FoodRepo: An Open Food Repository of Barcoded Food Products

Gianrocco Lazzari, Yannis Jaquet, Djilani J. Kebaili, Laura Symul and Marcel Salathé*

Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

INTRODUCTION

Metabolic disorders, such as diabetes or obesity, have become a major public health concern, with increasingly large parts of the global population affected (1, 2). Nutritional epidemiologists hope to better understand the underlying causes, the potential treatments and prevention strategies by analyzing population and individual patterns through studies that generally rely on surveying dietary habits. Traditional food-intake survey methods are based on questionnaires filled by participants at a given frequency. The frequency of diet records is an important factor contributing to the accuracy of the study (3). Multiple-day diet records might provide good accuracy when not based on memory, but require strong motivation and time commitment by the participants. Approaches like multiple/single 24-h recalls—involving a specialized interviewer performing surveys in person or on the phone with the participants—require less engagement, but pose issues with missing data as they rely on short-term memory. Finally, so-called Food Frequency Questionnaires, where participants are asked to indicate the frequency of intake of certain foods over long periods of time (typically 1 year), demand minimal participants' commitment, therefore allowing for large cohort studies on long-term dietary habits. However, the likelihood of missing or incorrect data increases as they count on participants' long-term memory. Overall, self-reported dietary data present biases which limit their applications, especially when they heavily rely on participants' memory (4). Such limitations, which should be properly addressed in further epidemiological studies, may be overcome with more advanced recording methodologies such as dietary biomarkers and digital technologies (5).

Recent technological advances, and in particular the emergence and almost complete market penetration of smartphones, have offered interesting surveying alternatives. In particular, mobile phones have been successfully deployed in several food-related studies (6), for example using food photography (7–12). Other research has also explored the possibility of recording dietary habits by asking participants to scan the barcodes of their consumed food (13, 14). Although further investigations are required to assess self-reporting biases, these advances in nutritional research have triggered the release of mobile apps oriented mainly toward diabetes and weight-loss self-management (15–19), showing the willingness and interest of users to monitor their food intake if it provides potential health benefits.

The further expansion of self-monitoring for research and medical purposes relies on comprehensive and continuously updated food databases. A few databases of barcoded products already exist, for example Open Food Facts (20) or the USDA Food Composition Databases (21). While they each have their strength, not all of them are openly accessible or, and they often have a limited product coverage, and are often not regularly updated. For Switzerland, we did not find any database whose product coverage was sufficiently high, where the data was completely open, and easily accessible through an Application Programming Interface (API). The last point was particularly important to us, as APIs are necessary for third parties to dynamically use the data in their products and services. Our approach was therefore to build an openly accessible database of barcoded food products with sufficiently high coverage, accessible through a stable API. Rather

than focusing on a wide geographic range, we focused on a small country (Switzerland) in order to obtain the necessary coverage. The focus on the Swiss market further benefits from the need to support multiple languages from the beginning, thus making the system readily expandable to other countries, which we are now planning to do.

Here, we present this system, which we call FoodRepo (https://www.foodrepo.org), an openly accessible database of barcoded food products, and we describe the data-acquisition framework, its quality control and maintenance. Here, the word repository is meant to be understood as a data repository, where the community can deposit an increasing number of datapoints on food products. The growing community around FoodRepo and the validation of new products make our database robust, scalable and self-sustainable in the long run. Currently, the FoodRepo database mostly holds products sold in Switzerland, from the main grocery stores in the country. Its international expansion is under development.

Any item in the database is accessible through the FoodRepo website (for an example of products contained in the FoodRepo database, please see **Figure 1A**) or via our API, described in section Usage Notes. The CC-BY-4 license under which our database is released will allow its exploitation by different type of users, from academic researchers to commercial partners. For instance, a Swiss consumers association is using FoodRepo data in their NutriScan mobile app (22) to make the food package information more accessible, and to provide their users with an overall nutritional score.

Beyond this specific example, the FoodRepo database opens the way for promising research opportunities in the field of digital epidemiology and personalized nutrition. Notably, we foresee that, through dietary live-tracking, this database can support studies which combine other recent technological developments and new findings in our understanding of the human metabolism. For example, phone-connected devices for continuous monitoring of blood glucose levels have recently been made available to diabetic patients (23, 24), as well as numerous direct-to-consumer devices to estimate glucose levels have appeared on the market. A plethora of other wireless sensors are now also available to record various physiological parameters such as heart rate or blood pressure, marking a new era of "high-throughput human phenotyping" (25). Studies that would simultaneously track participants' parameters, food intake, glycemic response and physical activity might provide detailed insights on the variability of individual metabolic responses. Interestingly, one of the factors which has recently been found to account for a large part of this variability is microbiota (26–30). Large-scale testing of these hypotheses through self-tracking could contribute to the assessment of the complex metabolic response of the human body to different energy sources. This requires detailed records of food intake that includes nutritional information as well as eating times (31) and food portion sizes (32–34), all challenges that FoodRepo may help to overcome.

