

Proposed files, elements and definitions for the interchange of compositional data in food

INTRODUCTION.....	2
FILE 1: SENDER (SENDER).....	2
FILE 2: AUTHORITY (AUTHORITY).....	3
FILE 3: BIBLIOGRAPHY (BIBLIO).....	5
FILE 4: COMPONENT GROUP (COMPGROUP).....	8
FILE 5: COMPONENTS (COMPONENT).....	8
FILE 6: CONVERSION FACTOR VALUES (FACTOR).....	9
FILE 7: FOOD GROUPS (FOODGROUP).....	10
FILE 8: FOOD ITEMS (FOODITEM).....	10
FILE 9: FOOD NAMES (FOODNAME).....	11
FILE 10: FOOD DESCRIPTION (FOODDESCR).....	12
FILE 11: CODED FOOD DESCRIPTION (CODEDESCR).....	12
FILE 12: SAMPLE (SAMPLE).....	13
FILE 13: METHOD (METHOD).....	14
FILE 14: COMPONENT VALUES (VALUE).....	15
FILE 15: CONTRIBUTING VALUES (CONTRIB).....	16
FILE 16: COMMON MEASURE OR PORTION SIZE (PORTION).....	18
FILE 17: LINK BETWEEN FOOD CODES FROM DIFFERENT DATABASES/DATASETS (FOODLINK)	19
FILE 18: ABBREVIATED DATASET (ABBREVIATE).....	19
ANNEX 1: Source type.....	20
ANNEX 2: INFOODS Component TAGNAMES.....	21
ANNEX 3: Unit.....	21
ANNEX 4: Base units.....	22
ANNEX 5: DERIVATION TYPE.....	22
ANNEX 6: Nutrient Definition Headline.....	26
ANNEX 7: Method Headline.....	26

INTRODUCTION

There is general agreement that XML would be the most suitable way to interchange data. However, technical limitations exist which make it difficult for some compilers, including those in North America and Europe, to accommodate an XML format. Therefore, as an intermediate step towards the XML interchange, a set of relational files is proposed in this document with standardized elements (with definitions) which could then be used as XML tags. The proposed files should therefore be regarded as collections of elements and their contents that can be presented in a single XML file or as a set of relational files.

In the construction of the files, attempts were made to be as non-prescriptive as possible while outlining the desired levels of metadata.

All 18 proposed files have the same structure with basic [definitions and terms](#). The column titled 'Element' identifies the discrete piece of information for which data entries are possible. In the column 'Prio', which stands for priority, 'M' indicates that this data entry is mandatory and that it has to be completed for interchanging data. 'Scope Note' provides definitions of the elements and explanations for their completion. The elements ending with '-id' are keys with which the relational files are linked with other, they are therefore always mandatory. In XML, elements would be identified as tags, which identify and surround contents. Content corresponds to the data entry for the element and may be values, keywords, other elements or a combination of them.

If XML is not used, the relational files may be sent as delimited ASCII text files. This means that each data record will correspond to the data entries of each element of a given file, separated by a delimiter as specified by the sender. If information for any element is missing, a delimiter of this element must be inserted in order to maintain the format. For consecutive missing elements the corresponding number of delimiters has to be indicated.

For more information see <ftp://ftp.fao.org/es/esn/infoods/interchange.pdf>

FILE 1: SENDER (SENDER)

This file (SENDER) is a **mandatory** file referencing the sender of the interchange files, e.g. the sender of the national database or of the regional data center. The fields are:

Element	Prio	Scope Note
senderid	M	Sender identification code, includes data sets retrieved from the web. It should be formed: region or country.publisher.specificDBandDate.sequencenumber. May be same as authorityid (File 2).
versionid		Version of the interchange package sent (e.g. as used in CEECFOODS interchange).
sendername		Name (last name, first name) of the sender of the data in ISO 646 characters, e.g. ASEANFOODS coordinator's name sending Thai data.

organization	M	Organisation of the sender of the data in ISO 646 characters, e.g. an INFOODS regional data centre coordinator's organisation sending national data set(s).
street		Street name and number of sender
city		Name of the city
postcode		Postal code.
state		State, department, county, region, province etc as indicated in postal address
country		Country code according to ISO 3166-1 (see http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html)
phone		Use the form +country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen.
fax		Use the form +country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen.
email		Internet e-mail address, e.g. nutrition@fao.org
<u>URI</u>		Complete URI of website hosting data set(s), e.g. http://www.fao.org/infoods
<u>available</u>	M	Date the interchange package is sent (format of this field is YYYY-MM-DD), or when downloaded from a website.
recordcount		Recordcount indicates the number and short names of files together with the number of records in every file. This will help the receiver of the data to evaluate if all files and all records have been received. E.g. file 1 AUTHORITY (2); file 2 BIBILIO (300); file 5 COMPONENT (120); File 6 FACTOR (12); file 7 FOODGROUP (15); file 8 FOODITEM (500); file 9 FOODNAME (1500); file 10 FOODDESC (500); file 14 VALUE (10000);
comment		Any further comment

FILE 2: AUTHORITY (AUTHORITY)

