

Branch: master ▾ FoodFeedDataStandard / Data Standard.md

Find file Copy path

 adamlocker Update Data Standard.md

5ca731a 2 days ago

1 contributor

316 lines (263 sloc) 15.1 KB

Food and Feed Sampling Data Standard (Proposed)

What Is This Document?

This document describes the data standard for submitting food sampling results data to the Food Standards Agency. It describes the fields, their data types and order as well as providing specific guidance on acceptable values and which fields use specific reference data values.

This document does not describe the mechanism by which data can be submitted to the Food Standards Agency (the FSA).

Who Is This Document For?

This document is written for Local Authority and Laboratory users who need to submit sampling and results data to the FSA.

How This Document Is Structured

- [Food Standard Overview](#) Contains a brief overview of all the fields in the standard.
- [Field Definitions](#) Complete definitions for each field in the standard, includes constraints and specific data type formatting requirements.

1. [Local Authority](#)
2. [Sample Date](#)
3. [Sample Identifier](#)
4. [Sample Category](#)
5. [Food Description](#)
6. [Premises Name](#)
7. [Premises Postcode](#)
8. [Premises Type](#)
9. [Sample Reason](#)
10. [Sample Type](#)
11. [Follow Up Required](#)
12. [Follow Up Identifier](#)
13. [Food Poisoning](#)
14. [Food Poisoning Details](#)
15. [Survey Identifier](#)
16. [Manufacturer Name](#)
17. [Distributor Name](#)
18. [Importer Name](#)
19. [Country of Origin](#)
20. [Laboratory](#)
21. [Laboratory Comments](#)
22. [Determinand](#)
23. [Result \(Numeric\)](#)
24. [Resut \(Text\)](#)
25. [Satisfactory](#)

[26. Reported Date](#)[27. Prosecution](#)[28. Result Fail Code](#)

- [Supported File Types](#)
- [Other Requirements](#)
- [File Naming Conventions](#)

Food Standard Overview

The following table lists the fields (name and description), their data types, whether they are optional, and whether they use a controlled vocabulary.

Index	Field Name	Description	Data Type	Optional	Controlled Vocabulary	Source
1	local_authority	Local Authority identifier	Text	No	Yes	LA
2	sample_date	Date sample taken	Date	No	No	LA
3	sample_id	Sample unique identifier	Text	No	No	LA
4	sample_category	Sample category reference	Text	No	Yes	LA
5	food_description	Free text food description	Text	Yes	No	LA
6	premises_name	Name of the premises where the sample was taken	Text	No	No	LA
7	premises_postcode	Post code of the premises where the sample was taken	Text	No	No	LA
8	premises_type	Premises type reference	Text	No	Yes	LA
9	sample_reason	The reason for sampling	Text	No	Yes	LA
10	sample_type	The type of sample	Text	No	Yes	LA
11	follow_up	Follow up visit required	Boolean	No	No	LA
12	follow_up_id	Original sample unique identifier	Number	Yes	No	LA
13	food_poisoning	Food poisoning result	Text	No	No	LA
14	food_poisoning_details	Test details of food poisoning	Text	Yes	No	LA
15	survey_id	Survey unique identifier	Number	Yes	No	LA
16	manufacturer	Name of the manufacturer	Text	Yes	No	LA
17	distributor	Name of the distributor	Text	Yes	No	LA
18	importer	Name of the importer	Text	Yes	No	LA
19	country	Country of origin	Text	Yes	Yes	LA
20	laboratory	Laboratory identifier	Text	No	Yes	Lab
21	laboratory_comments	Examiner's opinion of the sample	Text	Yes	No	Lab
22	determinand	Determinand identifier	Text	Yes	Yes	Lab
23	result_numeric	Numerical value outcome of test	Number	No	No	Lab
24	result_text	Text explanation of outcome of test	Text	Yes	No	Lab
25	satisfactory	Shows if sample was satisfactory	Text	No	Yes	Lab
26	reported_date	Date results were reported to the Local Authority	Date	No	No	Lab

Index	Field Name	Description	Data Type	Optional	Controlled Vocabulary	Source
27	prosecution	Does the sample form part of a prosecution	Boolean	No	Yes	Lab
28	result_fail_code	The result fail code	Text	Yes	No	Lab

Field Definitions

1. Local Authority

Field Name: local_authority

Data Type: Text (controlled vocabulary)

Optional: No

Source: Local Authority

Comments: This is the three character Local Authority code as listed in the Government Digital Service (GDS) register for [England](#), [Northern Ireland](#), and [Wales](#) respectively. It should be noted that the Northern Ireland register is currently only in discovery, but GDS have advised us that the three character code is unlikely to change.