However, we highlight an important limitation of all food databases. Generally, the curators of such repositories cannot ensure the validity of the data reported by the producers on the nutrition facts labels. It is indeed well known in the literature that there might be large discrepancies between the reported nutrients and the actual food content, due to different factors, such as food pre-processing or the different industry standards (35–40). Therefore, all studies using databases such as the one presented here would do well to assess the validity of such data and ideally quantify the reporting errors, especially when using the reported data on nutritional values.
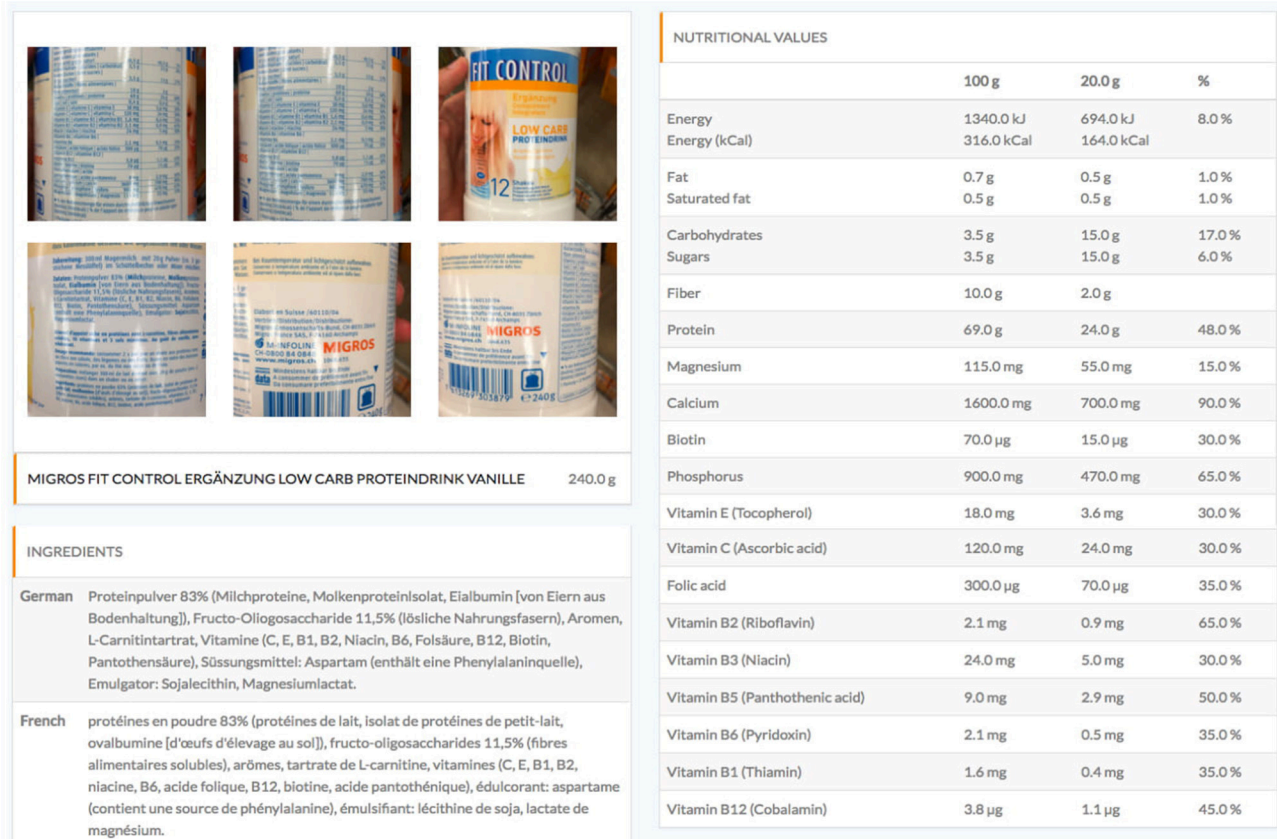
Analyses of the database evolution will give interesting indication on the dietary trends and on the overall modification of the nutritive quality of packaged food. Although the database itself does not inform on the buying frequency, the continuous introduction of specific products in the market and thus in the database can potentially indicate how retailers react to customer demands and changing dietary habits.
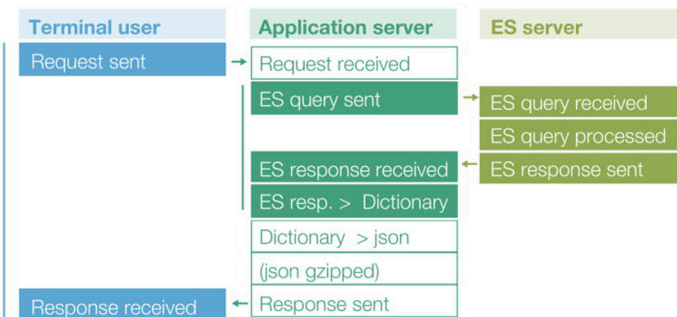
METHODS

The database building and maintenance process relies on the following steps: (i) collection of product pictures from local retailers, (ii) data extraction from the pictures, (iii) validation of the extracted data, and (iv) permanent storage in the database (**Figure 2**). For the initial build of the database, we designed a specific pipeline (bootstrap workflow, **Figure 2A**, which allowed us to validate the first 20,000 food products in a few months. Given the dynamic nature of our data and the cost of the bootstrap workflow, we designed a second pipeline (currently under development) which relies on the growing FoodRepo community. This workflow (community-based, **Figure 2B**) allows us to keep up with the new and seasonal products introduced to the market by the retail shops, as well as to ensure the scalability and self-sustainability of FoodRepo in the long run.