This file (AUTHORITY) is a **mandatory** file referencing the authority responsible for the data in the interchange files, e.g. the compiler of the national database. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code. The identification is unique in the file set and indicates who holds the responsibility for the interchanged data. It should be formed : region or country.publisher.specificDBandDate.sequencenumber
name		Name (last name, first name) of the responsible person of the sent data in ISO 646 characters, e.g. Miller, Anton
organization	M	Name of the responsible organisation of the sent data in ISO 646, e.g. National Institute of Nutrition
version	M	Actual version of the data set, e.g. USDA SR 17

ref		Data set citation: a consolidated field for the entire reference of the data set, e.g. U.S. Department of Agriculture, Agricultural Research Service. 2003. USDA National Nutrient Database for Standard Reference, Release 17.1. Nutrient Data Laboratory Home Page, http://www.nal.usda.gov/fnic/foodcomp
street		Street name and number of sender
city		Name of the city
postcode		Postal code
state		State, department, county, region, province etc as indicated in postal address
country		Country code according to ISO 3166-1
phone		Use the form +country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen.
fax		Use the form +country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen.
email		Internet e-mail address
URI		Complete URI for website hosting data set, e.g. http://www.nal.usda.gov/fnic/foodcomp
complang	M	Compilation language, used for the 'original language' fields of the interchange package, according to ISO 639 (see http://www.loc.gov/standards/iso639-2/langcodes.html#ab)
charset		To be used when this original food name requires a character set other than the ISO 646 Basic Character Set
compdate		Date the version is compiled (format of this field is YYYY-MM-DD) or when it is downloaded from the web
rights		Information about rights held in and over the resource. Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource. Therefore, indicate if the publication is protected by copyright. CR=copyright; NCR=no copyright issue involved, PD=Public domain; XX= do not know.
qualityassm		Quality assessment, i.e. how the Quality Identifier (applied to nutrient values) is defined by the data provider
biblioid		Bibliographic reference code of the data collection (key as in 'BIBLIO' file).
comment		Any further comment including glossary of terms, abbreviations or acronyms; meaning of non-numeric characters used in the 'value' field of VALUE file (e.g. 'tr' means trace); any other files (e.g. *.pdf, *.doc, *.xls, *.txt) are found useful to be interchanged e.g. the introduction section of the food composition table or indications of changes in food codes or names; or file with values (it would however be preferable to use for data interchange the files as outlined below).

FILE 3: BIBLIOGRAPHY (BIBLIO)

This file (BIBLIO) is **optional** and holds bibliographical information of the published and unpublished references cited in the dataset. See also the AgMES web site for the description of the metadata elements <http://www.fao.org/agris/agmes/Documents/Elements.html#top> and the user guide (<ftp://ext-ftp.fao.org/agris/agmes/AGRISAP-UserGuide.doc> (e.g. information on how to enter data for ‘creatorpersonal/creatorcorporate/creatorconference’ page 21). These may include the original sources, marked with ‘biblioid’, of the values contained in the national database (VALUE or CONTRIB files), references describing methods used (METHOD file), or the sampling (SAMPLE file), conversion factors (FACTOR file), the food name (FOODNAME file), the food description (CODEDESCR file), or common measurements (PORTION file), or any references cited within comments and other textual fields. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file).
biblioid	M	Bibliographic reference codes in original dataset, e.g. codes used in the national database to identify the published and unpublished references cited in the dataset. Each code is unique and an unambiguous reference to the location or identifier of resources. If a single reference list is interchanged, assign the number 1 and put all references in the ‘consolidated’ field.
consolidated		Consolidated field for complete citation of bibliographic reference, not broken down into the different fields in any style and language.
<u>title</u>		It is the title of the resource described in the original language, e.g. the title of a monograph, of a chapter, of a book or a journal article. In the rare cases where no title appears on the resource, e.g. unpublished data or personal communication, enter ‘No title’. Secondary information related to the title can be added, e.g. the title of the book or report in which the cited publication appears.
<u>creatorpersonal</u>		To be used if the author is a person(s) who is responsible for creating the intellectual content of the resource. There may be more than one, to be separated by a semicolon. Additions can be made after the author’s name: (ed) for editor, (comp) for compiler. Data should be entered in the <u>following sequence</u> : <i>surname, forename initial(s), prefixes, particles, role, affiliation. Surname and forename are essential, the other fields can be entered as supplementary information</i>
<u>creatorcorporate</u>		To be used if the author is an organization or a group of people identified by a name responsible for creating or making contributions to the intellectual content of the resource. Data should be entered in the <u>following sequence</u> : <i>Name of institution, Place (Country)</i>
<u>creatorconference</u>		To be used if the author is a conferences, e.g. named meetings of individuals or representatives of various bodies for the purpose of discussing and acting on topics of common interest, or meetings of representatives of a corporate body that constitute

its legislative or governing body. Data should be entered in the [following sequence](#): *Number of the conference. Name of the conference, place (country), date of conference*
ex. Instituto Nacional de Investigaciones Agrarias, Madrid (Spain)

[publishername](#)

Name of publisher or laboratory. A publisher is the individual, group, or organization named in the document as being responsible for that document's publication, distribution, issuing, or release.

[publisherplace](#)

Address of publisher or laboratory. The place of publication is the city, town, or other locality associated with the name of the publisher entered in the 'publishername' field.

[issued](#)

Date of formal issuance (e.g. publication) of the resource. Data should be entered in the form YYYY-MM-DD. As minimum information, the year YYYY should be included, even for unpublished material. For a static webpage containing a date, this should be the preferred date. Otherwise, indicate the date when URL was accessed. For unknown date use 'XXXX'.

[descriptionedition](#)

Edition number is the formally designated version of the data set or information resource being described. An edition is known by a word or phrase appearing in the item that normally indicates a difference in either content or form, and it denotes the existence of other versions of the work. This field should be completed **only when an edition statement appears on the resource**. Therefore, do not provide an edition statement for the **first edition** of a publication if not stated on the resource.

[language](#)

The language of the intellectual content of the resource. Best practice is to select from the ISO 639-2 (see <http://www.loc.gov/standards/iso639-2/langcodes.html#ab>). Enter one code or semicolon separated string. Use 'xxx' if language is unknown.

[ISBN](#)

International Standard Book Number. This field maps to (DC) identifier.

