



USDA Global Branded Food Products Database Defining Data Quality

March 2019

The utility and benefit of any database is found in the quantity and quality of the data it captures. The partners in the Public-Private Partnership to deliver the USDA Global Branded Food Products Database (BFPD), work with food and beverage companies to deliver quality data for public transparency and use by the research community and others.

What is Data Quality?

Data quality is critical to ensure the nutrient composition of branded products that is shared is complete, standardized, time stamped, and accurate. These factors are so central to the use and exchange of data they have been written into the operational standards and “validation rules” of the partners that provide the data, Label Insight and 1WorldSync, leveraging GS1 Global Standards.

The data partners collect and assesses product data provided by the product owner. They assess the data to ensure it is properly captured in a complete, consistent, and standardized manner to be used in the USDA Global Branded Food Products Database. This is achieved through rigorous quality assurance processes, including use of the following measures:

1. *“Hard” and “Soft” Data-entry Validations:* “Hard” Validations are strict restrictions on the type of data entered. For example, a product’s net weight must be entered as an exact numerical value and accompanied by a unit of measure, e.g., oz, gal, lbs. Where hard validations are not met, further data entry is forbidden. “Soft” Validations highlight information that is likely incorrect based on labeling regulations or entry. For example, the absence of a required nutrient(s) and an invalid Global Trade Item Number (GTIN – number contained in the U.P.C. barcode) are considered soft validations. Soft validations will allow data entry to continue to occur.
2. *Outliers and Profiling:* This accounts for data that is provided and falls outside of a typical product range (or standard deviation) and is flagged for review. For example, a chocolate covered cookie with less than 40 grams or greater than 90 grams of total carbohydrates per 100 grams would fall significantly outside a typical chocolate-covered cookie product and would merit a flag and review.
3. *Random Sampling:* In addition to the above quality assurance measures, each month thousands of products are pulled and sampled for overall data accuracy.
4. *Timeliness:* As products are reformulated and change over time, data is resubmitted and reanalyzed for these products to ensure a continuous flow of new and updated product data.

Product Dates

As stated above, each product reflected in the USDA Global Branded Food Products Database carries a date in addition to which data partner was used for the submission of product data. These dates are reflected and defined as:

- **Date Available:** This is the date when the product record was available for inclusion in the database.
- **Date Published:** This is the actual date the product record was added to the USDA Global Branded Food Products Database.



- **Date Modified:** This date reflects when the product data was last modified by the data provider, i.e., the manufacturer.

As the goal of the public-private partnership is to enhance public health and the sharing of open data by complementing USDA FoodData Central with nutrient composition and ingredient information of branded foods and private label data provided by the food industry, the partners work collaboratively to ensure the right validations are in place to maintain quality data. As the database continues to evolve, the partners may add to the current data quality processes to further enhance the product records.