The bootstrap workflow (**Figure 2A**) consists of 3 main steps. The first step entailed a massive manual data collection from three large groceries stores in Switzerland upon approval from the shops (specifically Migros, Coop, and Lidl). We hired students to take pictures of all barcoded food items in retail shops located in the Lausanne area. To facilitate the data collection, we specifically designed a simple phone app with which students could scan the products' barcode and take pictures of the front and back of the package, the product's name, ingredients list, and nutrition facts. These pictures were then automatically uploaded to the database. At the end of this step, students had collected on average 4.4 pictures per item. The second step focused on the extraction of information contained in the pictures. Due to the presence of multi-language ingredients and the often wrinkled surfaces of item packaging, Optical Character Recognition (OCR) systems could not achieve a reliable accuracy. We therefore opted for a crowd-sourced solution and in particular we decided to recruit workers on Amazon Mechanical Turk (41) (AMT). AMT is a platform connecting *requesters* to *workers*, the latter being financially compensated to achieve tasks requiring human intelligence (HITs—Human Intelligence Tasks). Here, we designed a graphical user interface (GUI) allowing workers to transcribe the text they could read from product pictures. Specifically, the GUI presented text boxes where AMT workers

a. FoodRepo website - product example

MIGROS FIT CONTROL ERGÄNZUNG LOW CARB PROTEINDRINK VANILLE   240.0 g

INGREDIENTS

German Proteinpulver 83% (Milchproteine, Molkenproteinisolat, Eialbumin [von Eiern aus Bodenhaltung]), Fructo-Oliogosaccharide 11,5% (lösliche Nahrungsfasern), Aromen, L-Carnitintartrat, Vitamine (C, E, B1, B2, Niacin, B6, Folsäure, B12, Biotin, Pantothensäure), Süssungsmittel: Aspartam (enthält eine Phenylalaninquelle), Emulgator: Sojalecithin, Magnesiumlactat.

French protéines en poudre 83% (protéines de lait, isolat de protéines de petit-lait, ovalbumine [d'œufs d'élevage au sol]), fructo-oligosaccharides 11,5% (fibres alimentaires solubles), arômes, tartrate de L-carnitine, vitamines (C, E, B1, B2, niacine, B6, acide folique, B12, biotine, acide pantothénique), édulcorant: aspartame (contient une source de phénylalanine), émulsifiant: lécithine de soja, lactate de magnésium.

NUTRITIONAL VALUES

	100 g	20.0 g	%
Energy	1340.0 kJ	694.0 kJ	8.0 %
Energy (kCal)	316.0 kCal	164.0 kCal	
Fat	0.7 g	0.5 g	1.0 %
Saturated fat	0.5 g	0.5 g	1.0 %
Carbohydrates	3.5 g	15.0 g	17.0 %
Sugars	3.5 g	15.0 g	6.0 %
Fiber	10.0 g	2.0 g	
Protein	69.0 g	24.0 g	48.0 %
Magnesium	115.0 mg	55.0 mg	15.0 %
Calcium	1600.0 mg	700.0 mg	90.0 %
Biotin	70.0 µg	15.0 µg	30.0 %
Phosphorus	900.0 mg	470.0 mg	65.0 %
Vitamin E (Tocopherol)	18.0 mg	3.6 mg	30.0 %
Vitamin C (Ascorbic acid)	120.0 mg	24.0 mg	30.0 %
Folic acid	300.0 µg	70.0 µg	35.0 %
Vitamin B2 (Riboflavin)	2.1 mg	0.9 mg	65.0 %
Vitamin B3 (Niacin)	24.0 mg	5.0 mg	30.0 %
Vitamin B5 (Panthothenic acid)	9.0 mg	2.9 mg	50.0 %
Vitamin B6 (Pyridoxin)	2.1 mg	0.5 mg	35.0 %
Vitamin B1 (Thiamin)	1.6 mg	0.4 mg	35.0 %
Vitamin B12 (Cobalamin)	3.8 µg	1.1 µg	45.0 %