[ISSN](#)

International Standard Serial Number. As some journals have different ISSN if printed or online, please indicate [print] or [online] after ISSN; and if both ISSN are given, separate with a semicolon. Example: 0001-3072 [print];1467-6281[online]. This field maps to (AGS) citationIdentifier. See example <http://www.blackwellpublishing.com/xml/dtds/4-0/help/journalcodes.htm>

[DOI](#)

DOI (digital object identifier) is a permanent identifier given to a Web file or other Internet document so that if its Internet address changes, users will be redirected to its new address. This field maps to (DC) identifier.

[URI](#)

URI (Uniform Resource Identifier) is the internet address of the web site. This field maps to (AGS) citationIdentifier.

[RN](#)

RN (Report Number) is the number uniquely identifying a report, results of an investigation, proceedings at a meeting, or of any occurrence, event or matter on which definite information is required. This field maps to (DC) identifier.

<u>extent</u>	The pages, i.e. size or duration of the resource. For an entire book, enter the total number of pages. E.g. 134 p. For a chapter in a book, or article in journal or parts of a report, enter the pages covered, e.g. p. 19-32
<u>citationtitle</u>	The serial title, e.g. of journal, also known as the key-title, is a name ascribed to a serial publication and is inseparably associated with its ISSN. If possible, to be selected from the list from the Authority file. In the presence of full and abbreviated journal title, enter the full title first followed by the abbreviated title in brackets ().
abjrname	This is the abbreviated journal name or acronym. This should be completed for journals and journal articles, and should uniquely identify the journal by its <u>standard abbreviations according to International Serial Data System</u> (ISDS). This field maps to (AGS) citationtitle. E.g. JFCA stands for the 'Journal of Food Composition and Analysis'
<u>citationnumber</u>	The issue, part or number which denotes a particular issue of a journal, as it appears on the cover. In many cases this indicates a part of a journal volume and issue. Enter volume (issue). Ex. 3(2)
<u>valid</u>	Date (often a range) of validity of a resource. In the form YYYY-MM-DD to YYYY-MM-DD, for e.g. law or regulation or other authoritative document
<u>rights</u>	Information about rights held in and over the resource. Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights, Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource. Therefore, indicate if the publication is protected by copyright. CR=copyright; NCR=no copyright issue involved, PD=Public domain; XX= do not know.
<u>subject</u>	The topic of the content of the resource. Typically, these are keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a term from a controlled vocabulary or formal classification scheme. See website: <u>http://www.fao.org/agris/agmes/Documents/Elements.html#subject</u>
<u>decriptionnotes</u>	A brief statement, annotation, comment, or elucidation concerning any aspect of the resource. Information in a general note is information about the item that the compiler has deemed important to add to the record.

Note: The following fields are common for all the publication types: 'title', 'one of the author fields: 'creatorpersonal/creatorcorporate/creatorconference', 'publishername' 'publisherplace', 'issued', 'descriptionedition', 'language', and 'right' Specific fields may also be used, depending on the publication type such as ISBN, ISSN, DOI, valid etc.

FILE 4: COMPONENT GROUP (COMPGROUP)

This file (COMPGROUP) is an **optional** file describing the grouping of components. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
compgroupid	M	Component group code indicating to which component group(s) each component belongs to; e.g. selenium grouped in "trace element" and "antioxidant"
compgroupname		Name of the component group in the original data set.
compnameng		English name of the component group. It is highly recommended to complete this field for languages other than English.
comment		Attached notes, comments

FILE 5: COMPONENTS (COMPONENT)

This file (COMPONENT) is a **mandatory** file describing the components and linking the original component codes to the INFOODS component identifiers (TAGNAMES). This file can be used to identify the analytical method(s) of the component through the methodid, either applicable to the entire dataset/database (in this case leave the element 'foodgroupid' empty) or to specific food group(s) (in this case complete *foodgroupid* field). The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
componentid	M	INFOODS component identifier, also called TAGNAMES. (See annex 2)
origpcpd		Original component code, ID or abbreviation used to identify the component in the original dataset.
compname		Name of the component in the original dataset. Should always be included if ORIGPCPD is not a descriptive text code for the component.
compnameng		English name of the component of the original dataset. It is recommended to always be included .
compdef		Component definition text or code, following Annex 6, e.g. on summation or calculation.
methodid		Identification of the analytical method (key as in METHOD file).
foodgroupid		Food group code used in the original dataset (key as in file FOODGROUP). Can be a single group or semicolon separated string
unit	M	Unit following international recommendations (see Annex 3)
baseunit	M	Base unit following international recommendations , e.g. per 100 mL, per g N (see Annex 4)

compgroupid	Component group code (key as in COMPGROUP file). Can be a single group or semicolon separated string.
decimal	Number of decimal places displayed
comment	Attached notes, comments, i.e. any further remarks, algorithms, significant digits

Note: When many TAGNAMEs correspond to one original component code, as many rows as necessary should be entered. When attributing the 'componentid' to the values attention should be paid to choose the correct TAGNAME. The option 'unknown or multiple method(s)' is existing for most TAGNAMEs and it might be useful to add this combination of TAGNAME-origpcd. E.g. FIB- can mean that the method is unknown or that the value for this component is derived from aggregation of e.g. Southgate fibre, Prosky fibre and Englyst fibre values.

