.

FSANZ (Food Standards Australia New Zealand), 2019. Australian Food Composition Database. Release 1. Canberra. http://www.foodstandards.gov.au.

Grande, F., Ueda, Y., Masangwi, S., Holmes, B., 2024. Global Nutrient Conversion Table for FAO Supply Utilization Accounts. FAO, Rome, Italy. https://doi.org/10.4060/cc9678en.

Greenfield, H., Southgate, D.A.T., 2003. Food Composition Data: Production, Management, and Use. FAO, Rome.

Hinojosa-Nogueira, D., Pérez-Burillo, S., Navajas-Porras, B., Ortiz-Viso, B., de la Cueva, S.P., Lauria, F., Fatouros, A., Priftis, K.N., González-Vigil, V., Rufián-Henares, J.Á., 2021. Development of an unified food composition database for the European project "Stance4Health". Nutrients 13, 4206. https://doi.org/10.3390/nu13124206.

Ispirova, G., Eftimov, T., Koroušì, B., 2017. Mapping Food Composition Data from Various Data Sources to a Domain-Specific Ontology 8. https://doi.org/10.5220/0006504302030210.

Ispirova, G., Eftimov, T., Seljak, B.K., 2020. Evaluating missing value imputation methods for food composition databases. Food Chem. Toxicol. 141, 111368. https://doi.org/10.1016/j.fct.2020.111368.

Joy, E.J.M., Kumssa, D.B., 2022. Nutrient accounting in global food systems, 678–678 Nat. Food 3. https://doi.org/10.1038/s43016-022-00593-w.

Kapsokefalou, M., Roe, M., Turrini, A., Costa, H.S., Martinez-Victoria, E., Marletta, L., Berry, R., Finglas, P., 2019. Food composition at present: new challenges. Nutrients 11, 1714. https://doi.org/10.3390/nu11081714.

Klensin, J.C., Feskanich, D., Lin, V., Truswell, A.S., Southgate, D.A.T., 1989. Identification of food components for INFOODS data interchange. The United Nations University.

Longvah, T., Ananthan, R., Bhaskarachary, K. & Venkaiah, K., 2017. Indian Food Composition Tables. Hyderabad, India, National Institute of Nutrition, Indian Council of Medical Research.

Luo, H., Dodd, K.W., Arnold, C.D., Engle-Stone, R., 2021. Introduction to the simple macro, a tool to increase the accessibility of 24-hour dietary recall analysis and modeling. J. Nutr. https://doi.org/10.1093/jn/nxaa440.

MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2015. Standard Tables of Food Composition in Japan 2015. 7th revised edition. Tokyo, Official Gazette Co-operation of Japan. https://www.mext.go.jp/en/policy/science_technology/policy/ title01/detail01/1374030.htm.

Micha, R., Coates, J., Leclercq, C., Charrondiere, U.R., Mozaffarian, D., 2018. Global dietary surveillance: data gaps and challenges. Food Nutr. Bull. 39, 175–205. https://doi.org/10.1177/0379572117752986.

Møller, A., Ireland, J., 2018. LanguaL™ 2017 - Thesaurus. https://doi.org/10.13140/RG.2.2.23131.26404.

Moltedo, A., Álvarez-Sánchez, C., Grande, F., Charrondiere, U.R., 2021. The complexity of producing and interpreting dietary vitamin A statistics. J. Food Compos. Anal. 100, 103926. https://doi.org/10.1016/j.jfca.2021.103926.

Moshfegh, A.J., Rhodes, D.G., Martin, C.L., 2022. National food intake assessment: technologies to advance traditional methods. Annu. Rev. Nutr. 42, 401–422. https://doi.org/10.1146/annurev-nutr-062320-110636.

