Data fields

A **data field** is a structured information that has at least a specific usage. For example, the "product name" field allows us to easily recognize the main name printed on the packaging.

Open Food Facts manages different kinds of data fields:

1. fields that can be completed by users, such as the name of the product, the brand, etc.
2. fields that are always computed by machines such as the name of the contributor or the date of the contribution
3. fields that are sometimes computed based on other fields, such as the Nutri-Score, the Nova score, etc.

This page deals only with "Fields completed by users". All these fields can be entered or modified by hand by the users.

Product characteristics

Product name

The product name is the main name **printed on the packaging**. It can be a registered trademark such as Nutella. This data is important and useful as it's one of the most used data.

If it's not a part of the name, it shouldn't contain the number of portions or a quantity: bad examples are "1 Onglet", "10 Burgers", "1L Sirop cerise"; "100% Cacao"; it shouldn't contain registered trademark symbols ®, HTML code such as "; it shouldn't be in capital letters except if they are used on the product; it shouldn't contain brands except if it's included in the name ("Kinder Bueno" is good while "Kronembourg 1664", "Stella" or "Vodka Smirnoff" are not); it shouldn't contain price.

At the beginning of 2020, more than 95% of Open Food Facts products have a product name:

- the ones who have a name (https://world.openfoodfacts.org/state/product-name-completed)
- and the ones who doesn't (https://world.openfoodfacts.org/state/product-name-to-be-completed)

The product name shouldn't include any other information such as the brand of the product, the weight, etc.

In the database, the technical name for this field is `product_name`.

Good examples

- `Nesquick` (link (https://world.openfoodfacts.org/product/3033710065967/nesquik-poudre-cacaotee-boite-nestl e))
- `111 Filini` (link (https://world.openfoodfacts.org/product/8008845071119/111-filini-maltagliati))
- `Yazoo strawberry` (link (https://world.openfoodfacts.org/product/5410438044845/yazoo-strawberry)) => in this example, "strawberry" has a lower font size than "Yazoo" but the flavour here distinguishes between other products, so we can consider it is part of the name

Bad examples

- <u>Petit déjeuner</u> `Nesquick` => you don't have to explain, just put the name from the packaging
- `Nutella` <u>by Ferrero</u> => you shouldn't fill the brand here, there's a field for that :)
- `Nesquick`<u>®</u> => Don't use symbols ®, ™, © or similar in product name data field or even in other fields.

Common name

The common name defines the product. It is the name used when you don't want or can't use the product name. This is the place where you say `Cocoa and hazelnuts spreads` instead of `Nutella`. This name is very useful for our AI (artificial intelligence): it helps to guess the category of the product.

The common name might be equivalent to product category but sometimes not [examples].

In the database, the technical name for this field is `generic_name`.

Example in Coca-Cola Cherry: **"Sparkling Cherry Flavour Soft Drink with Vegetables Extracts"**. This info should be in "Common name" filed.

Quantity

This is the quantity of the product, with the corresponding number of portions or unit. The best way to fill it is to enter the value as indicated on the product. Don't forget the units! If we can deduce the quantity in grams it can be used to calculate some things such as the carbon impact.

Examples:

- `230g`
- `230 g`
- 6 (for 6 eggs)
- `3 x 150g` (for a product with 3 boxes, each of 150g)

A complete wiki page is dedicated to <u>products quantities</u>.

In the database, the technical name for this field is `quantity`. If the system recognizes the product's quantity, it stores it as a numeric value in the `product_quantity` field: e.g. `quantity = 10 * 9.8 ml` will lead to `product_quantity` = 98.

See <u>quantities that are not recognized (https://world.openfoodfacts.org/data-quality/quantity-not-recognized)</u>.

See issues related to `quantity` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+i
s%3Aopen+quantity+label%3Aquantity).

Packaging

This is the packaging of the product. Multiple values are allowed. There is no taxonomy for this field, so you can enter anything you find relevant including:

- the substance of the packaging: glass, metal, plastic, etc.
- the shape: bottle, can, etc.

You can write it in your own language. Don't hesitate to add as much data as you find relevant.