FILE 6: CONVERSION FACTOR VALUES (FACTOR)

This file (FACTOR) is an **optional** file defining conversion factors applicable to the entire database, a specific food group, or at the food level. When the conversion factors are treated as 'components' and the factors are provided in the VALUE file, this file should not be used. This file can be used for different purposes:

- to indicate that the conversion factor(s) is applicable for specific food group(s) through completing the 'foodgroupid',
- to indicate that the conversion factor(s) is applicable for specific food(s) through completing the 'fooditemid'
- to indicate that the conversion factor(s) is applicable for the entire database/dataset through leaving 'foodgroupid' and 'fooditemid' empty

However, values entered in 'foodgroupid' or 'fooditemid' overwrite default value for the database. The fields are:

Field label	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
factorid	M	INFOODS identifier, also called TAGNAMEs. (See Annex 2) (XCT, XCA, XCAM, XCF, ...)
biblioid		Bibliographic reference code (key as in BIBLIO file) to indicate the origin of the conversion factor value
foodgroupid		Food group code used in the original dataset (key as in FOODGROUP file). Can be a single group or semicolon separated string.
fooditemid		Original food code (key as in FOODITEM file). Can be a single food or semicolon separated string.
value	M	Factor value
comment		Attached notes, comments, i.e. any further remarks

FILE 7: FOOD GROUPS (FOODGROUP)

This file (FOODGROUP) is an **optional** file listing the codes used for the food groups in the original dataset. It is strongly recommended to complete this file as it can be helpful for the food identification and for the use of conversion factors and other factors. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file).
foodgroupid	M	Original food group code used in the dataset.
origfdgp		Name of the food group in original dataset
engfdgp	M	English name of food group
comment		Any further comment

FILE 8: FOOD ITEMS (FOODITEM)

This file (FOODITEM) is a **mandatory** file defining all food identification codes used in the complete data set. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
fooditemid	M	Original food code, ID or abbreviation used to identify the food in the original food composition database or publication. If the identifier is a line number in a printed table, this code should be presented as a combination of identifier for publication plus line number, in order to avoid duplicates.
foodgroupid		Original food group code used in the dataset (key as in FOODGROUP file)
imagefile		Give file name of image with commonly used format or URL link
comment		Attached notes, comments, i.e. any further remarks.

FILE 9: FOOD NAMES (FOODNAME)

The file (FOODNAME) is a **mandatory** file. It is a definition file listing food names and alternative food names. It can include alternative food names, taxonomic names or synonyms, or food names in different languages for foods listed in the 'FOODITEM' file. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
fooditemid	M	Original food code (key as in FOODITEM file)
name	M	Name of the food includes the translation of the original food name in the specific languages. If English is not the original language the translation into English is strongly recommended. Can be a single name or semicolon separated string.
lang	M	Language of the food name, according to ISO 639 (see Annex 8). Equals 'TX' for the taxonomic names (For taxonomic names see: http://www.langual.org/langual_linkcategory.asp?CategoryID=3&Category=Food+Classification%2C+Description%2C+Nomenclature+and+Taxonomy). Can be a single language or semicolon separated string corresponding to the string in 'name'.
charset		To be used when this original food name requires a character set other than the ISO 646 Basic Character Set. Can be a single character set or semicolon separated string corresponding to the string in 'name'.
namstatus		Used to distinguish between preferred and other names or for the food names. It should be indicated as O (original), P (preferred), A (abbreviation) or blank FORsynonyms or other names. Can be a single 'namstatus' or semicolon separated string corresponding to the string in 'name'.
edibdesc		Description of the nature of edible proportion of the food and how the value is calculated
edibdesceng		Description of the nature of edible proportion of the food and how the value is calculated in English
refusedescorg		Description of the nature of inedible proportion of the food and how the value is calculated in original language
refusedesceng		Description of the nature of inedible proportion of the food and how the value is calculated in English
biblioid		Bibliographic reference code (key as in BIBLIO file). Can be a single biblioid or semicolon separated string corresponding to the string in 'name'.
comment		Attached notes, comments, i.e. any further remarks

FILE 10: FOOD DESCRIPTION (FOODDESCR)

This file (FOODDESCR) is an **optional** file listing the food description of the foods listed in FOODITEM file. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
fooditemid	M	Original food code (key as in FOODITEM file)
descriptortag	M	Name(s) of the property or facet(s) of the food description(s). ex: Grade. Can be a single 'descriptortag' or semicolon separated string.
descriptorval	M	description(s) in text form e.g.: prime. Can be a single 'descriptorval' or semicolon separated string as defined in the 'descriptortag'.
charset		character set used for term e.g.: ISO 646
comment		Attached notes, comments, i.e. any further remarks

FILE 11: CODED FOOD DESCRIPTION (CODEDESCR)

This file (CODEDESCR) is an **optional** file listing the food description using a internationally recognized coded system such as LanguaL for the foods listed in FOODITEM file. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
fooditemid	M	Original food code (key as in FOODITEM file)
thesaurus		Name and version of coding system, i.e. LanguaL, Codex
biblioid		Bibliographic reference code (key as in BIBLIO file).
facetcode	M	The LanguaL facet code, ex. F0014. Can be a single code or semicolon separated string.
comment		Attached notes, comments, i.e. any further remarks

FILE 12: SAMPLE (SAMPLE)

The file (SAMPLE) is an **optional** file. It is a file identifying the sample and listing the sample details, as well as the fooditem code to which it refers. It is a complementary file and should be used only if additional information on the food sample is available. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
sampleid	M	Sample code, which may, in some cases, be the same as the fooditemid, where each code indicates a different sample reference code for a given food. If no sampleid exists, a code should be created.
fooditemid	M	Original food code (key as in FOODITEM file)
biblioid		Bibliographic reference code (key as in BIBLIO file)..
sampplan		Succinct description of sampling method, if available. Indicate if nationally representative, proportional etc
sampdate		Date of sample collection of the food, if available. The format of this field must be one of YYYY-MM-DD, YYYY-MM, YYYY or YYYY-YYYY (for year ranges, e.g. for an analytical project).
sampdesc		Description of the food sampled, not already included in the food description files (FOODDESCR or CODEDESCR), e.g. agronomic conditions.
sampcoll		Place(s) where the food sample(s) was obtained, purchased, harvested, etc.
sampfdnr		Number of food samples collected to construct analytical sample
sampwght		Weights of food samples as collected in gram
sampannr		Number of analytical samples, either individual or composite
sampanrep		Number of analytical replicates per sample. Can be semicolon separated string corresponding to 'sampannr'
samphand		Food specific handling of sample before arrival at laboratory, e.g. sample transport.
samparriv		Sample handling when the sample arrives at the laboratory
sampstor		Storage conditions in the laboratory before the start of the analytical process, e.g. temperature.
sampreason		Reason for analysis, i.e. context of investigation e.g. for clinical, comprehensive, control, or contamination study
comment		Attached notes, comments, i.e. any further remarks