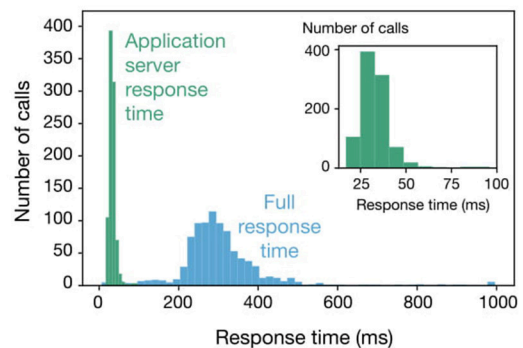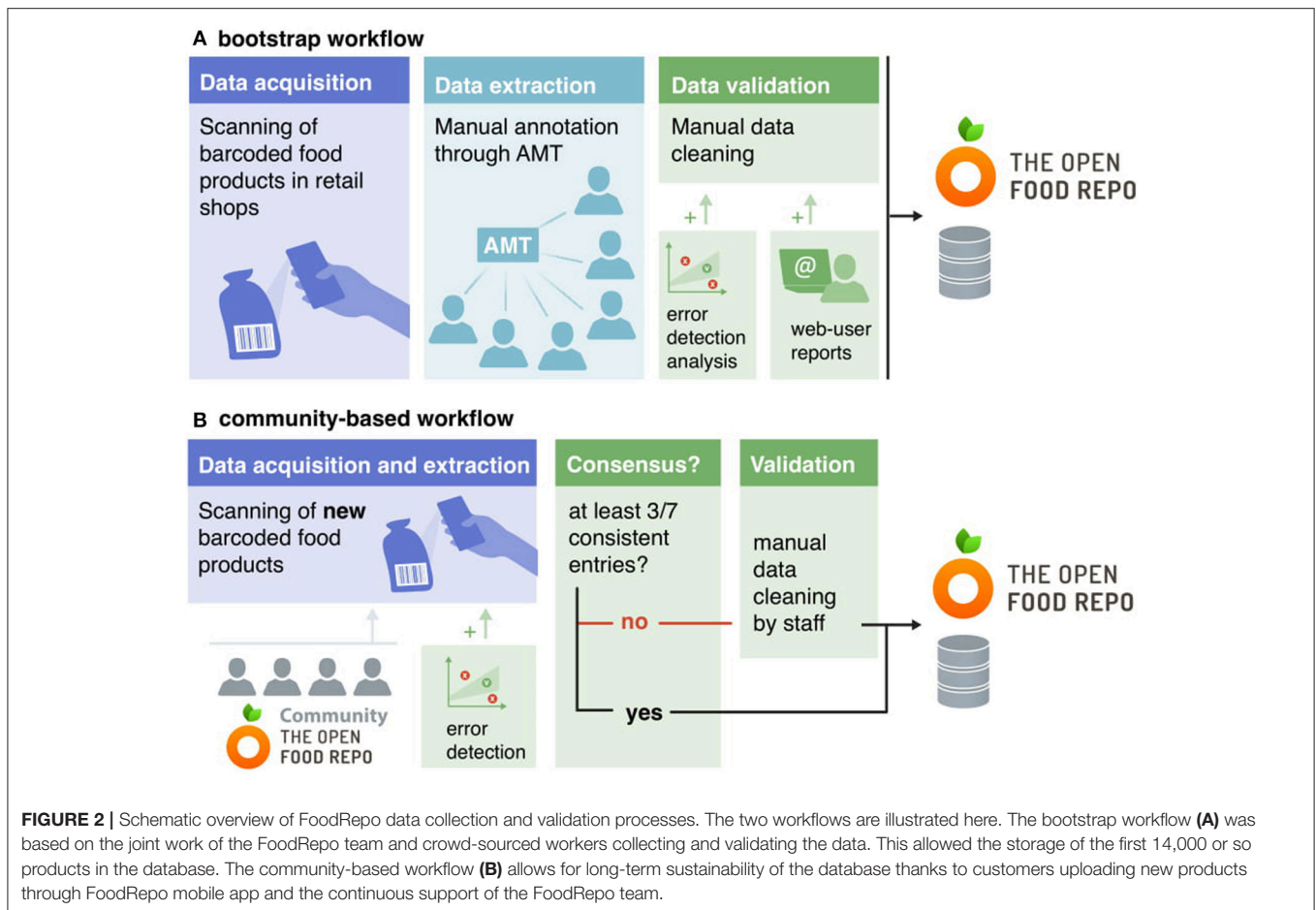## b. API workflow

c. API response times

FIGURE 1 | (A) Screenshot from the webpage of a product on the FoodRepo website. **(B)** Schematic representation of the pipeline behind our API. When a user or an application (left column) sends a call to the API, the request is handled by the server that hosts the API (middle column). This sends then a query to the server which hosts the FoodRepo database (right column), where the query is handled by the Elastic Search engine. The data is returned to the API server which performs final formatting before giving it back to the user or the application. **(C)** Distribution of API response times, color-coded according to different sections of the back-end pipeline, as shown in **(B)**. In green (main plot and inset) the response-times of the Elastic Search server to the application server; in blue the full time needed for a user to have the data after a call to our API.

provided the product name, nutritional values (in a table format) and ingredients, in every language present on the label (German and/or French for almost all items; Italian and/or English in addition for some products). Three different HITs were set up: one for nutrients, one for product name and one for ingredients. For the last two, we set up qualification rounds for AMT workers as their transcription involved some language skills. AMT workers could choose to either enter from scratch the information they saw on the pictures, or to approve/modify the suggestions given by an OCR (42) system. At the end of the second step, all annotated products were uploaded into the database, flagged as ready for validation.

FIGURE 2 | Schematic overview of FoodRepo data collection and validation processes. The two workflows are illustrated here. The bootstrap workflow **(A)** was based on the joint work of the FoodRepo team and crowd-sourced workers collecting and validating the data. This allowed the storage of the first 14,000 or so products in the database. The community-based workflow **(B)** allows for long-term sustainability of the database thanks to customers uploading new products through FoodRepo mobile app and the continuous support of the FoodRepo team.