Two draft taxonomies has been proposed, don't hesitate to comment or help:

- a global taxonomy
- and a packaging logos taxonomy

In the database, there's two fields related to packaging:

- `packaging`: the list of values entered in the packaging field; e.g.: `Bottle,Glass`
- `packaging_tags`: an array of different packaging tags

See issues related to `packaging` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+i
s%3Aopen+packaging+label%3Apackaging).

Brands

This is the brands of the product. The main brand, generally clearly displayed on the front pack, should be entered first. A product can have other brands:

- when a product is a brand sold by a big company: `Actimel` is sold by `Danone`, see
  https://world.openfoodfacts.org/product/4009700036810/actimel-granatapfel
- when a product is sold with its brand translated in two languages: `Nature Valley` is sometimes written `Val Nature`; see https://world.openfoodfacts.org/product/0065633280267/barre-granola-erable-et-cassonade-nature-valley

There's no taxonomy for brands for the moment, so just do your best and don't waste too much time to enter brands.

A complete wiki page is dedicated to brands.

In the database, this field is called `brands`.

See: issues related to `brands` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3
Aopen+ingredients+label%3Abrands).

Categories

[to be completed]

This field list the specific category of a product and its related parents. **This field is very important**: it needs to be completed for Nutri-Score computation, as this computation makes differences between types of products (beverages, ...).

→ Indicate only the most specific category. "Parents" categories will be automatically added.

Examples: `Sardines in olive oil`, `Orange juice from concentrate`

See also:

- our categories taxonomy on our repository (https://github.com/openfoodfacts/openfoodfacts-server/blob/master/taxonomies/categories.txt).
- issues related to `categories` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+category+label%3Acategories).

Labels

Here we put any characteristic of the product which is factual and different from the other fields: organic, vegetarian, religious, etc.

Provenance: many contributors add the origins here: eg, `Made in Belgium`, `Produced in Brittany`, ...

Religious labels: namely halal and kosher labels.

Rating labels when they are displayed on the product (even if it's not the same as we compute): Nutri-Score (6+ European countries), Five Start Rating system (Australia, New-Zealand), etc.

[to be completed]

"Labels, certifications, awards" on the website.

Example: Organic.

See the Labels page for more informations.

See also:

- our labels taxonomy on our repository (https://github.com/openfoodfacts/openfoodfacts-server/blob/master/taxonomies/labels.txt).
- issues related to `labels` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+category+label%3Alabels).

Manufacturing or processing places

[to be completed]

This field lists the places where the product has been manufactured or processed.

It is often displayed on the packaging by the traditional "Made in XXX". The packaging can also mention the region or even the address of the place. The origins of the ingredients have nothing to do with this: they can come from all over the world without any impact on the place where they have been processed.

You should enter this data like an address: the most precise information first and the most general at the end.

Eg.

- France
- Bretagne, France
- Laiterie de Saint-Denis-de-l'Hôtel (LSDH) - 10 Route de l'Aérodrome - Les Grandes Beaugines - 45550 Saint-Denis-de-l'Hôtel, Loiret, Centre-Val de Loire, France

In the database, this field is called `manufacturing_places`; it has a normalized version (small caps, spaces removed) under the name `manufacturing_places_tags`.

EMB code

[to be completed]

This field is dedicated for various codes related to packaging marks, identification marks or health marks:

- EC identification and health marks (https://en.wikipedia.org/wiki/EC_identification_and_health_marks) in use in the European Community to identify food producers or food packagers. Example: FR `72.264.002` CE.
- EMB codes (https://fr.wikipedia.org/wiki/EMB_(code)) (fr): packaging codes in use in France. Example: EMB `72264`.

Open Food Facts gathered 25,000+ codes (https://world.openfoodfacts.org/packager-codes) as of 2021-08-18. These codes allow to localize the places on a map: https://world.openfoodfacts.org/packager-code/fr-72-264-002-ce

We use it to produce the world map of products: https://cestemballepresdechezvous.fr/ (fr) and https://madenear.me/ (en).