Notes:

- 'sampleid' should be reported in VALUE or CONTRIB file only if the sample description is available of the contributing values.
- 'samplefdnr' and 'sampleannr' are useful information but difficult to define. If 12 food samples were collected and put into 1 composite sample, samplefdnr is 12 and sampleannr is 1.

FILE 13: METHOD (METHOD)

This file (METHOD) is **optional** and contains information about the analytical and derivation method(s) associated with a given component through the bibliographic reference.

Even though this file is optional it is highly recommended to include method information at least for some nutrients such as folate, fat, water, fat-soluble vitamins for which some of the analytical methods generate different values.

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file).
componentid	M	INFOODS component identifier (key as in COMPONENT file).
methodid	M	Method code. Code or text as captured in the database.
biblioid		Bibliographic reference code of the method (key as in BIBLIO file).
methcode		AOAC or other national official method code for complete analytical method starting from preparation to determination. The identification of the coding system has to be included in the code, e.g. AOAC 985.29. If more than one code is used, the purpose can be specified with attributes, i.e. using [PREP] for preparation, [SEP] for separation, [QT] for quantification after the method code.
methhdl		Method Headline code(s) used for the entire analysis, following the COST99 Recommendations (see Annex 7). Several Method Headline codes can be entered if appropriate. They can be specified by putting [PREP] for preparation, [SEP] for separation, [QT] for quantification after the method headline.
methdesc		Description of the methods used (in the absence of codes) and/or of the modifications done compared to AOAC or other coded official method
analysedate		Date of analysis of the sample, if available. The format of this field must be one of YYYY-MM-DD, YYYY-MM, YYYY or YYYY-YYYY (for year ranges, e.g. for an analytical project). Further notes can be added in square brackets
prepsampl		Sample preparation description (e.g. blending, homogenisation, hydrolysis) for the analytical sample.
prepanadt		Date of preparation of sample for analysis, if available. The format of this field must be one of YYYY-MM-DD, YYYY-MM, YYYY or YYYY-YYYY (for year ranges, e.g. for an analytical project). Further notes can be added in square brackets
prepbiblioid		Bibliographic reference code (key as in BIBLIO file) describing the preparation method.
prepmodif		Description of the modifications done compared to AOAC or other national official method. The text may include reference citations (preferably with full information in the BIBLIO file), but should never just consist of a reference. A plain reference should always be entered as a bibliographic code in the 'prepbiblioid' field.
sepanadt		Date of the separation part of the analysis, if available. The format of this field must be one of YYYY-MM-DD, YYYY-MM, YYYY or YYYY-YYYY (for year ranges, e.g. for an analytical project). Further notes can be added in square brackets
sepbiblioid		Bibliographic reference code (key as in BIBLIO file) describing the

	separation method.
sepmoif	Description of the modifications done compared to AOAC or other national official method. The text may include reference citations (preferably with full information in the BIBLIO file), but should never just consist of a reference. A plain reference should always be entered as a bibliographic code in the 'sepbiblioid' field.
qtanadt	Date of the quantification part of the analysis for the value, if available. The format of this field must be one of YYYY-MM-DD, YYYY-MM, YYYY or YYYY-YYYY (for year ranges, e.g. for an analytical project). Further notes can be added in square brackets
qtbiblioid	Bibliographic reference code (key as in BIBLIO file) describing the quantification method.
qtmofif	Description of the modifications done compared to AOAC or other national official method. The text may include reference citations (preferably with full information in the BIBLIO file), but should never just consist of a reference. A plain reference should always be entered as a bibliographic code in the 'qtbiblioid' field.
LOD	Limit of detection of the method
LOQ	Limit of quantification of the method
comment	Notes and comments relating to the method for the component.

FILE 14: COMPONENT VALUES (VALUE)
--

This file (VALUE) is **mandatory** and contains the description of the value and the value as such. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file)
fooditemid	M	Original food code (key as in FOODITEM file)
sampleid		Sample code (key as in SAMPLE file).
componentid	M	INFOODS Component identifier (key as in 'COMPONENT' file).
value	M	Component value ('best location') that is considered the best representative value according to the decision of the data compiler. It should use decimal points (and not commas). It is strongly recommended that this field only contains numeric characters. If non-numeric characters are used the explanations for abbreviations should be entered in the comment field in the AUTHORITY file.
valtype		Value type, i.e. qualitative description of the value. The following vocabulary should be used: <ul style="list-style-type: none"> • &gt; for more than; • &lt; for less than; • tr for trace; • &lt;LOQ for under limit of quantification; • &lt;LOD for under limit of detection; • halfLOQ for half of limit of quantification; • halfLOD for half limit of detection;