2. Sample Date

Field Name: sample_date

Data Type: Date (format: YYYY-MM-DD)

Optional: No

Source: Local Authority

Comments: The date the sample was taken, this should follow the YYYY-MM-DD format as laid out in the International Standard ISO 8601.

3. Sample Identifier

Field Name: sample_id

Data Type: Text (32 character limit)

Optional: No

Source: Local Authority

Comments: The sample number, as recorded by the sampling body. This must be a unique value within the records of that sampling body. It can be any combination of numeric or alphanumeric characters as long as it is unique.

4. Sample Category

Field Name: sample_category

Data Type: Text (controlled vocabulary, format: 00.00.00.00)

Optional: No

Source: Local Authority

Comments: The [sampling taxonomy](#) defines the acceptable values for this field. The sample taxonomy is a hierarchy of sampling classification with four levels. They are organised as Clan > Family > Group > Type . Each item in each level is numbered, these values are concatenated into a single string with a . used to separate them. Because they have a strict hierarchical relationship the classification only needs to be reported at the type level (format 00.00.00.00), the group, family and clan do not need to be supplied.

5. Food Description

Field Name: food_description

Data Type: Text (255 character limit)

Optional: Yes

Source: Local Authority

Comments: A free text description of the food being sampled. Where possible, the description should include the brand name as well as a short description of the type of product, e.g. "Mr Yum's Chocolate Digestives".

6. Premises Name

Field Name: premises_name

Data Type: Text (50 character limit)

Optional: No

Source: Local Authority

Comments: The name of the premises where the sample was taken. For premises that are one of many locations with the same business name, it is advisabel to include the street where that specific premises is located, e.g. Frank's Deli, High Street.

7. Premises Postcode

Field Name: premises_postcode

Data Type: Text (Between 5 and 8 characters)

Optional: No

Source: Local Authority

Comments: The postcode of the premises where the sample was taken.

8. Premises Type

Field Name: premises_type

Data Type: Text (controlled vocabulary)

Optional: No

Source: Local Authority

Comments: The premises type as defined by the [Categories of Food Establishment Register](#), the notation should be provided, for example a premises designated as a take-away has the notation RC-TA .

9. Sample Reason

Field Name: sample_reason

Data Type: Text (controlled vocabulary)

Optional: No

Source: Local Authority

Comments: The acceptable values for this field are:

- Surveillance OR S
- Monitoring OR M
- Enforcement OR E
- Investigation OR I

10. Sample Type

Field Name: sample_type

Data Type: Text (controlled vocabulary)

Optional: No

Source: Local Authority

Comments: The acceptable values for this field are:

- Formal OR F
- Informal OR I

11. Follow Up

Field Name: follow_up

Data Type: Boolean (True/False)

Optional: No

Source: Local Authority

Comments: Indicates whether this sample was taken during a follow-up visit.

12. Follow Up Identifier

Field Name: follow_up_id

Data Type: Text (32 character limit)

Optional: Yes

Source: Local Authority

Comments: The unique sample reference number for the original sample to which this is a follow up.

13. Food Poisoning

Field Name: `food_poisoning`

Data Type: Boolean (True/False)

Optional: No

Source: Local Authority

Comments: This field indicates whether or not this sample was taken as part of a food poisoning investigation.

14. Food Poisoning Details

Field Name: `food_poisoning_details`

Data Type: Text (255 character limit)

Optional: Yes

Source: Local Authority

Comments: The type of food poisoning outbreak being investigated as well as any additional pertinent information.

15. Survey Identifier

Field Name: `survey_id`

Data Type: Text (50 character limit)

Optional: Yes

Source: Local Authority

Comments: A unique reference to identify the survey under which this sample was taken. An entry in the fields indicates that this sample was taken as part of a survey, a null entry indicates that it was not.

16. Manufacturer Name

Field Name: `manufacturer`

Data Type: Text (255 character limit)

Optional: Yes

Source: Local Authority

Comments: The name of the manufacturer of the sampled product.