The third step was thus dedicated to data validation, which was based on extensive manual checking by the FoodRepo team, and was additionally informed by manual reports from visitors to the FoodRepo website and with error-detection analyses of nutritional values. Such online reports are encouraged by the presence of a "`report an issue`" button on each product web-page, which prompts a visitor to file an issue when spotting a potential error. Details about the error-detection analyses are given in the Technical Validation section. Before the final validation of the data, the FoodRepo team as well as students manually checked all products thoroughly.

The community-based workflow (**Figure 2B**) is similar to the bootstrap workflow, but instead of counting on AMT workers, it relies on the growing FoodRepo community. As new products become available in retail shops, FoodRepo users can submit them by uploading the corresponding package pictures, using the FoodRepo smartphone app. Currently, the information extraction is still performed by the FoodRepo team, but additional features are being implemented in the app, which will allow users to directly type the product details contained on the package. Before user-provided information is permanently stored in the FoodRepo database, consistent entries will need to be submitted by at least three different FoodRepo users. If such consensus will not be reached after seven independent submissions (i.e., there are still less than three consistent entries),

the item will be manually analyzed by the FoodRepo team for definitive validation and inclusion into the database.

This procedure will ensure minimal intervention from our team, while still guaranteeing the reliability of the data. The FoodRepo team is currently fostering the development of an active community through which the continuity of FoodRepo is assured, and which will likely accelerate the birth of independent exploitations of the database, from both public and private partners.

DATA RECORDS

All FoodRepo data are stored in a PostgreSQL (43) database, physically hosted on a server in Ireland. For a quick overview of the dataset, a database dump can be downloaded from the dedicated folder in our API repository (44). However, these dumps are not generated regularly, and we strongly encourage the use of the API which delivers up-to-date information. For each product, which comes with a unique numerical identifier, the database contains pictures of the item as found in the shop (usually between three to seven `.jpg` files), together with the main information presented on the package, i.e., the product name, nutritional values, ingredients list, barcode, and country of origin. The database holds as well the dates of the creation and last modification of the related item in

TABLE 1 | Sample product from the FoodRepo database with its values for the most relevant fields.

Fields	Sample
Product ID	972
Barcode	7611654884033
Name	Chocolat au lait aux noisettes
Quantity	150
Units	g
Portion quantity	30
Portion unit	g
Alcohol by volume	0
Origin	Switzerland
Ingredients	(FR) Sucre de canne brut* (Paraguay), noisettes caramélisées et grillées 22% (noisettes [Turquie], sucre de canne brut* [Paraguay], eau), lait entier en poudre (Suisse), beurre de cacao* (République dominicaine), cacao en pâte* (Pérou), pâte de noisette 4.5% (Turquie), gousses de vanille*. Teneur en cacao du chocolat: 32% minimum. * Ingrédients conformes aux standards du commerce équitable Fairtrade. 58.6% du poids total. Dont sucre et produits à base de cacao avec bilan de masse. Tous les ingrédients agricoles sont issus de l'agriculture biologique. Allergies: Contient des noisettes, du lait. Peut contenir des traces d'amandes, de soja. (DE) Rohrohrzucker* (Paraguay), karamellisierte geröstete Haselnüsse 22% (Haselnüsse [Türkei], Rohrohrzucker [Paraguay], Wasser), Vollmilchpulver (Schweiz), Kakaobutter* (Dominikanische Republik), Kakaomasse* (Peru), Haselnusspaste 4.5% (Türkei), Vanilleschoten*, Kakaobestandteil in der Schokoladenmasse: mindestens 32%, * Nach Fairtrade-Standards gehandelte Zutaten. Gesamtanteil 58.6%. Davon Kakaoerzeugnisse und Zuckerarten mit Mengenausgleich. Alle landwirtschaftlichen Zutaten stammen aus biologischem Anbau. Allergie: Enthält Haselnuss, Milch. Kann spuren von Mandeln, Soja enthalten.
Nutrients (per 100 g)	Energy 2410.0 kJ; Energy (kCal) 577.0 kCal; Fat 40 g; Saturated fat 16 g; Carbohydrates 43 g; Sugars 42 g; Fiber 4 g; Protein 10 g; Salt 0.2 g
Created at	2016-05-31, 17:54:07
Updated at	2017-11-16, 10:13:31
Pictures	Url to the front picture of the sample product: e.g., https://goo.gl/PyjjNa

While here we only provide the link to the front image of the product, an API call would provide the links to all pictures available for the requested products. A complete description of the fields provided by the API is available in the API documentation, on the project's GitHub repository.

the database (see **Table 1**). The programmatic access to the database is allowed by an API, described in the section Usage Notes.

TECHNICAL VALIDATION

As described in the Methods section, during the bootstrap stage (**Figure 2A**) the final validation was performed manually by the FoodRepo team, while in the community workflow (**Figure 2B**), the accuracy of the data is ensured by the consensus test (the FoodRepo team intervenes only if fewer than three matches are achieved after the uploads of the same product by seven different users). We highlight here that FoodRepo strictly reflects the information printed on products packages, even when suspicious values are present on the labels. All validation processes have thus been set-up to detect transcription errors.

