FP7-KBBE-2010-4

Grant agreement no: FP7-265967

EuroFIR – Nexus

The EuroFIR Food Platform: Further integration, refinement and exploitation for its long-term self-sustainability

Deliverable 1.11

1st Update of EuroFIR Thesaurus

Due date of milestone or deliverable: M11

Actual submission date: M13

Start Date: April 01, 2011 Duration: 2 years

Organisation name of lead contractor for this deliverable or milestone: EuroFIR AISBL [IDUFIC]

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level (please check appropriate box)		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EUROFIR NEXUS WP1.3.1

REPORT ON THE 1ST UPDATE OF EUROFIR THESAURI PREPARED BY WORK PACKAGE 1.3.1

Ian Unwin (IDUFIC)

in collaboration with Anders Møller and Jayne Ireland (DFI)

***ABSTRACT**:*

Updates to the EuroFIR thesauri have been made in response to requirements of EuroFIR compilers for maintaining their food composition databases. The updates have involved the Component, Method Indicator and Matrix Unit thesauri and the revised versions of these thesauri have been numbered 1.3. Issues related to the thesaurus entries are described and areas requiring further work or discussion have been identified. For carbohydrates in the Component Thesaurus, work has involved polysaccharides and polyols. New terms have been added for individual fatty acids and their totals, as well as for a number of trace elements. Other additions have been made for amino acids and organic acids, and improvements have been made for various component terms. For the Method Indicator thesaurus, several issues relating to the classification, non-reported methods and recipe calculation are addressed. The main area of work for specific terms has been on energy calculation methods, but some other issues relating to calculation methods are also addressed. One new term has been added to the Matrix Unit thesaurus.

EuroFIR thesaurus updates

Report on the preparation of version 1.3

Ian Unwin (IDUFIC)
in collaboration with Anders Møller and Jayne Ireland (DFI)

Version date: 30 March 2012

Contents

Introduction 4
EuroFIR Component Thesaurus 4
Carbohydrate components 4
Polysaccharides and analogues 4
Polyols 5
Lipid components 5
Individual fatty acids 5
Fatty acid totals 6
Fats 6
Glycerol 7
Phospho- and glycolipids 7
Nitrogen components 7
Amino acids 7
Organic acids 8
Individual organic acids 8
Vitamins and related compounds 8
Carotenoids 8
Thiamin 8
Vitamin C 9
Minerals 9
Elements 9
Speciation 9
Food properties 10
Components added in version 1.3 10
Method Indicator thesaurus 11
General issues 11
Classification terms and hierarchy 11

Non-indexed and unknown methods 12
Recipe calculation methods 12
Carbohydrate calculations 13
Energy 13
Carbohydrate, available 14
Lipid components 15
Individual fatty acids 15
Fatty acid totals 15
Nitrogen components 15
Total nitrogen 15
Vitamins and related compounds 15
Vitamin A 15
Matrix Unit thesaurus 16
References 16

Introduction

This document records the changes made in preparing version 1.3 of the EuroFIR Thesauri. Version 1.3 is the first of two updates being prepared with the support of the EuroFIR Nexus project. It was finalised in March 2012, with the second update, version 2.0, due at end-February 2013. Many of the requested additions to the thesauri were identified through a questionnaire on data compilation completed by EuroFIR compilers in 2011 (Ireland *et al.*, 2012). Updates have only been necessary for the Component, Method Indicator and Matrix Unit thesauri and their new versions are numbered 1.3. The remaining EuroFIR thesauri remain as versions 1.1. The earlier updates for versions 1.1 and 1.2 of the Component Thesaurus are documented in Unwin (2010A) and Unwin (2010B), respectively.

In parallel work on component identification, INFOODS is establishing several working groups to review its component tagnames for specific types of component such as carbohydrates and fatty acids. It is unlikely that these will produce results for version 1.3 of the EuroFIR Component Thesaurus, but are likely to be a significant consideration for version 2.0.

Editing of the EuroFIR thesauri is performed using the Thesaurus Manager software, which also permits read-only access by a wider audience of thesaurus users. This is now available on the EuroFIR server at:

http://ethesaurus.eurofir.org/ (username: eThesaurusGuest, password: guest$Welcome).

Changes to the thesaurus have been documented in the Editor's notes field of Thesaurus Manager.

EuroFIR Component Thesaurus

Carbohydrate components

Polysaccharides and analogues

A term was added for beta-glucans. β-D-Glucans (CHEBI:28793) are polysaccharides containing only glucose units, linked by β-glycosidic bonds. The original INFOODS list (Klensin *et al.*, 1989) includes the tagname GLUCNB for so-called "betaglucan", which can be taken to be equivalent to β-D-glucans and therefore the EuroFIR identifier used is GLUCNB, with the descriptor being "beta glucans"[1].

An entry has been added for the polysaccharide analogue chitin. Chitin (CHEBI:17029) consists of 2-(acetylamino)-2-deoxy-D-glucose) units linked through β-1,4 bonds and thus is a cellulose with one hydroxyl group on each glucose unit replaced by an acetylamino group. The INFOODS tagnames, added in the 2008 update (INFOODS, 2008), is CHITIN and this is proposed as the EuroFIR identifier. It has been added to a new group for **Polysaccharide analogues** (and not **Modified polysaccharides** as this might be confused with modified starch).