	<ul style="list-style-type: none"> • ND for non detected; • MN for mean; • MD for median; • MO for mode
unit	Unit following international recommendations, e.g. g, mg, mcg (see Annex 3). This field should be completed only to overwrite the default unit from the COMPONENT file
baseunit	Base unit following international recommendations, e.g. per 100g edible portion, per 100 ml (see Annex 4). This field should be completed only to overwrite the default base unit from the COMPONENT file.
sourcetype	Source type of the value indicating in general how the value was obtained (see Annex 1)
derivtype	Derivation type provides more detailed information than 'sourcetype' on the process used to obtain the value. It may include a fooditemid separated by semicolon from derivtype (see Annex 5).
sd	Standard deviation of the value, for normal distributions only. SD can be used for average values, if n > or = 3 (for n=1 or 2 SD is not valid). It is allowed for trace.
se	Standard error of the value, for normal distributions only. SE can be used for average values, if n > or = 3 (for n=1 or 2 SE is not valid). It is allowed for trace.
min	The minimum value for n > 1 only. It is allowed for trace.
max	The maximum value for n > 1 only. It is allowed for trace.
mean	The mean value if it is different from the 'value' chosen.
median	The median value if it is different from the 'value' chosen.
lowerror	lower 95% error bound
higherror	higher 95% error bound
n	Number of samples analysed (= 'sampleannr'). Analytical replicates should not be included in the number.
methodid	Identification of the analytical method (key as in METHOD file).
qi	Quality indicator, i.e. result of any systematic quality assessment applied by the data provider (refers to the 'qualityassm' element of the AUTHORITY file)
comment	Attached notes, comments, i.e. any further remarks. E.g. fortification

FILE 15: CONTRIBUTING VALUES (CONTRIB)

This file (CONTRIB) is **optional** and indicates the necessary information of contributing values to a given food-nutrient pair, which applies for raw analytical data (n) or means of means (N). This file allows accommodating several scenarios:

- To indicate that **all** components of a food present in the original dataset are contributing to another food in the database (leave the componentid field empty)
- To indicate that **one** component of a food present in the original dataset are contributing to another food in the database (complete the componentid field)

- To indicate through **sampleid (and 'componentid')** the food (and component) which contribute to the aggregated food.
- To indicate the source reference, the food code and the contributing value if a contributing food and its value(s) is not part of the original dataset through *biblioid*, 'externalfdid' and 'contribvalue'
- This file can also be used to record recipe information if ingredients are all raw or if the cooked ingredients are part of the database/dataset and the proportions of the cooked ingredients are known

The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file).
fooditemid	M	Original food code in own database (key as in FOODITEM file) indicating the food to which contributing value(s) are attached
componentid		INFOODS component identifier (key as in COMPONENT file), indicating the component of the food to which a contributing value is attached. Can be a single value or a semicolon separated string. If this field is blank it means that all components of this food were used as contributing values
contribfooditemid		Original food code in own database (key as in FOODITEM file) from which the component value(s) are used to form an aggregated fooditem in the original dataset.
sampleid		Identification of the food sample (key as in SAMPLE file). Can be a single sample or a semicolon separated string.
biblioid		Bibliographic identification (key as in BIBLIO file). If the food of the contributing value is not part in the dataset/database. It can be an external food composition database
externalfdid		Food identification code in the bibliographic reference from the external food composition table indicating the food from which the value(s) are taken from.
contribvalue		Numerical value contributing to the aggregated food. It may be value(s) from the sample(s) or from foods external to the database. It can be a single value or a semicolon separated string corresponding to 'sampleid'.
valtype		Value type, i.e. qualitative description of the value. The following vocabulary should be used: <ul style="list-style-type: none"> • > for more than; • < for less than; • tr for trace; • <LOQ for under limit of quantification; • <LOD for under limit of detection; • halfLOQ for half of limit of quantification; • halfLOD for half limit of detection; • ND for non detected; • MN for mean; • MD for median; • MO for mode

unit	Unit of the 'contribvalue' e.g. g, mg, mcg (see Annex 3). It should be completed only if the unit of the contributing value is different from the default unit of the value.
baseunit	Base unit of the 'contribvalue' e.g. per 100g edible portion, per 100 ml (see Annex 4). It should be completed only if the base unit of the contributing value is different from the default base unit of the value.
weighting	In case of weighted aggregations. Values should be between 0 and 1.
sourcetype	Source type of the value indicating in general how the value was obtained (see annex 1)
derivtype	Derivation type, i.e. more detailed information than 'sourcetype' on how the value was obtained (see Annex 5)
methodid	Identification of method of the 'contribvalue' (key as in METHOD file).
qi	Quality indicator, i.e. result of any systematic quality assessment applied by the data provider (refers to the 'qualityassm' element of the AUTHORITY file)
comment	any further remarks

FILE 16: COMMON MEASURE OR PORTION SIZE (PORTION)
--

This file (PORTION) is **optional** and indicates the common measurements or portions used to quantify this food. The fields are:

Element	Prio	Scope Note
authorityid	M	Authority identification code (key as in AUTHORITY file).
fooditemid	M	Original food code (key as in FOODITEM file)
portionid	M	Code of the common measurement or portion.
amount	M	Number of measurements, e.g. 1 in '1 cup' or 30 in 30 nuts
portiondesc	M	Description of the common measurement or portion (slice, package, piece 5x5x3 cm, can (330ml), tablespoon, cup (200ml)-chopped, cup(200ml)-sliced, etc)
portionweight	M	Gram weight
portionsamplnr		Number of analysed data points
portionsd		Standard derivation of the analysed data points of this common measurement or portion
portiondensity		Density in kg/L for this specific common measurement or portion, e.g. density for grated vs. density for whole
biblioid		Bibliographic reference code (key as in BIBLIO file).
comment		any further remarks

FILE 17: LINK BETWEEN FOOD CODES FROM DIFFERENT DATABASES/DATASETS (FOODLINK)
--

This file (FOODLINK) is **optional** and indicates the links between food codes from different databases/datasets. The fields are:

Element	Prio	Scope Note
senderid	M	Sender identification code (key as in SENDER file)
biblioid1	M	Bibliographic identification of the own database/dataset or external food composition dataset /database corresponding to 'foodcd1' (key as in BIBLIO file).
foodcd1	M	Food identification code of the own database/dataset or external food composition dataset /database number 1 be linked to 'foodcd2'
biblioid2	M	Bibliographic identification of an external or own food composition dataset /database corresponding to 'foodcd2' (key as in BIBLIO file).
foodcd2	M	Food identification code of the external or own food composition dataset /database number 2 linked to 'foodcd1'
linktype		Indicate the type of link between foods 'foodcd1' and 'foodcd2'. The codes are as following: <ul style="list-style-type: none"> • 'S' for same if the foods are exactly the same (including food description) • 'R' for related if the foods are similar foods (may be two similar foods or the same food but with different food description) • 'N' for unknown
comment		any further remarks

FILE 18: ABBREVIATED DATASET (ABBREVIATE)
--

This file (ABBREVIATE) is a concession to formal interchange and requires further deliberation on elements and format. If an abbreviated dataset is interchanged, minimum standardisation should include 'componentid' using the INFOODS TAGNAMES.

ANNEX 1: Source type

The source type codes are based on [COST99 recommendation on data management and interchange](#) as well as on other examples, e.g. USDA. The source type is considered general information on the source of the value. If more information on imputation techniques are known the derivation type should be completed in addition.

Code	Descriptor	Scope note
AO	Original analytical data	Analytical values with full or partial documentation, from in-house or affiliated laboratories
ASC	Analytical data from scientific literature	Value from peer reviewed scientific study, published in a scientific journal.
AAG	Aggregated of analytical values	Value is the result of the aggregation of analytical values, including values coming from manufacturers'
AI	From non-scientific journal publication	Analytical values from published articles reports, books, posters, letters, personal communications, websites ... not necessarily peer reviewed,
ESTIM	Imputed/estimated	Value is imputed, estimated, or guesses. If possible, it should be further described using derivation codes
FCDB	From food composition tables, databases or datasets	Value from compiled food composition tables, databases or datasets or from interchange or software packages.
LABEL	Food label, product information	Value based on food label or product information as provided by the producer or distributor often without further information about the data sources.
LC	Claim	Value based on manufacturer's label claim for added nutrients. For fortificants, it can be just the fortificant level, or the sum of fortificant and naturally occurring nutrient.
MANUF	Calculated by manufacturer	Value is calculated by manufacturer, not adjusted or rounded for national nutritional labelling programs, such as NLEA in the United States
HOST	Value created within host-system	Values created within own database using calculation or estimation, e.g. linear regression, concentration adjustment, average value of food group, based on other ingredients, based on physical composition, calculation by difference. Note: simple unit conversion does not fall into this category. It should be further described using derivation codes
RECIP	Calculated through recipes	Recipe calculation, ideally including complete list of ingredients with description and quantities, and list of applied yield and retention factors at ingredient or recipe level.

ASSUM	Presumed value	Assumed or presumed value according to information on food or food and labelling regulations or law, not including zero
Z	Assumed zero value	The component is not found in this food, naturally or by regulation, e.g. alcohol in meat, or fat in mineral water.
OTHER	Other source type	Value is based on other source type not mentioned in this list
X	Source type not known	

ANNEX 2: INFOODS Component TAGNAMES

A set of revised TAGNAMES are available on the INFOODS website: http://www.fao.org/infoods/tagnames_en.stm. They are currently being revised. A revised version is expected to be posted soon on the same website.

ANNEX 3: Unit

Unit description is influenced by International Standard, ISO 1000:1992 (incl. Draft Amendment 1, ISO 1000:1992/DAM 1(1997)). The standard is extended with food composition specific units.

Code	Descriptor	Scope note
kg	kilogram	
g	gram	
mg	milligram	
ug	microgram	
ng	nanogram	
L	Liter	
mL	milliliter	
uL	microliter	
mmol	millimol	
kJ	kilojoule	
kcal	kilocalorie	
ppm	parts per million	mg per gram
R	Ratio	
X	not applicable	to be used for e.g. conversion factors

ANNEX 4: Base units

The information is based on the [COST99/EUROFOODS recommendations](#) of the working group on data management and interchange.

Code	Descriptor	Scope note
W	per 100g edible portion	
T	per 100g total food	As purchased including any waste e.g. chicken wing with bones, banana including peel, etc.
D	per 100g dry weight	
WKG	per kg edible portion	
TKG	per kg total food	
DKG	per kg dry weight	
VL	per l food volume	
V	per 100 ml food volume	
N	per g total nitrogen	
NH	per g protein nitrogen	
AS	per 100 g total amino acid	if the corresponding unit is g, then this expression is the same as % of total amino acids?
P	per 100 g protein	
TFA	per g total fatty acids	The corresponding unit should normally be mg fatty acid
FA	per 100 g total fatty acids	If the corresponding unit is g fatty acid, then this expression is the same as % of total fatty acids
FT	per g total fat	
FTG	per g fat as triglyceride equivalent	
TF	per 100g total fat	
TTG	per 100g fat as triglyceride equivalent	
X	not applicable	

ANNEX 5: Derivation Type

The derivation types are based on [USDA derivation codes](#). The derivation types are providing more detailed information on imputation techniques than the source type. The derivation type can therefore be completed in addition to the source type or in some cases instead of the source type.