Within this rationale, computational analyses were implemented for the detection of outliers, in particular regarding the nutritional values. These tests reflect basic constraints, such as the mass upper-limit:

$$p + f + c \leq 100 \tag{1}$$

where p, f, c are respectively the product's protein, fat and carbohydrates concentrations expressed in grams per 100 g of product. From Equation (1), one can also derive other linear inequalities for a single nutrient or couples of nutrients, namely $p + f \leq 100$, $p + c \leq 100$, and $c + f \leq 100$. These simple tests

allowed us to detect transcription errors in earlier versions of the database, as illustrated by the outliers in **Figure 3A** which shows the distribution of products in the fat-carbohydrates space with the joint mass boundary.

Similarly, other typos could be spotted by checking that the concentration of a subclass of nutrient is smaller than the one of the parent-class. This is the case for instance of sugars vs. carbohydrates, or saturated-fat vs. fat, shown in **Figure 3B**.
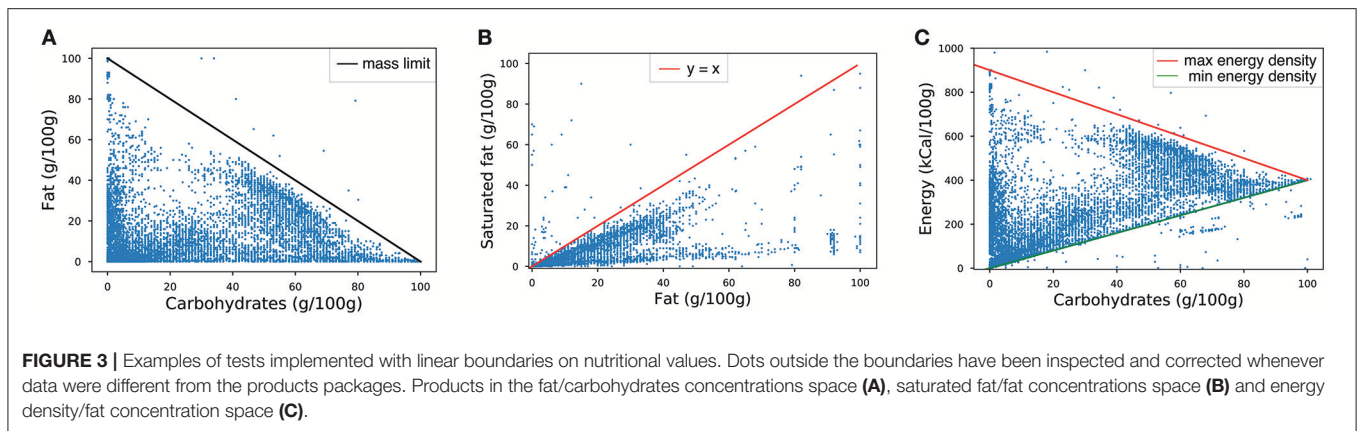
Another simple relation that helps check products' nutrition facts can be derived from the standard approximation of energy density based on nutrients composition (45):

$$E \sim 4p + 9f + 4c, \tag{2}$$

where the product's energy content E is expressed in $kCal/100\,g$. Combining expressions 1 and 2 provides upper and lower boundaries for the energy content (for example **Figure 3C**). In this case however, not all dots that fall outside the boundaries were due to typos in transcription. Indeed, the approximation in Equation (2) does not take into account the different contribution to energy of complex carbohydrates such as polyols, which account for less than 4 $kCal/g$. This is why products such as candies and chewing gums would fall below the energy boundaries.

USAGE NOTES

In order to facilitate the access to the database, we built an openly accessible API. Any terminal user, including third party

FIGURE 3 | Examples of tests implemented with linear boundaries on nutritional values. Dots outside the boundaries have been inspected and corrected whenever data were different from the products packages. Products in the fat/carbohydrates concentrations space **(A)**, saturated fat/fat concentrations space **(B)** and energy density/fat concentration space **(C)**.

apps or services, can send API requests to retrieve specific data. The API pipeline is illustrated in **Figure 1B**. User's requests are handled on an application server, where an Elastic Search (ES) application handles the queries on another cloud computing service, based in Ireland. The ES response is then returned to the user after JSON formatting and compression (on demand). We checked that handling the request between the two servers does not critically compromise the total user-response time. We run series of single-page API calls, every 6 h, over a week, in order to measure the full response-time and the application server response-time. We observed that the latter was consistently fast across all experiments (in the range of 20–50 ms) and that the bottleneck was rather the transmission between the terminal user and the application server (the average full response time was about 250 ms—see **Figure 1C**).

For a quick introduction to the API endpoints, users are welcome to try them out on the API Playground page (46). Furthermore, on the project's GitHub repository, one can also find usage cases (47) in Python, Ruby, Curl and JavaScript, as well as examples of complex queries which include fuzzy searches (48). When fetching a large amount of data, we suggest using the option of compressed data[1] and the possibility to include/exclude specific fields of each product [see for details the API documentation (46)]. In this way, one could reduce the response payload size by up to a factor of 10.