BLS requested a term for "polysaccharides (>10M)", a total for polysaccharides with a degree of polymerisation (DP) above 10 sugar units. INFOODS introduced the tagname POLYSAC for "polysaccharides, total" in their 2008 update, with the definition "Sum of starch and non-starch-polysaccharides. DP >10". EuroFIR can introduce the equivalent identifier, but this is a difficult area. If this is indeed a summation, there should be identifiers for the contributing components, in this case STARCH and NSP. STARCH includes dextrins, which may be (or may include) carbohydrates with DP <11. Definitions and methods for polysaccharides (and

[1] Use of the space is consistent with the descriptor "alpha galactosides" and in contrast to use of the hyphen in specific compounds such as "beta-carotene". However these policies affect the alphabetical listing and should be reviewed further.

thus oligosaccharides) need further work and indeed one of the above-mentioned INFOODS working groups is addressing these issues. Thus any terms new to the Component Thesaurus should await the results of these discussions.

It is noted that there is a major problem with the dividing-line between oligo- and polysaccharides. Following FAO (1998), Eurofoods (Schlotke *et al.*, 2000) and Cummings & Stephen (2007), EuroFIR has defined oligosaccharides as DP 3-9, whereas the BLS and INFOODS definitions use DP 3-10. A further complication is that in some disciplines disaccharides are considered oligosaccharides, for example in Belitz *et al.* (2000).

BLS also has a component that is “non-resorbable oligosaccharides”. An entry for "unavailable oligosaccharides", perhaps using the identifier OLSACNA, might be proposed for the Component Thesaurus. However, this will require that the definition of oligosaccharides is resolved and those that are unavailable can be specified. BLS also has the component "available oligosaccharides", which is in the Thesaurus as OLSAC, except that the BLS definition does not match the scope note that specifies DP 3-9.

Polyols

The records for *Sugar alcohols* [GRP_SUGOH] and *Polyols, total* [POLYL] were reviewed as a result of the revisions to the Method Indicator thesaurus for energy calculation methods, as well as in checking that *Glycerol* [GLYRL] (see below) is clearly excluded from this group of components.

For the classification term for *Sugar alcohols* [GRP_SUGOH], the definition was moved from Additional Information to the Scope Note and the Additional Information text was amended to “Polyols are excluded from total sugar, but their contribution to digestible carbohydrate may be reflected in energy calculations required by food labelling regulations. The various energy calculation methods are documented as terms in the Method Indicator thesaurus.”

The Additional Information for the term *Polyols, total* [POLYL] has been amended to read “POLYL can be a contributing value in energy calculation, but some polyols may be assigned differing factors from that applied to polyols in general. For example, Commission Directive 2008/100/EC specifies a lower energy factor for erythritol and a factor of zero is used for it in the corresponding Method Indicator calculation method.”

Such changes in energy calculations affect the definitions of the components contributing to the calculations. Further consideration needs to be given to the handling of polyol values for use in calculations, since it may be necessary to introduce a term for the total of polyols making the standard contribution to energy. Also, a term for *Erythritol* [(2*R*,3*S*)-butane-1,2,3,4-tetrol] may be required. Erythritol is the sugar alcohol derived from the tetrose erythrose. By analogy, glycerol is the sugar alcohol derived from any of the three possible trioses L-glyceraldehyde, D-glyceraldehyde and dihydroxyacetone. Possibly definitions in the area of sugar alcohols need to be further refined. As a result, both glycerol and erythritol might be either included in, or excluded from, the definition.

Lipid components

Individual fatty acids

Terms have been requested for the unsaturated fatty acids “15:1 omega-9”, “17:1 omega-9”, “cis 18:2 (n 6 or 9 not known)”, “19:1” and “24:6”. For all of these, some aspect of the double bond configuration or position is not explicit and the additional detailed terms have been decided on the assumptions noted in this section.

As a generalisation, the existing identifiers for monounsaturated fatty acids are either totally non-specific or specify the configuration. For example, for the C_{17} compounds there are the three codes F17:1, F17:1CIS and F17:1CN8. Apart from the first of these, they specify the *cis*

configuration, with the three-letter CIS form used for total *cis*-heptadecenoic acids. If it is assumed that the omega designation implies *cis* configuration, identifiers can be assigned that are consistent with the other monounsaturate terms. Thus the codes F15:1CN9 and F17:1CN9 have been added for the first two requested fatty acid terms[2], and the scope note "Use for cis-7-pentadecenoic acid" was added for F15:1CN8. Fatty acid 19:1 was added as the code F19:1, with the lack of further detail being compatible with the general codes for other fatty acids. For "24:6", the all-*cis* form is the well-established compound known as nisinic acid. Thus two terms with the identifiers F24:6 and F24:6CN3 were added, allowing the use of whichever best fits the information available.

With regard to the request for "cis 18:2 (n 6 or 9 not known)", the general term F18:2 has the scope note "Use for total 18:2 or if isomer unknown". This is too general as it covers all 18:2 isomers, including those with *trans* bonds or with *cis* bonds not methylene separated. An alternative might be to introduce a term F18:2CIS analogous to F18:2TRS, although at present the latter is defined for the total of *trans* isomers. However, a new convention has been adopted in which the unknown position is indicated by "NX", in this case F18:2CNX. This accurately represents the information implied by the expression "cis 18:2 (n 6 or 9 not known)".

In 2010, BLS requested terms for nonadecatrienoic acid and docosatrienoic acid. The general terms F19:3 and F22:3, respectively, have been added for these. The existing term F22:3CN3 has been moved under F22:3 in the hierarchy. ChEBI does not contain entries for F19:3 or F22:3.

Fatty acid totals

Terms have been requested for C20:>1 and C22:>1, which are interpreted as totals of the polyunsaturated compounds. These have been added with the descriptors "fatty acids, polyunsaturated, total C20" and "... C22", respectively, with the codes F20:P and F22:P, using P to represent *polyunsaturated*.