Best before date (expiration date)

The expiration date is a way to track product changes over time and to identify the most recent version. It's a data for manual usages. At this moment (2020-03), Open Food Facts apps and website don't make any usage of this field. An issue is open to throw off very old products in averages (https://github.com/openfoodfacts/openfoodfacts-server/issues/76), it could be useful for it.

Be aware that, for the moment, this field is NOT normalized, so it probably contains dates in various formats that can be ambiguous (31/12/2019, 12/31/2019, 13 mai 2018, etc.). Also, the meaning of "best before", "expiry date", "use-by" notions might be different from different countries. You can help us to gather information about it in your own country, participating to the page Dates on the products.

It is possible to see:

- how many products do have an expiration date (https://world.openfoodfacts.org/state/expiration-date-completed) (a bit more than 10% at the beginning of 2020)
- and how many don't (https://world.openfoodfacts.org/state/expiration-date-to-be-completed)

In the database and in Product Opener software, the technical name for this field is `expiration_date`.

Countries where sold

This field contains all the countries where the product is widely available (not including stores specialising in foreign products). If the field contains France, the product will be listed on the https://fr.openfoodfacts.org/ website.

The list of all existing countries can be found here: https://world.openfoodfacts.org/countries

In 2021, the United Nation recognize 193 countries (https://en.wikipedia.org/wiki/List_of_sovereign_states) but Open Food Facts also recognize other territories, such as overseas regions/territories such as French Guiana, Guadeloupe, French Polynesia, etc. This field is taxonomized (source code (https://github.com/openfoodfacts/openfoodfacts-server/blob/main/taxonomies/countries.txt)). Though, Open Food Facts accepts all values and some people or some bogus tools enter other values, which leads to have bad values.

The users or third party apps should enter:

- either the name of the country under the same locale as the web site:
    - on fr.openfoodfacts.org (French locale) you should enter `Belgique`,
    - while on world.openfoodfacts.org (English locale) you should enter `Belgium`
- either the name of the country with the locale as a prefix; e. g. `en:Belgium` or `fr:Belgique`.

In the database, this field is called `countries`.

This field allows to compute the `countries_tags` field, the normalized version of countries, which allows to show the name of a country in the desired language, thanks to countries taxonomy. Eg: when `Italy` is entered from a page in english (say, world.openfoodfacts.org), `countries_tags` becomes `en:italy`.

Ingredients

This field lists the ingredients of the product. This field is **one of the most important** as it is used for:

- Nova calculation
- Nutri-Score calculation: it is used to calculate the proportion of fruit and nuts
- evaluation of some food preferences such as vegetarians, vegans, etc.
- automatic identification of allergens (both substances and traces, see below)
- etc.

Thanks to automatic Optical Characters Recognition (OCR), this field can be filled by softwares. But OCR is not always good, and you should always verify the result.

Ingredients analysis also produce an array of all ingredients which allows translation in other languages. To allow analysis and translation, Open Food Facts community has build a taxonomy (https://github.com/openfoodfacts/openfoodfacts-server/blob/master/taxonomies/ingredients.txt). You can contribute to it.

In the database, this field is called `ingredients_[country code]`.

See issues related to `ingredients` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+ingredients+label%3Aingredients).

Substances or products causing allergies or intolerances

The substances are ingredients that are actually in the product, which could cause common allergies. This field can be filled by hand, but is also completed by automatic ingredients analysis.

Examples:

- `Milk`
- `Gluten`
- `Nuts`

In the database, this field is an array of tags called `allergens_tags`.

See issues related to `allergens_tags` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+traces+label%3Aallergens).

Traces

The traces are ingredients which are not used for the product itself but lay in the factory or the production process: the product might contain traces of these ingredients. Traces are really important if you are allergic.

This field can be filled by hand, but is also completed by automatic ingredients analysis.

Examples:

- `Milk`
- `Gluten`
- `Nuts`

In the database, the technical name for this field is `traces`.

See issues related to `traces` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+traces+label%3Atraces).

Nutrition facts

Nutrition facts not specified

Sometimes nutrition facts are not specified on the packaging or on a document given with the product. In this case, and only in this case, you have to fill the checkbox called `Nutrition facts are not specified on the product`.