Code	Description	Scope note
A	Analytical data	
AI	Analytical data; from the literature or government; incomplete documentation	Incomplete documentation in published or unpublished document
AR	Analytical data; derived by linear regression	
AS	Analytical data; derived by summation of analytical data and assumed zero	

BU	Based on another form of the food or similar food, unknown further procedures	If based on one food of the own database its 'fooditemid' should follow the derivation type code, separated by a semicolon. For all other cases the CONTRIB file should be used.
BD	Based on same food; Drained solids from solids and liquids or vice versa (canned fruits and vegetables)	see scope note of BU
BFAN	Based on another form of the food or similar food; Concentration adjustment for Ash; Retention factors not used	see scope note of BU
BFCN	Based on another form of the food or similar food; Concentration adjustment for Carbohydrate; Retention factors not used	see scope note of BU
BFFN	Based on another form of the food or similar food; Concentration adjustment for Fat; Retention factors not used	see scope note of BU
BFFY	Based on another form of the food or similar food; Concentration adjustment for Fat; Retention factors used	see scope note of BU
BFNN	Based on another form of the food or similar food; Concentration adjustment for Non-fat solids; Retention factors not used	see scope note of BU
BFNY	Based on another form of the food or similar food; Concentration adjustment for Non-fat solids; Retentions factors used	see scope note of BU
BFPN	Based on another form of the food or similar food; Concentration adjustment for Protein; Retention factors not used	see scope note of BU
BFPY	Based on another form of the food or similar food; Concentration adjustment for Protein; Retention factors used	see scope note of BU
BFSN	Based on another form of the food or similar food; Concentration adjustment for Solids; Retention factors not used	see scope note of BU
BFSY	Based on another form of the food or similar food; Concentration adjustment for Solids; Retention factors used	see scope note of BU
BFYN	Based on another form of the food or similar food; Yield adjustment; Retention factors not used	see scope note of BU
BFZN	Based on another form of the food or similar food; No concentration adjustment; Retention factors not used	see scope note of BU
BFZY	Based on another form of the food or similar food; No concentration adjustment; Retention factors used	see scope note of BU
BNA	Based on another form of the same food or similar food: constituents normalized to total weight; vitamin A adjusted	

CAU	Calculated from different food or average values of food category, unknown further procedures	
CAAN	Calculated average values of food category for Ash; Retention factors not used	
CACN	Calculated average values of food category for Carbohydrate; Retention factors not used	
CAFN	Calculated average values of food category for Fat; Retention factors not used	
CASN	Calculated average values of food category for Solids; Retention factors not used	
CAZN	Calculated average values of food category; No adjustment; Retention factors not used	
DA	Concentration adjustment using factor; derived from analytical data	
DI	Concentration adjustment using factor; derived from imputed data	
FLA	Estimated formulation based on ingredient list; Linear program used to estimate ingredients based on Analytical data	
FLC	Estimated formulation based on ingredient list; Linear program used to estimate ingredients based on Claim on label/serving	
FLM	Estimated formulation based on ingredient list; Linear program used to estimate ingredients based on manufacture's calculation for data per 100g edible food	
JA	Aggregated data involving combinations of analytical data	
JO	Aggregated data involving combinations of data with partial analytical data or aggregation of analytical data	
LC	Label claim serving (g or %RDI or RDA) are back calculated by compiler to per 100g food	
MA	Manufacturer supplied (industry or trade association) analytical data, incomplete documentation	
MC	Manufacturer supplied; Calculated by manufacturer or unknown if analytical or calculated	
ML	Manufacturer supplied; Value upon which manufacturer based label claim for fortified/enriched nutrient	
NC	Nutrient that is based on other nutrient/s; calculated rather than analyzed	
NP	Nutrient that is based on other nutrient/s; calculated by difference or summed (with or without activity factors)	
NR	Nutrient that is based on other nutrient/s; value used directly, ex. fat from total fatty acids	
PAE	Based on estimated segment reconstitution; Derived from analytical data;	E.g. Used for meat cuts to calculate nutrient values by adding the amounts of nutrients in the lean and the fat

		portion
PAK	Based on known segment reconstitution; Derived from analytical data; Known segment reconstitution	see scope note of PAE
PIE	Based on estimated segment reconstitution; Derived from imputed data;	see scope note of PAE
PIK	Based on known segment reconstitution; Derived from imputed data;	see scope note of PAE
RA	Recipe; Approximate ingredient proportions (ex. combination of several recipes)	
RC	Recipe; Cookbook	
RF	Recipe; Formulary of standard products (formulary or standards of identity)	
RK	Recipe; Known formulation (dissection data or proprietary formulation)	
RKA	Recipe; Known formulation; No adjustments applied, combination of analytical data or aggregation of analytical data.	
RKI	Recipe; Known formulation; No adjustments applied, combination of analytical data or aggregation of analytical data	
RP	Recipe; Per package directions (e.g. refrigerated dough, toast, cake mix)	
RPA	Recipe; Per package directions; No adjustments applied, combination of analytical data or aggregation of analytical data	Ingredients could be a mixture of different data sources, but is that too much detail
RPI	Recipe; Per package directions; no adjustments applied, with partial analytical data or aggregation of analytical data	
S	Product standard, e.g. enrichment level	
T	Taken from another source-other food composition tables/databses	Source should be documented using CONTRIB file
Z	Assumed zero	The component is not found in this food, naturally or by regulation, e.g. alcohol in meat, or fat in mineral water
O	Other procedure used for imputing	

ANNEX 6: Nutrient Definition Headline

The Nutrition Definition Headlines correspond to information captured in the TAGNAME concept through keywords and secondary TAGNAMES. This file is still under construction and will allow to report separately nutrient definitions and calculation modes and information on analytical methods and component identifiers. It is however still undecided if nutrient definition headlines will be kept separated or remain part of the TAGNAME concept.

ANNEX 7: Method Headline

This file is still under construction and will be developed by the EuroFir project (<http://www.eurofir.net/>). It will provide generic method headlines (ME), sometimes specified by an additional reference or information. A preliminary list is available in the [COST 99/EUROFOODS recommendations](http://www.eurofir.org/COSTAction99/COSTAction99Publications.htm) (<http://www.eurofir.org/COSTAction99/COSTAction99Publications.htm>).