Requests have also been received for totals of saturated fatty acids as by the representations *C8:0+C10:0*, *C12 to C16 saturated* and *>C18 saturated*. The first, C8:0+C10:0, follows the convention for mixtures and thus the composite code F8:0+F10:0 might be used for identification of the sum of the components. There exist two terms for ranges of saturates, i.e. F4-10:0 and F4-8:0, so that continuing the convention, the new total would use the code F12-16:0 and this has been added to the thesaurus. A proposal for handling *greater than* and *less than* in fatty acid identifiers is documented in Unwin (2010A, p 13), but further discussion and final agreement on this is needed. Under these proposals, the code for *>C18 saturated* would be either F19A:0 or F20A:0, depending whether the starting point is C19:0 or C20:0.

BLS requested terms for short-, medium- and long-chain fatty acids, but their definitions are based on chemical rather than nutritional definitions of chain length. The first two categories include just two fatty acids each and might be represented using the convention for mixtures by C4:0+C6:0 and C8:0+C10:0, respectively. The third total includes saturated and unsaturated fatty acids with a C12 chain and above. The term *fatty acids, total C12 and above* has been added to the thesaurus, using F12:A as a provisional code.

Fats

Requests for fat total unsaturated and fatty acids total unsaturated could be met by again using the convention for mixtures, i.e. FATMU+FATPU and FAMS+FAPU , respectively.

---

[2] It must be assumed that F17:1CN8 was intended as an alternative isomer to F17:1CN9, although there is a slight concern as 8 + 9 =17 and thus F17:1CN8 is *cis*-9-heptadecenoic acid.

Glycerol

The component with the BLS code FG has been called “glycerol + lipoids” in English and it was queried whether the corresponding EuroFIR identifier was GLYRL (glycerol). Previously there was no scope note for this component, which has now been added as “Use for free propane-1,2,3-triol”. However, glycerol is not considered a polyol and, as documented above in the section on *Polyols*, this has been clarified at the entry for *Polyols, total* [POLYL]. Further consideration should be given to the need for a term for Lipoids. If this is introduced, the requested term could be represented using the convention for mixtures.

Phospho- and glycolipids

Specific terms within the *Phospho- and glycolipids* group have scope notes, but the two terms for totals, namely “phospholipids, total” [PHOLIP] and “glycolipids, total” [GLYLIP], did not. In IUPAC/IUB (1997), the term glycolipid is defined as any compound containing one or more monosaccharide residues bound by a glycosidic linkage to a hydrophobic moiety such as an acylglycerol, a sphingoid, a ceramide (N-acylsphingoid) or a prenyl phosphate. A simple definition for a phospholipid is a lipid in which phosphoric acid as well as a fatty acid is esterified to glycerol. These definitions have been used as the basis for provisional scope notes. However, it may also be necessary to consider the methods used for PHOLIP and GLYLIP when the Method Indicator Thesaurus is revised and the conclusions may suggest that amendments are necessary to the scope notes of these “totals” components.

Nitrogen components

Amino acids

The German BLS database reports “essential amino acids” and “non-essential amino acids”. For the Component Thesaurus version 1.2, “amino acids, total essential; eight essential amino acids” [AAE8] was added, corresponding to the equivalent INFOODS tagname. If the value is for the total of isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine, this identifier should be used. If the total conforms to one of the INFOODS tagnames AAE10A, AAE10B or AAE12, the identifier AAE10B exists or the other identifiers should be considered for the EuroFIR Thesaurus. If the definition for the total is unknown, the identifier AAE- should be used. It is noted that the April 2010 INFOODS proposals (INFOODS, 2010) add the codes AAE7 and AAE9 for essential amino acid totals and AAT19, AAT20 and AAT24 for overall amino acid totals.

INFOODS introduced the tagname AANE for “non-essential amino acids”in their 2008 updates (INFOODS, 2008), with the definition "Sum of 11 non-essential amino acids: alanine, asparagine, aspartic acid, cysteine, glutamic acid, glycine, hydroxyproline and hydroxylysine, proline, serine, tyrosine". The EuroFIR identifier AANE can be proposed, but the definition may need to be reviewed. Overall, no new terms for aggregated amino acid values have been added to the Component Thesaurus version 1.3, but the further BLS and INFOODS terms should be considered for version 2.0.

A term has been added for “hydroxylysine”. 5-Hydroxylysine is (2S,5R)-2,6-diamino-5-hydroxyhexanoic acid (CHEBI:18040[3]), which is an important constituent of collagen. The INFOODS tagnames, added in the 2008 update (INFOODS, 2008), is HYL. Although this does not appear on standard lists of amino acid three-letter identifiers, a Google search for “hydroxylysine HYL” showed that the code is in common usage and HYL is used as the EuroFIR identifier.

[3] The structure without the stereochemistry at C-5 specified is CHEBI:60175.

The term Taurine (2-aminoethanesulphonic acid) has been added. The normal classification of taurine as an amino acid has been used, although this is not strictly correct, its acid group being sulphonic rather than carboxylic. The INFOODS tagname for taurine is TAU and this has been used as the EuroFIR identifier. The ChEBI identifier 15891 and the synonym "aminoethylsulphonic acid" were included.

Organic acids

Individual organic acids

Inclusion of terms for pimelic acid and uric acid was requested. Pimelic acid (CHEBI:30531) is 1,5-pentanedicarboxylic acid, i.e. heptanedioic acid, the next higher dicarboxylic acid homologue after adipic acid. The identifier PIMAC was defined.