We remind readers that all contents (other than computer software) made available by FoodRepo on its websites, apps or services are licensed under the Creative Commons Attribution 4.0 International License. We however would like to highlight the fact that product images may contain copyrighted data such as brand logos.

AUTHOR CONTRIBUTIONS

GL performed the descriptive and validation analysis of the dataset. YJ built the FoodRepo database, website, API and AMT HITs. DK maintained the API, coordinated the manual data validation and built the framework for the FoodRepo community. GL, LS, and MS wrote the manuscript. MS initiated and supervised the project.

NOMENCLATURE

- API: Application Programming Interface—an set of tools and methods that allow to types of software to communicate. The FoodRepo API allows other applications to get and use the data.
- CC-BY-4: Creative-Commons public license, with the "Attribution" term. It implies that anyone is free to share and transform the content of FoodRepo, even for commercial purposes, with the obligation to properly give credit to FoodRepo, and to display any modification without claiming direct endorsement from FoodRepo. For a detailed description, see the license text at https://creativecommons.org/licenses/by/4.0/
- OCR: Optical Character Recognition—tools that allow for automatic conversion of text contained in images to machine-readable formats.
- AMT: Amazon Mechanical Turk—web platform providing a marketplace, where workers perform tasks set up by requesters, usually in exchange of money.
- HIT: Human Intelligence Task—task related performed by workers in crowd-sourcing platform, such as AMT.
- PostgreSQL: A popular and freely available relational database.
- JSON: a JavaScript-based file format commonly used for browser-server data exchange.
- Elastic Search: a very popular open-source search-engine.

---

[1]This can be done by simply setting in the request header: `Accept-Encoding: gzip`

REFERENCES

1. WHO. *Diabetes*. Available online at: http://www.who.int/mediacentre/factsheets/fs312/en/.

2. WHO. *Obesity and Overweight*. Available online at: http://www.who.int/mediacentre/factsheets/fs311/en/.

3. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. *Adv Nutr Int Rev J*. (2015) 6:5–18. doi: 10.3945/an.114.007492

4. Archer E, Pavela G, Lavie CJ. The inadmissibility of what we eat in America and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines. *Mayo Clin Proc*. (2015) 90:911–26. doi: 10.1016/j.mayocp.2015.04.009

5. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing current criticism regarding the value of self-report dietary data, 2. *J Nutr*. (2015) 145:2639–45. doi: 10.3945/jn.115.219634

6. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. *Nutrition* (2014) 30:1257–66. doi: 10.1016/j.nut.2014.02.020

7. Chae J, Woo I, Kim S, Maciejewski R, Zhu F, Delp EJ, et al. Volume estimation using food specific shape templates in mobile image-based dietary assessment. In: *Proceedings of SPIE. Vol. 7873. NIH Public Access*. San Francisco, CA (2011). p. 78730K.

8. Kong F, Tan J. DietCam: automatic dietary assessment with mobile camera phones. *Pervas Mob Comput*. (2012) 8:147–63. doi: 10.1016/j.pmcj.2011.07.003

9. Lee CD, Chae J, Schap TE, Kerr DA, Delp EJ, Ebert DS, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. *J Diab Sci Technol*. (2012) 6:428–34. doi: 10.1177/193229681200600231

10. Dibiano R, Gunturk BK, Martin CK. Food image analysis for measuring food intake in free living conditions. In: *Medical Imaging: Image Processing*. Orlando, FL (2013). p. 86693N.

11. Zhu F, Bosch M, Khanna N, Boushey CJ, Delp EJ. Multilevel segmentation for food classification in dietary assessment. In: *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*. Dubrovnik: IEEE (2011). p. 337–42.

12. Zhu F, Bosch M, Woo I, Kim S, Boushey CJ, Ebert DS, et al. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE J Select Top Signal Proces*. (2010) 4:756–66. doi: 10.1109/JSTSP.2010.2051471

13. Siek KA, Connelly KH, Rogers Y, Rohwer P, Lambert D, Welch JL. When do we eat? An evaluation of food items input into an electronic food monitoring application. In: *Pervasive Health Conference and Workshops, 2006*. Innsbruck: IEEE (2006). p. 1–10.

14. Eyles H, Jiang Y, Mhurchu CN. Use of household supermarket sales data to estimate nutrient intakes: a comparison with repeat 24-hour dietary recalls. *J Am Diet Assoc*. (2010) 110:106–10. doi: 10.1016/j.jada.2009.10.005

15. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. *Am J Prevent Med*. (2013) 45:576–82. doi: 10.1016/j.amepre.2013.04.025

16. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. *JMIR mHealth uHealth*. (2014) 2:e37. doi: 10.2196/mhealth.3230

17. Stephens J, Allen JK, Himmelfarb CRD. "Smart" coaching to promote physical activity, diet change, and cardiovascular health. *J Cardiovasc Nurs*. (2011) 26:282. doi: 10.1097/JCN.0b013e31821ddd76

