

Overview

There are five types of data that need to be imported/exported at the same time: SKUs, ingredients, product lines, formulas, and manufacturing lines. At the customer's request, manufacturing lines do not need to be imported, but the relation between SKUs and manufacturing lines still does. In addition, for interoperability with existing spreadsheets, the formulas and their ingredients are specified in the same file.

CSV files do a good job representing table data, and are extremely easy to parse. In addition, CSVs are among the most space-efficient formats, so long as the tables they are representing are themselves the minimal. As discussed above, **four** CSV (comma-separated values) files are specified, and their headers are defined below:

Filename Prefix	Header
skus	SKU# , Name, Case UPC, Unit UPC, Unit size, Count per case, PL Name, Formula#, Formula factor, ML Shortnames, Rate, Comment
ingredients	Ingr# , Name, Vendor Info, Size, Cost, Comment
product_lines	Name
formulas	Formula# , Name, Ingr# , Quantity, Comment

Definitions

- Evolution requirements: the requirements set forth in the original requirements for the various evolutions. **In case of conflict, the evolution requirements supersede those in this document.**
- CSV file: a file compliant with [RFC4180](#).
- CSV cell: a single cell in a CSV file that contains in itself multiple values separated by comma(s). Note that these cells are necessarily quoted.
- CSV cell item: an individual item in a CSV cell. More formally, items the result of splitting the raw, quoted string value of a CSV cell by the regular expression `,\s*`.
- Type of file: each row in the table above is a type of file.
- Record: a row in any of the CSV files.
- Attribute: a column in any of the CSV files.
- Unique identifier: a column designated as unique, either in this document or in Evolution requirements.
- Primary key: an attribute that is **bolded** in the table above.
- Autogeneratable: an attribute is autogeneratable, and may be autogenerated, if it is so permitted in the Evolution requirements.
- Collision: a record that is identical in one or more unique identifier(s) to one existing in the system.
- Duplicate record: a record that is identical in any of its unique identifier(s) to another record in the same CSV file.
- Ambiguous record: a record whose values of unique identifiers refer to more than one record in the system.
- Session: a group of one or more request-response cycles in which a user submits/obtains a logical set of file(s) to the system.
- Batch: all files in a session.
- Batched Mode: an import/export workflow, in which multiple files may be submitted/obtained per session.
- Individual Mode: an import/export workflow in which only one file may be submitted/obtained per session.
- Referential integrity: the property that all unique identifier(s) must refer to exactly one record, that already exists in the system, or, when in Batched Mode, exists in another file of the same batch.
- Mixed-unit expression: a numeric expression followed by the name or symbol of a unit of measurement, optionally separated by spaces. E.g., "0.3 oz." or "300 Kilogram". The numeric expression is also referred to as "the number part", and the unit of measurement "the unit part".

File format

1. The names of the files must be prefixed by an identifying string, as defined above. The filenames must end in ".csv". More precisely, the filenames must constitute matches for the following regular expressions conforming to PCRE standard:
 - o SKUs: `^skus(\S)*\.csv$`
 - o Ingredients: `^ingredients(\S)*\.csv$`
 - o Product Lines: `^product_lines(\S)*\.csv$`
 - o Formulas: `^formulas(\S)*\.csv$`
2. The CSV files must be properly delimited and escaped. RFC4180 specifies the format of a CSV file. Note specifically:
 - o Spaces (and other space-like characters) must not be ignored if they are part of a field.
 - o Null values, if they are used, must be denoted by an empty string (i.e., two consecutive commas, or a comma followed by a `CRLF` character).
 - o The double quote symbol (ASCII 034) must be properly quoted.
3. A valid header, as defined above, must be included as the first line in each file. The header is case-insensitive.
4. Data types, uniqueness, requiredness, autogenerability of attributes in the files must conform to specifications in the **Evolution requirements**. If an attribute is not specified, it should follow the requirements below:
 - o All-string-typed attributes must be no more than 1000 characters.
 - o All Integer-typed attributes must be within range of $[-2^{31}, 2^{31})$ (i.e., a 32-bit signed integer).
 - o All floating-point numbers must be in decimal, and have a leading 0 (before the decimal place) if the value is less than 1.0.
 - o The column `ML_Shortnames` in `skus.csv` contains CSV cells. Each of the CSV cell items must be a string with no whitespace.
5. Each file must contain no more than 1,000,000 lines. The records may be in no particular order.
6. For **mixed-unit expressions**, e.g., ones for the "Size" attribute of an ingredient, the following processing algorithm must be used.
 - o The numbers (whole or decimal) are separated from non-number (the unit part). More formally, the expressions must constitute full matches for the following PCRE-standard regular expression: `1(\d*\.\d*)\s*(\D.*|$)`, where the number part will be taken from group 1 of the match, and the unit part from group 2.
 - o The unit part is then sanitized as follows:
 - i. Trimmed of whitespace on both ends
 - ii. Dots (".") and whitespaces (" ") are replaced with empty string
 - iii. Converted to lowercase
 - iv. Any trailing 's' is removed