Uric acid (CHEBI:27226) is not carboxylate, being the purine derivative 7,9-dihydro-1*H*-purine-2,6,8(3*H*)-trione (CHEBI17775) and its tautomers. It is a diprotic acid, but at biological pH only the first ionisation occurs, giving the singly charged hydrogen urate ion. It might be classified as an organic acid or a purine, or both, but provisionally is only classified under "purines" as it perhaps should not be included in "total organic acids" [OA]. The identifier URAC has been defined and the synonyms "lithic acid" and "trihydroxypurine" included.

Vitamins and related compounds

Carotenoids

A term for "Carotenoids excluding β-carotene" has been requested (Ireland *et al.*, 2012), for the total of carotenoids that are considered to contribute half the vitamin A activity contributed by β-carotene. Thus the component excludes non-vitamin A active carotenoids. Such a term is also useful in formulae for the overall activity and for activities expressed as β-carotene equivalents, for example in specifying the calculation equations in the Method Indicator thesaurus. The term has been added, with the descriptor *carotenoids, provitamin A, excluding beta-carotene* and the code CAROTPAXB. This was preferred to the possible alternative identifier, CARTXB, as CAROT represents Carotenoids better than CART, which also is used for Carotene. Provisionally the term appears in the hierarchy immediately below *Vitamin A and related components*, rather than with the individual carotenoids under *Carotenoids*.

Thiamin

It has been noted that Vitamin B1 represents total thiamin, i.e. free thiamin and thiamin phosphorylated forms (thiamin mono-, di- and tri-phosphate), making the definition incorrect. Also, it is generally expressed as thiamin chloride hydrochloride as this is the form of the standard used in all methods. Thus thiamin values are usually presented as mg thiamin chloride HCl per 100g food. It is proposed to add the scope note "Defined as total thiamin, i.e. free thiamin and thiamin phosphorylated forms (thiamin mono-, di- and tri-phosphate)" and to change the Additional Information text to read "Values are normally expressed as thiamin chloride hydrochloride (ChEBI:49105), although values are sometimes expressed as the free thiamin cation (ChEBI:18385) or as thiamin monochloride (ChEBI:33283)". The existing synonym was changed to "thiamin dichloride" (from "thiamin chloride") and the further synonym "vitamin B1" should be added.

Logically there is an argument for using the descriptor "vitamin B1" [VITB1] for values representing the activity of "total thiamin". However, this would be a major change for an important nutrient and it might be difficult to maintain harmonisation with (a) INFOODS tagnames and (b) riboflavin and niacin.

Vitamin C

It has been noted that the descriptor for total vitamin C, namely "vitamin C (ascorbic acid)", is incorrect in referring specifically to ascorbic acid, because dehydroascorbic acid is also included in the component *vitamin C*. The parenthetic note has been deleted from the descriptor.

Minerals

Elements

Inclusion of the elements caesium, lithium, strontium, tin and vanadium was requested. All these are included in the INFOODS lists (caesium in the October 2008 update and all the others in the original publication), using the established policy of atomic symbols for the tagnames. The elements have been added to the Component Thesaurus using the appropriate identifiers CS, LI, SR, SN and V, respectively. It should be noted that the descriptor for CS differs between the EuroFIR and INFOODS terms, being "caesium" and "cesium", respectively. The synonym "cesium" has been included.

Other elements that had INFOODS tagnames, but not EuroFIR identifiers were barium (BA), titanium (TI), silver (AG), gold (AU), antimony (SB), bismuth (BI) and lanthanum (LA). These have been added to the Component Thesaurus and it is suggested that in future the two lists are harmonised in respect of the coverage of elements.

All the new terms for elements have been added within the *Trace elements* category of components.

Speciation

There may be a need to identify specific species of elements, such as oxidation states, charged ions, radicals and isotopes. INFOODS update lists (INFOODS, 2008; INFOODS, 2010) include examples of ionic/oxidation states and isotopes. At a more general level, there may be differentiation between the "inorganic" and "organic" forms of an element.

The INFOODS 2008 and 2010 updates added many tagnames for elemental species such as "sulfur 4+", "selenium 6+" and "silicon 4+", using corresponding tagnames S4+, SE6+ and SI4+. Recently, it was recognised that the plus sign cannot be used in tagnames[4] and one option under discussion is to use Roman numerals. Decisions on the identifiers and corresponding descriptors should be based on accepted chemical conventions.

Three distinct concepts are involved, namely ionic state, oxidation state and oxidation number, and each has a separate convention by which it is represented. Ionic state relates to charged species, cations or anions, that are not bonded covalently to another moiety but are counterbalanced by oppositely charged ions. The ionic charge is represented by the magnitude of the charge preceding the sign, e.g. Fe^{2+} and SO_4^{2-}. Ionic species such as *selenium 6+* do not exist in a normal environment.

Oxidation state and oxidation number are general, closely related concepts that relate to individual atoms in compounds, whether these are ionic or covalent. Oxidation state provides a measure of the degree of oxidation of an atom in a substance (IUPAC, OxState). In contrast to ionic state, it is represented by the sign, positive or negative, followed by the magnitude, e.g. for sulphur it is -2 in hydrogen sulphide and +4 in sulphur dioxide. The term oxidation number relates specifically to the central atom in a coordination entity. It is normally the same

[4] The use of the plus sign might also on occasions clash with the EuroFIR convention of using the form X+Y for mixtures, for example if FE2++FE3+ were to be needed for reporting ferrous plus ferric.

as the atom's oxidation state, although in special circumstances it can differ. It is represented by a Roman numeral.