17. Distributor Name

Field Name: `distributor`

Data Type: Text (255 character limit)

Optional: Yes

Source: Local Authority

Comments: The name of the distributor of the sampled product.

18. Importer Name

Field Name: `importer`

Data Type: Text (255 character limit)

Optional: Yes

Source: Local Authority

Comments: The name of the importer of the sampled product.

19. Country of Origin

Field Name: `country`

Data Type: Text (controlled vocabulary)

Optional: Yes

Source: Local Authority

Comments: The country of origin. The field `country` from the [GDS Country Register](#) must be used. This is a two character ISO 3166 alpha 2 code.

20. Laboratory

Field Name: laboratory

Data Type: Text (controlled vocabulary)

Optional: No

Source: Laboratory

Comments: The name of the laboratory completing the testing on the sample. The notation for each lab will be published in due course. In the interim period a plain text name can be used.

21. Laboratory Comments

Field Name: lab_comments

Data Type: Text (2000 character limit)

Optional: Yes

Source: Laboratory

Comments: This free text field shows the examiners opinion of the sample.

22. Determinand

Field Name: determinand

Data Type: Text (controlled vocabulary)

Optional: Yes

Source: Laboratory

Comments: Describes the exact test completed on the sample. The [determinands taxonomy](#) contains this information. It is hierarchical and has four levels. Only the most detailed level should be used. For example, the test for Tartrazine assessed on a miligrams per kilogram basis has the notation CE102-05 .

23. Result (Numeric)

Field Name: result_numeric

Data Type: Number (any numeric data type)

Optional: Yes

Source: Laboratory

Comments: The result of the test completed on the sample, where the result is numerical, e.g. 1.4

24. Result (Text)

Field Name: result_text

Data Type: Text (255 character limit)

Optional: Yes

Source: Laboratory

Comments: The result of the test completed on the sample, where the result cannot be expressed numerically, e.g. <1.4 or less than one

25. Satisfactory

Field Name: satisfactory

Data Type: Text (controlled vocabulary)

Optional: No

Source: Laboratory

Comments: Indicates whether a sample is deemed satisfactory for a specific test. The acceptable values are:

- Satisfactory OR S
- Unsatisfactory OR U
- Borderline OR B

26. Reported Date

Field Name: report_date

Data Type: Date (format: YYYY-MM-DD)

Optional: No

Source: Laboratory

Comments: The date the results were reported, this should follow the YYYY-MM-DD format as laid out in the International Standard ISO 8601.

27. Prosecution

Field Name: prosecution

Data Type: Boolean (True/False)

Optional: No

Source: Laboratory

Comments: Indicates whether the sample forms part of a prosecution. It is critical to get this field correct as it will make sure that those samples involved in prosecutions are not published.

28. Result Fail Code

Field Name: result_fail_code

Data Type: Text (255 character limit)

Optional: Yes

Source: Laboratory

Comments: This field requires a full definition.

Supported File Types

Currently we are only supporting the standard for comma separated values (CSV) files. We have chosen this format as it is open, widely supported and easy to understand. Our new process for submitting the data to the FSA means that using CSV we can do some validation on files as they are submitted.

The [current standard for CSV](#) gives a detailed explanation of the common format for CSV files. It's concise and clear and we recommend reading it.

We expect to be able to support other formats (e.g. JSON or XML) in future, and are actively exploring it, but this would mean the sender would be responsible for validating their data against the standard.

Other Requirements

Text Fields

The majority of the fields in the standard are text, which needs to be treated carefully when stored in a CSV file. All text fields must be enclosed within double quotes "this is the text". You should try to avoid using double quotes within a text field as this can cause the field to be misread, but the RFC4180 standard allows it if handled appropriately.

If double-quotes are used to enclose fields, then a double-quote appearing inside a field must be escaped by preceding it with another double quote. For example: "aaa", "b""bb", "ccc"

Encoding

Where possible please use UTF-8 encoding.

File Naming Conventions

In order to make it easy for us to manage the files longer term, it will be important to name files so that we can tell who has submitted them and the time period that each file covers. The format will be the laboratory code from the laboratories register, and the date of submission in YYYYMMDD format.

For example, the file from the laboratory with code LAB10 for sampling data submitted the 31st of January 2017 would be named LAB1020170131.csv .