18. Tsai CC, Lee G, Raab F, Norman GJ, Sohn T, Griswold WG, et al. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. *Mob Netw Appl*. (2007) 12:173–84. doi: 10.1007/s11036-007-0014-4

19. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. *Am J Prev Med*. (2013) 45:583–9. doi: 10.1016/j.amepre.2013.07.005

20. Open Food Facts. Available online at: https://world.openfoodfacts.org/.

21. USDA Food Composition Database. Available online at: https://ndb.nal.usda.gov/ndb/.

22. Application NutriScan. Available online at: https://www.bonasavoir.ch/nutriscan.

23. Pfeiffer E. The glucose sensor: the missing link in diabetes therapy. *Horm Metab Res Suppl Ser*. (1989) 24:154–64.

24. Aljasem LI, Peyrot M, Wissow L, Rubin RR. The impact of barriers and self-efficacy on self-care behaviors in type 2 diabetes. *Diab Educ*. (2001) 27:393–404. doi: 10.1177/014572170102700309

25. Elenko E, Underwood L, Zohar D. Defining digital medicine. *Nat Biotechnol*. (2015) 33:456–61. doi: 10.1038/nbt.3222

26. Griffin NW, Ahern PP, Cheng J, Heath AC, Ilkayeva O, Newgard CB, et al. Prior dietary practices and connections to a human gut microbial metacommunity alter responses to diet interventions. *Cell Host Microbe*. (2017) 21:84–96. doi: 10.1016/j.chom.2016.12.006

27. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* (2006) 444:1027–31. doi: 10.1038/nature05414

28. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* (2013) 500:541–6. doi: 10.1038/nature12506

29. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. *Cell* (2015) 163:1079–94. doi: 10.1016/j.cell.2015.11.001

30. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* (2016) 535:376–81. doi: 10.1038/nature18646

31. Scheer FA, Hilton MF, Mantzoros CS, Shea SA. Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc Natl Acad Sci USA*. (2009) 106:4453–8. doi: 10.1073/pnas.0808180106

32. Ello-Martin JA, Ledikwe JH, Rolls BJ. The influence of food portion size and energy density on energy intake: implications for weight management. *Am J Clin Nutr*. (2005) 82:236S–41S. doi: 10.1093/ajcn/82.1.236S

33. Ledikwe JH, Ello-Martin JA, Rolls BJ. Portion sizes and the obesity epidemic. *J Nutr*. (2005) 135:905–9. doi: 10.1093/jn/135.4.905

34. Young LR, Nestle M. The contribution of expanding portion sizes to the US obesity epidemic. *Am J Public Health* (2002) 92:246–9. doi: 10.2105/AJPH.92.2.246

35. Ng SW, Popkin BM. Monitoring foods and nutrients sold and consumed in the United States: dynamics and challenges. *J Acad Nutr Diet*. (2012) 112:41–5. doi: 10.1016/j.jada.2011.09.015

36. Ng S, Dunford E. Complexities and opportunities in monitoring and evaluating US and global changes by the food industry. *Obes Rev*. (2013) 14:29–41. doi: 10.1111/obr.12095

37. Ahuja JK, Lemar L, Goldman JD, Moshfegh AJ. The impact of revising fats and oils data in the US Food and Nutrient Database for Dietary Studies. *J Food Compos Anal*. (2009) 22:S63–7. doi: 10.1016/j.jfca.2009.02.005

38. Merchant AT, Dehghan M. Food composition database development for between country comparisons. *Nutr J*. (2006) 5:2. doi: 10.1186/1475-2891-5-2

39. Phillips KM, Patterson KY, Rasor AS, Exler J, Haytowitz DB, Holden JM, et al. Quality-control materials in the USDA national food and nutrient analysis program (NFNAP). *Anal Bioanal Chem*. (2006) 384:1341–55. doi: 10.1007/s00216-005-0294-0

40. Deharveng G, Charrondiere U, Slimani N, Southgate D, Riboli E. Comparison of nutrients in the food composition tables available in the nine European countries participating in EPIC. *Eur J Clin Nutr*. (1999) 53:60. doi: 10.1038/sj.ejcn.1600677

41. Amazon Mechanical Turk. Available online at: https://www.mturk.com/.

42. Text Recognition API Overview. *Google Developers*. Available online at: https://developers.google.com/vision/text-overview.

43. PostgreSQL. The World's Most Advanced Open Source Database. Available online at: https://www.postgresql.org/.

44. FoodRepo Database Dumps. Available online at: https://github.com/salathegroup/foodrepo_api/tree/master/data.

45. COUNCIL DIRECTIVE of 24 September 1990 on Nutrition Labelling for Foodstuffs. Available online at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1990L0496:20081211:EN:PDF.
46. OpenFood API Documentation. Available online at: https://www.foodrepo.org/api-docs/swaggers/v3.
47. OpenFood API GitHub Repository. Available online at: https://github.com/salathegroup/foodrepo_api/tree/master/v3/code.
48. Elasticsearch Queries Example. Available online at: https://github.com/salathegroup/foodrepo_api/blob/master/v3/code/meta/es_sample_queries_product.md.
49. Kristian Gerhard Jebsen Foundation. Available online at: http://www.kgjf.org/.