Although strictly incorrect, it is probably acceptable to use Roman numerals to represent oxidation state in component identifiers, although this leaves the question of differently representing positive and negative oxidation states unresolved. Also there are likely to be confusing cases and indeed ambiguities. The representation of atomic species in the EuroFIR and INFOODS component identification systems will need very careful consideration.

Food properties

BLS requested Component Thesaurus terms for P/S ratio and bread units. P/S ratio, the ratio of polyunsaturated fatty acids to saturated fatty acids, is a property that can be considered for adding to the Component Thesaurus, perhaps with the identifier FAPSRAT. However, the addition needs further discussion and has not been included in version 1.3.

A value for Bread Units[5] provides a measure of available carbohydrate, which is mainly used in Germany, Switzerland and Austria. Different calculation formulae may be used and therefore the use of a component identifier (perhaps BRUNIT) would need to be accompanied by an appropriate Method Indicator. For example, the formula used by BLS is given as "(CHO (carbohydrates, resorbable) - MANTL (mannitol)) / 12.000 + 0.005". An alternative view is that Bread Unit is indeed a unit, used with the identifier CHO for Available Carbohydrate. The thesaurus entries required to document Bread Units also need further discussion.

Components added in version 1.3

The new terms added to the Component Thesaurus for version 1.3 are listed in the following table.

Group	Identifier	Descriptor
Polysaccharides	CHITIN	chitin
	GLUCNB	beta glucans
Fatty acids	F15:1CN9	fatty acid 15:1 n-9 cis
	F17:1CN9	fatty acid 17:1 n-9 cis
	F18:2CNX	fatty acid 18:2, n-6 or n-9 not known
	F19:1	fatty acid 19:1 (nonadecenoic acid)
	F19:3	fatty acid 19:3
	F22:3	fatty acid 22:3
	F24:6	fatty acid 24:6 (tetracosahexaenoic acid)
	F24:6CN3	fatty acid 24:6 n-3 all-cis
Fatty acid totals	F12:A	fatty acids, total C12 and above
	F20:P	fatty acids, polyunsaturated, total C20
	C22:P	fatty acids, polyunsaturated, total C22
Amino acids	HYL	hydroxylysine
	TAU	taurine

[5] See: http://de.wikipedia.org/wiki/Broteinheit

Group	Identifier	Descriptor
Purines	URAC	uric acid
Organic acids	PIMAC	pimelic acid
Vitamins	CAROTPAXB	carotenoids, provitamin A, excluding beta-carotene
Trace elements	AG	silver
	AU	gold
	BA	barium
	BI	bismuth
	CS	caesium
	LA	lanthanum
	LI	lithium
	SB	antimony
	SN	tin
	SR	strontium
	TI	titanium
	V	vanadium

Thus version 1.3 of the EuroFIR Component Thesaurus contains 30 new terms for components. In addition, the information held for many existing terms has been improved.

Method Indicator thesaurus

General issues

Classification terms and hierarchy

In version 1.1 of the Method Indicator thesaurus, general terms such as *Simple summation* [MIR006] were flagged as classification terms. This generated the standard sentence "*This term is for CLASSIFICATION ONLY; DO NOT USE term in indexing. Use a more precise narrower term.*" in the Scope Note. In practice there are circumstances where it is more appropriate to assign a generic term that reports some information, rather than forcing the use of a term that provides no information and may not be true, such as *Analytical or calculation method not known* [MIR003]. Also, where the meaning is unambiguous, the use of a generic term can avoid the need to define extra specific terms, as in the case of most totals of fatty acids as noted below in the section on *Fatty acid totals*. The standard sentence in the Scope Notes field for classification terms has been changed to "*This term is a Broad Term for Classification. It may be used in indexing, but only when there is no more precise narrower term that is appropriate.*" for the version 1.3. The use of the generic terms for indexing normally will relate to the application of a single method and therefore the classification terms have been made singular, with the word *methods* changed to *method*.

However, in the terms for non-indexed and unknown methods, use of the expression *analytical or calculation* raises the question of whether these include imputation methods. Previously, *Imputation* [MIR007] was a child term below *Calculation methods* [MIR002], but this implied that imputation always involves a calculation, which may not be true. The descriptor for MIR003 was changed from *Analytical or calculation method not known* to *Method not known*

and the scope note "*Use when no information is available about whether the method used was analysis, calculation, imputation or recipe calculation*" was added. The position of *Imputation* [MIR007] in the hierarchy has been raised so that it becomes a sibling alongside *Analytical methods* [MIR001], *Calculation methods* [MIR002] and *Recipe calculation methods* [MIR008].

On the other hand, *Recipe calculation methods* [MIR008] might logically be placed as a narrow term under *Calculation methods* [MIR002], although this has not been done in version 1.3. Some definitions might be added for the generic terms calculation, recipe calculation and imputation. This will need further discussion and consideration of the corresponding terms in the Method Type thesaurus. For example, the Method Type term *Calculated as recipe* [R] has the Scope Note "Used for Selected Values obtained by calculation - e.g. recipe calculation - from ingredients (or a single-ingredient raw-to-cooked calculations including Toast from Bread)." and the Additional Information "Further information should normally be reported through food description in the Recipe and Ingredient information.". However, the phrase *e.g. recipe calculation* implies that recipe calculation is not the only possibility and no indication is given of the Method Indicators to be assigned when recipe calculation or other options (e.g. single-ingredient raw-to-cooked calculations) have been used. This area needs to be developed further in collaboration with the work on recipe yield and retention factors.