NEPA-UNICAMP (Núcleo de Estudos e Pesquisas em Alimentação – Universidade Estadual de Campinas), 2011. Tabela Brasileira de Composição de Alimentos [Brazilian Food Composition Table] (TACO). 4th edition. Campinas, Brazil. https://www.nepa.unicamp.br/publicacoes/tabela-taco-pdf.
New Zealand Institute for Plant and Food Research Limited & Ministry of Health. 2019. New Zealand FOODfiles 2018 Manual. New Zealand Food Composition Database. Auckland. https://www.foodcomposition.co.nz/foodfiles.
Norwegian Food Safety Authority, 2021. Norwegian Food Composition Database. Oslo. https://www.matvaretabellen.no.
Pennington, J.A.T., Stumbo, P.J., Murphy, S.P., McNutt, S.W., Eldridge, A.L., McCabe-Sellers, B.J., Chenard, C.A., 2007. Food composition data: the foundation of dietetic practice and research. J. Am. Diet. Assoc. 107, 2105–2113. https://doi.org/10.1016/j.jada.2007.09.004.
Posit team, 2023. RStudio: Integrated Development for R.
R Core Team, 2023. R: A language and environment for statistical computing.
Rittenschober, D., Stadlmayr, B., Nowak, V., Du, J., Charrondiere, U.R., 2016. Report on the development of the FAO/INFOODS user database for fish and shellfish (uFiSh) – Challenges and possible solutions. Food Chem. 193, 112–120. https://doi.org/10.1016/j.foodchem.2014.08.055.
Segovia de la Revilla, L., Ferguson, E.L., Dooley, C., Osman, G., Ander, E.L., Joy, E.J.M., 2023. The availability and geographic location of open-source food composition data used to estimate micronutrient intakes in sub-Saharan Africa: a scoping review. J. Food Compos. Anal. 120, 105322. https://doi.org/10.1016/j.jfca.2023.105322.
Shaheen N., Rahim A.T.M.A., Mohiduzzaman Md, Banu C.P., Bari Md L., Tukun A.B., Mannan M.A., Bhattacharjee L., Stadlmayr B., 2013. Food Composition Table for Bangladesh. Institute of Nutrition and Food Science, Centre for Advanced Research in Sciences, University of Dhaka. https://www.fao.org/infoods/infoods/tables-and-databases/asia.
Slimani, N., Deharveng, G., Unwin, I., Southgate, D. a T., Vignat, J., Skeie, G., Salvini, S., Parpinel, M., Møller, A., Ireland, J., Becker, W., Farran, A., Westenbrink, S., Vasilopoulou, E., Unwin, J., Borgejordet, Å., Rohrmann, S., Church, S., Gnagnarella, P., Casagrande, C., van Bakel, M., Niravong, M., Boutron-Ruault, M.C., Stripp, C., Tjønneland, A., Trichopoulou, A., Georga, K., Nilsson, S., Mattisson, I., Ray, J., Boeing, H., Ocké, M., Peeters, P.H.M., Jakszyn, P., Amiano, P., Engeset, D., Lund, E., Santucci de Magistris, M., Sacerdote, C., Welch, A., Bingham, S., Subar, A. F., Riboli, E., 2007. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur. J. Clin. Nutr. 61, 1037–1056. https://doi.org/10.1038/sj.ejcn.1602679.
Tierney, N., Cook, D., 2023. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. J. Stat. Softw. 105, 1–31. https://doi.org/10.18637/jss.v105.i07.
Traka, M.H., Plumb, J., Berry, R., Pinchen, H., Finglas, P.M., 2020. Maintaining and updating food composition datasets for multiple users and novel technologies: current challenges from a UK perspective. Nutr. Bull. 45, 230–240. https://doi.org/10.1111/nbu.12433.
United States Department of Agriculture (USDA), 2019. Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Legacy. Version Current: April 2019. Washington, DC, Agricultural Research Service, Nutrient Data Laboratory. https://fdc.nal.usda.gov/ download-datasets.html.
Vincent, A., Grande, F., Compaoré, E., Amponsah Annor, G., Addy, P.A., Aburime, L.C., Ahmed, D., Bih Loh, A.M., Dahdouh Cabia, S., Deflache, N., Dembélé, F.M., Dieudonné, B., Edwige, O.B., Ene-Obong, H.N., Fanou Fogny, N., Ferreira, M., Omaghomi Jemide, J., Kouebou, P.C., Muller, C., Nájera Espinosa, S., Ouattara, F., Rittenschober, D., Schönfeldt, H., Stadlmayr, B., van Deventer, M., Razikou Yiagnigni, A. & Charrondière, U.R., 2020. FAO/INFOODS Food Composition Table for Western Africa (2019) User Guide & Condensed Food Composition Table / Table de composition des aliments FAO/INFOODS pour l'Afrique de l'Ouest (2019) Guide d'utilisation & table de composition des aliments condensée. Rome, FAO. http://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases.
Zeb, A., Soininen, J.-P., Sozer, N., 2021. Data harmonisation as a key to enable digitalisation of the food sector: a review. Food Bioprod. Process. 127, 360–370. https://doi.org/10.1016/j.fbp.2021.02.005.