It's often (if not always) the case for aromatic herbs (https://world.openfoodfacts.org/category/aromatic-herbs).

In the database, the technical name for this field is `no_nutrition_data`.

Serving size

Serving size has a specific goal: to let Open Food Facts app make a proportional calculation of each nutrient per serving size. If a candy's weight is 5 g, it can be chosen as the serving size: if these candies has 66 g of sugar per 100 g, it has about 3 g per candy. Allowed units (https://github.com/openfoodfacts/openfoodfacts-server/blob/f25308b7d47255be83210f699f897cba87c9517f/lib/ProductOpener/Food.pm#L3835) are: `kg, g, mg, µg, oz, l, dl, cl, ml, fl.oz, fl oz, г, мг, кг, л, дл, кл, мл, 毫克, 公斤, 毫升, 公升, 吨`.

`grammes`, `liter`, etc., are NOT recognized.

Decimals can be written with a comma (`,`) or a point (`.`).

In the database and in Product Opener software, the technical name for this field is `serving_size`. Based on `serving_size`, Open Food Facts computes a `serving_quantity` float number for 100g or 100ml. The `serving_quantity` can be found in the API or the data exports.

See issues related to `serving_size` (https://github.com/openfoodfacts/openfoodfacts-server/issues?q=is%3Aissue+is%3Aopen+serving+label%3Aportions).

Good examples

- `60 g` (preferred, for readability reasons)
- `30g`
- `35G`
- `90 ml`

- 1L

Possible examples (while not recommended)

- `cookie 25g`
- `One Slice (50g)`
- `97 g (0.5 cup)`

Bad examples

- `30 gr` => `gr` is not a correct unit
- `9 candies and 2 biscuits` => it's not possible to calculate a ratio because we don't know the weight of this portion
- `30` => there is no unit

Entering values

Question: is there any difference between comma "," and dot "."? (example "2,5" vs "2.5")

Answer: no

Question: if the number on the product ends with ".0" ("8.0", for example). Does it make any difference, if we enter "8.0" or "8"?

Answer: yes

Question: if the number on the products starts with "<". ("<0.5", for example). Does it make any difference, if we enter "<0.5" or "0.5"?

Answer: yes

Question: what do I enter when the label says something like "contains trace amounts of ..."?

Answer: you can enter the word "traces" as the value, and the system recognizes this as "approximately zero".

Missing value

When the nutrition facts table is missing some values, enter the "-" (hyphen symbol) for that field. For example the *Fiber* value is sometime missing.

Photos and data check

At the end of the page, there is a specific section called "Photos and data check", documented as follow:

"Product pages can be marked as checked by experienced contributors who verify that the most recent photos are selected and cropped, and that all the product data that can be inferred from the product photos has been filled and is correct."

When `[x] Photos and data checked` is checked, Product Opener save the following data into the database:

- the value on in the field `checked`
- the last checked date (Unix timestamp format) into `last_checked_t`. Eg. `last_checked_t: 1704906819`.
- the last user who checked the product into `last_checker`. Eg. `last_checker: benbenben`.

- the users who have checked the products are stored in the field `checkers_tags`. Eg. `checkers_tags: ["benbenben","stephane"]`.

When `[x] I checked the photos and data again` is checked, Product Opener update the following data:

- the last checked date (Unix timestamp format) into `last_checked_t`.
- the last user who checked the product into `last_checker`.
- the users who have checked the products are stored in the field `checkers_tags`.

When a product is checked, this information is given on the product page, in the **Data sources** section, eg. "Last check of product page on January 10, 2024, 6:13:39 PM CET by charlesnepote".

These data can also be retrived via the API.

It's possible to explore the checked products with the dropdown menu "Explore products by...", selecting "Last checked dates", leading to the page: https://world.openfoodfacts.org/last-check-dates

It's possible to combine facets, for example, to get last checked dates of the TOP-1000 products in a country, eg. https://world.openfoodfacts.net/popularity/top-1000-fr-scans-2022/last-check-dates

---