Non-indexed and unknown methods

There can be circumstances where the method may be known, but it is decided that there is not an appropriate thesaurus term that correctly records the method. In this case, it is not correct to use the term *Method not known* [MIR003]. The Method Type thesaurus includes the term *Other method type* [E] and an analogous term has been added to the Method Indicator thesaurus. By further analogy with the above generalisation of the term *Analytical or calculation method not known* [MIR003] with the change to *Method not known*, the term has been added as *Other method* rather than *Other analytical or calculation method*. Although separate terms could be introduced for *Other analytical method* and *Other calculation method*, this is perhaps unnecessary at this stage as the distinction can be recorded using the Method Type flag. Thus the new term *Other method* [MIR009] has been added with the Scope Notes "Use when the method is known, but no appropriate Method Indicator term can be assigned" and the Additional Information "This term can be used in combination with a Method Type term to give more specific information, for example its use together with Method Type AG (Analytical, generic) indicates "Other analytical method". The need to use the term may indicate that a new specific term should be proposed for addition to the Method Indicator thesaurus."

Recipe calculation methods

The work on verifying data documentation in accordance with EuroFIR standards has indicated problems with the recipe classification in version 1.1. No compilers used the narrower terms, which moreover lack scope notes/definitions. They classified the recipe calculation procedure with four options under the general term *Recipe calculation method* [MIR008], namely:

- EuroFIR recipe calculation procedure [MI0002]
- Recipe level calculation procedure [MI0003]
- Ingredient level calculation procedure [MI0004]
- Other recipe calculation procedure [MI0005]

These apparently categorise on two alternative criteria (Recipe level vs. Ingredient level and EuroFIR vs. Other), it might be more interesting to know if the recipe calculation involved yield factors and/or nutrient retention factors. This is non-judgemental information (as opposed to a "EuroFIR method" stamp), because not all recipe calculations require yield factors and retention factors. Further, the EuroFIR recipe calculation procedure applies weight yields at the recipe level and nutrient retention factors at the ingredient level. Considering this mixed

procedure, the meanings of *Recipe level calculation procedure* and *Recipe level calculation procedure* were not clear.

For version 1.3, recipe calculation terms have been extended and reorganised according to an alternative approach that should allow clearer documentation of the recipe calculation procedure used. The hierarchy is now:

- **Recipe calculation method [MIR008]**
  - Simple recipe calculation [MIR010, new]
  - Recipe calculation with factors [MIR011, new]
    - Recipe level calculation procedure [MI0003]
    - Ingredient level calculation procedure [MI0004]
    - Mixed level calculation procedure [MI0006, new]
      - EuroFIR recipe calculation procedure [MI0002]
  - Other recipe calculation procedure [MI0005]

The term for other recipe calculation procedures, MI0005, has been retained at present, although it might be considered redundant as the situation could be handled by the new term for *Other method* [MIR009], together with the Method Type *Calculated as recipe* [R], as discussed in the previous section, above. The term specifically for the EuroFIR procedure, MI0002, has been retained in version 1.3. Scope Notes have been added for methods MIR0010, MIR0011, MI0002, MI0003, MI0004, MI0005 and MI0006.

Carbohydrate calculations

Energy

Several new types of energy calculation method are required to document total metabolisable energy [ENERC] values as they may be reported, leading to new thesaurus terms. As these reflect requirements for nutritional labelling at different times and in Europe or other parts of the world, it is proposed to include routinely necessary identifying information in each energy calculation method's descriptor, which also involved some revisions to existing descriptors. Thus the text "EC Nutrition Labelling Directive" in the descriptors for MI0108, MI0109 and MI0110 has been amended to "Nutrition Labelling Directive 90/496/EEC" (and the identifier for *polyols, total* was updated to POLYL in the Additional Information formulae).

The new calculation methods are:

Energy (kJ) =
17 x g PROT + 17 x (g CHO - g POLYL) + 37 g FAT + 29 g ALC + 13 x g OA + 10 x g POLYL

This formula for Energy expressed in **kJ** units uses protein calculated from Jones NCF, not 6.25 as in the EEC Labelling Directive 90/496/EEC. The proposed descriptor is *Energy calculated according to Nutrition Labelling Directive 90/496/EEC (kJ, NCF protein exception)* [MI0113].

Energy (kcal) =
4 g PROT + 4 x (g CHO - g POLYL) + 9 x g FAT + 7 x g ALC + 3 x g OA + 2.4 x g POLYL.

This formula for Energy expressed in **kcal** units uses protein calculated from Jones NCF, not 6.25 as in the EEC Labelling Directive 90/496/EEC . The new descriptor is *Energy calculated according to Nutrition Labelling Directive 90/496/EEC (kcal, NCF protein exception)* [MI0112].

Energy (kJ) =
17 x g PROT + 17 x (g CHO - g POLYL) + 37 x g FAT + 29 x g ALC + 8 x g FIBT + 13 x g OA + 10 x g POLYL

This formula for Energy expressed in **kJ** units incorporates all contributions to total metabolisable energy specified in the EC Labelling Directive 2008/100/EC (EC, 2008),

including that of dietary fibre. The proposed descriptor is *Energy calculated according to Nutrition Labelling Directive 2008/100/EC (kJ)* [MI0114].

Energy (kcal) =

4 x g PROT + 4 x (g CHO - g POLYL) + 9 x g FAT + 7 x g ALC + 2 x g FIBT + 3 x g OA + 2.4 x g POLYL

This formula for Energy expressed in **kcal** units incorporates all contributions to total metabolisable energy specified in the EC Labelling Directive 2008/100/EC (EC, 2008), including that of dietary fibre. The proposed descriptor is *Energy calculated according to Nutrition Labelling Directive 2008/100/EC (kcal)* [MI0115].