- The unit part (after sanitization) is then interpreted by matching it exactly to the list below.

Unit	Accepted forms
Mass-based units	
Ounce	"oz", "ounce"
Pound	"lb", "pound"
Imperial ton	"ton"
Gram	"g", "gram"
Kilogram	"kg", "kilogram"
Volume-based units	
Fluid ounce	"floz", "fluidounce"
Pint	"pt", "pint"
Quart	"qt", "quart"
Gallon	"gal", "gallon"
Milliliter	"ml", "milliliter"
Liter	"l", "liter"
Count-based units	
Count	"ct", "count"

Upload Specifications

1. Only one file may be submitted in one import **session**. This file may be of any type.
2. All submitted **CSV file(s)** may be assumed to have MIME type `text/csv`, as stipulated in [RFC4180 §3](#).
3. The following checks will be performed on each file individually. If any of the check fails, the file will be rejected as a whole, and an appropriate error message will be displayed for the administrator.
 - File validity: the file must be standards-compliant, and contain the required headers in the order specified. The matching of headers should be case-insensitive.
 - **Duplicate records**: the file must not contain **duplicate records** in itself. The file may contain **collision(s)** to **records** existing in the system. Refer to the Collision Resolution algorithm in the [next section](#) when a **collision** is detected.
 - Data validity: where required, inputs must conform to standards.
 - UPC numbers must conform to the UPC-A standard
 - Sanitized units in **mixed-unit expressions** must be one in the accepted list. Further, units used in `formulas.csv` must be logically convertible to the original unit used in the ingredients. Similarly, if a record constitutes a **collision** in `ingredients.csv`, the unit, if different than that of the existing record, must be logically convertible to the existing unit.
 - Empty attributes: all required attribute(s) must be supplied, except when both of the following are true:
 - It is marked as **autogeneratable** in the **Evolution requirements**;
 - The record whose required attribute is empty contains no **collision**.
4. **Referential integrity** must be preserved in the file.

Note that the following relations exist among the files / corresponding entries in the system:

- In `skus.csv`,
 - PL Name must be existing in the system.
 - Formula# must be existing in the system.
 - All **CSV cell items** in `ML Shortnames` must be existing in the system.
- In `formulas.csv`:
 - Ingr# must be existing in the system.

Example formulas.csv:

Formula#,Name,Ingr#,Quantity,Comment

1001,Formula1,2001,0.1250000000,

1001,Formula1,2002,0.2500000000,

1002,Formula2,2001,0.1250000000,

1002,Formula2,2006,0.0005000000,

Example ingredients.csv:

2001,Bread,Butternut,340g,2.29,This bread is used to make all the burger

2002,Beef,Beef Choice,2lb,12.25,

2006,Turkey,,424 pounds,666.66,

Example product lines.csv:

Hamburger

Salad

Pasta

Example skus.csv:

SKU#,Name,Case UPC,Unit UPC,Unit size,Count per case,PL Name,Formula#,Formula factor,ML
Shortnames,Rate,Comment

1001,Burger Beef,042272010264,725272730706,600g,50,Hamburger,1001,1,"line1,,,,",1,

1002,Burger Turkey,987654321098,705632085943,250g,80,Hamburger,1001,1,"line1",1,