Energy (kJ) =

16.7 x g PROT + 37.4 x g FAT + 16.7 x g CHO + 29.3 x g ALC

This formula for Energy expressed in kJ units is based on the Atwater calculation (Atwater and Woods, 1896; Atwater and Bryant, 1900) and the unrounded factors cited in FAO (2003), p. 23. This adaption, used in the New Zealand database (Sivakumaran and Huffman, 2011), uses available rather than total carbohydrate and a slightly different factor for alcohol. The new term has the code MI0116 and the provisional descriptor *Energy calculated according to adapted factors (kJ)* [MI0116]. Further discussion will be needed to find a suitable and acceptable permanent descriptor, but method documentation using the code MI0116 will always be linked to the factors specified. A further term may be required as the NZ database also provides an energy value with a contribution from dietary fibre.

When this method is used, calculation of ENERC expressed as **kcal** is performed by dividing the energy content expressed in **kJ** by 4.18 kJ/kcal. Therefore an additional term for kcal factors from proximates is not currently needed, but a new term has been added for Energy (kcal) = Energy (kJ) ÷ 4.18, i.e. *Energy (kcal) calculated from energy (kJ) as kJ/4.18* [MI0117]. Incidentally, FAO (2003) also provides an exact value for carbohydrate as monosaccharide equivalent of 15.7 kJ/g, but at present this variant has not been added as a further term.

The original EEC Council Directive of 1990 (EEC, 1990) on nutrition labelling has had two amendments that impact on energy calculation. EC Labelling Directive 2003/100/EC (EC, 2003) introduced energy factors for salatrims of 6 kcal/g and 25 kJ/g. EC Labelling Directive 2008/100/EC (EC, 2008) added energy factors for fibre of 2 kcal/g and 8 kJ/g. It also noted that the energy provided by the polyol erythritol was less than 0.9 kJ/g (less than 0.2 kcal/g) and therefore it is not appropriate to use the standard factor for polyols as the conversion factor for erythritol. Commission Directive 2008/100/EC is in force from 31 October 2009 and all energy calculations for nutrition labelling must include the contribution from fibre from 31 October 2012. These details have been noted in the Additional Information fields for the energy calculation terms relating to Labelling Directive 2008/100/EC, i.e. MI0114 and MI0115.

Carbohydrate, available

A term is required for the calculation by difference of “available” carbohydrate [CHO], subtracting the value of dietary fibre from that for total carbohydrate [CHOT]. The CHOT value will have been calculated using *Carbohydrate, total, calculated by difference* [MI0131]. Thus the method differs from *Carbohydrate, available, calculated by difference* [MI0183], which involves the subtraction of each non-carbohydrate component. A new term *Carbohydrate, available, calculated from total carbohydrate by difference* [MI0184] has been added. Expressed as component identifiers, the equation is:

CHO= CHOT - FIBT

The Scope Note “Use when total carbohydrate has been calculated by difference using values for the individual proximates protein, fat, water, alcohol and ash” was added for *Carbohydrate, total, calculated by difference* [MI0131].

Lipid components

Individual fatty acids

A question has been raised about which method documentation should be favoured for fatty acids, either the analytical method, e.g. *GLC* [MI1205] or *Fatty acid content calculated on fatty acid profile (%)* [MI0201]. Fatty acids are almost always calculated from a profile (like amino acids) and this is the method used to generate the individual values, so the appropriate Method Indicator is MI0201. When the full documentation specifications are used, the analytical method will be documented by the linked Method Specification record. However, as yet, this is not always used and the analytical method used to determine the total fatty acid content for the food will only be available if that analytical value is also reported. The current Explanation in the Additional Information for MI0201 is probably sufficient. It indicates that the unit conversion is part of the analytical method, which thus is normally documented for the individual fatty acids. However, this guideline may need to be changed if the policy is at any stage further clarified.

Fatty acid totals

A method term has been requested for the calculation of *fatty acids, total cis n-6* [FACN6] by summation of the values for individual *cis* n-6 fatty acids. Also, a method term is required to document the calculation of *fatty acids, total trans* [FATRS] by summation of the values for individual *trans* fatty acids.

There are a considerable number of different fatty acid totals that are obtained by the simple summation of the individual acids that belong to the relevant category. It may not be necessary or helpful to define separate methods terms for each totalled component, but to simply document the calculation method as *Simple summation* [MIR006]. At present, specific summation methods for totalling fatty acids usually are defined when alternative definitions are possible, for example the saturated fatty acids total including and excluding branched chain isomers, MI0209 and MI0208 respectively. It is proposed that in the cases where summations of fatty acid totals do not include alternative definitions, these are recorded using the generic term *Simple summation* [MIR006]. This is in accord with allowing the use of classification terms for indexing in appropriate circumstances, as proposed in the section on *Classification terms*, above.

Nitrogen components

Total nitrogen

A term was requested for values for *Nitrogen, total* [NT] calculated from the *Protein* [PROT] value using a defined Nitrogen Conversion Factor. The factor used should be recorded as a Method Parameter and its source in Method Reference field. The method may be used, for example, when the PROT value is obtained from food label data that will have been calculated from an unreported NT value using the standard factor of 6.25. The new term has been added as *Nitrogen, total, calculated from protein using specific NCF* [MI0127], with the Scope Note "Use for calculation of total nitrogen from protein using a specific Nitrogen Conversion Factor. Record factor used as Method Parameter and its source in Method Reference."

Vitamins and related compounds

Vitamin A

A new term *Vitamin A activity calculated from retinol and beta-carotene (factor 1/12)* [MI0325] has been added for the calculation of vitamin A without the contribution of pro-vitamin A

carotenoids other than β-carotene, to be used when the other carotenoids are not monitored. The calculation is:

VITA = RETOL + 1/12 CARTB

which corresponds to the existing, more comprehensive term MI0324, which is:

VITA = RETOL + (CARTB/12) + (other pro-vitamin A carotenoids / 24)

A term *Beta-carotene equivalent calculation including alpha-carotene* [MI0304] has been added for the calculation of beta-carotene equivalents by multiplying only α-carotene by 0.5 and adding β-carotene, in contrast to MI0301, which is *Beta-carotene equivalent calculation including alpha-carotene and cryptoxanthins*.

Matrix Unit thesaurus

A new term *per unit* [U] has been added to the Matrix Unit thesaurus. Its use is described by its Scope Note, which is "Use as the matrix unit for values that are expressed for a unit of the food item (e.g. per tablet, per drop, per teaspoon), not as a metric measure of the food". The Additional Information is "Food items such as dietary supplements may have their composition expressed 'per unit'. The 'unit' measure should be explained in the Food Remarks field, with more specific information in the Value Remarks field if more than one non-metric 'unit' measure is used in the values for the food item."

References

Atwater, W.O. and Bryant, A.P. (1900). *The availability and fuel values of food materials*. Connecticut (Storrs) Agricultural Experiment Station 12th Annual Report, 1899.

Atwater, W.O. and Woods, C.D. (1896). *The chemical composition of American food materials*. US Official Experiment Stations, Experiment Station Bulletin No. 28. Washington, DC.

Belitz, H.-D., Grosch, W. and Schieberle, P. (2000). *Food chemistry*. 3rd revised edition, Springer, Berlin.

Cummings, J.H. and Stephen, A.M. (2007). Carbohydrate terminology and classification. *Eur. J. Clin. Nutr.*, **61** (Suppl 1), S5–S18.

EC (2003). Commission Directive 2003/120/EC of 5 December 2003 amending Council Directive 90/496/EEC on nutrition labelling for foodstuffs. *Official J. Eur. Union*, L 333/51. http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_333/l_33320031220en00510051.pdf, accessed 2010-05-19.

EC (2008). Commission Directive 2008/100/EC of 28 October 2008 amending Council Directive 90/496/EEC on nutrition labelling for foodstuffs as regards recommended daily allowances, energy conversion factors and definitions. *Official J. Eur. Union*, L 285/9-285/12. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:285:0009:0012:EN:PDF, accessed 2012-02-16.

EEC (1990). Council Directive 90/496/EEC of 24 September 1990 on nutrition labelling of foodstuffs. *Official J. Eur. Union*, L 276/40-276/44. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1990:276:0040:0044:EN:PDF accessed 2012-02-23.

FAO (1998). *Carbohydrates in human nutrition.* FAO Food and Nutrition Paper 66, Food and Agriculture Organization of the United Nations, Rome. http://www.fao.org/docrep/W8079E/W8079E00.htm, accessed 25 January 2012.

FAO (2003). *Food energy – methods of analysis and conversion factors*. FAO Food and Nutrition Paper 77, Food and Agriculture Organization of the United Nations, Rome. ftp://ftp.fao.org/docrep/fao/006/y5022e/y5022e00.pdf. Downloaded 2010-05-04.

INFOODS (2008). http://www.fao.org/infoods/TAGREV.xls, accessed 24 January 2012.

INFOODS (2010). http://www.fao.org/infoods/Tagname%20new%20April%202010-web.xls, accessed 25 January 2012.

Ireland, J., Westenbrink, S., Oseredczuk, M. and Møller, A. (2012). *EuroFIR Compiler Questionnaire 2011: Pitfalls and Challenges in Reporting Food Composition Data*. EuroFIR Nexus - Report on the EuroFIR Compiler Questionnaire 2011, dated 2012-01-04.

IUPAC (OxState). Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook

IUPAC/IUB (1997). Nomenclature of glycolipids (IUPAC Recommendations 1997). IUPAC Commission on the Nomenclature of Organic Chemistry and IUPAC-IUB Commission on Biochemical Nomenclature. Pure Appl. Chem., **69** (*12*), 2475-2487. http://www.chem.qmul.ac.uk/iupac/misc/glylp.html, accessed 25 January 2012.

Klensin, J.C., Feskanich, D., Lin, V., Truswell, A.S. and Southgate, D.A.T. (1989). Identification of food components for INFOODS data interchange. The United Nations University, Tokyo, Japan. (See also: http://www.fao.org/infoods/tagnames_en.stm)

Schlotke, F., Becker, W., Ireland, J., Møller, A., Ovaskainen, M.-L., Monspart, J. and Unwin, I. (2000). *EUROFOODS Recommendations for Food Composition Database Management and Data Interchange*. COST Action 99 – EUROFOODS Research Action on Food Consumption and Composition Data. COST Action 99 Report EUR19538, European Commission, Luxembourg. See also: *J. Food Compos. Anal.*, **13**, 709-744.

Sivakumaran, S. and Huffman, L. (2011). *New Zealand FOODfiles 2010 Manual*. New Zealand Institute for Plant & Food Research Limited, Palmerston North, N.Z., 2011.

Unwin, I.D. (2010A). *EuroFIR component thesaurus report. Report on the preparation of version 1.1*. EuroFIR Deliverable D1.8.33; submitted 12 May 2010.

Unwin, I.D. (2010B). *EuroFIR component thesaurus report. Report on the preparation of version 1.2*. EuroFIR Deliverable D1.8.33; submitted 21 June 2010.