



FP6-CT-2005-513944

EuroFIR

European Food Information Resource Network of Excellence

Network of Excellence

Sixth Framework Programme Priority 5 – Food Quality and Safety

**Deliverable D1.6.7
Position Paper on the Eurocode 2 Coding System**

Due date of milestone or deliverable: April 2006

Actual submission date: December 2006

Start Date: January 01, 2005

Duration: 5 years

Organisation name of lead contractor for this deliverable or milestone: IDUFIC

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
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 Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

EUROCODE 2

POSITION PAPER ON THE EUROCODE 2 CODING SYSTEM

PREPARED BY

IAN UNWIN

FOOD INFORMATION CONSULTANCY

IN COLLABORATION WITH

ANDERS MØLLER

DANISH VETERINARY AND FOOD ADMINISTRATION

AND

JAYNE IRELAND

AFSSA, FRANCE

July 2006

Contents

Contents.....	ii
Position Paper on the Eurocode 2 Food Coding System.....	1
Summary	1
1. Background	3
1.1. Eurocode 2 evaluation, 1993	4
1.2. Brief summary of Review report.....	4
1.3. Workshop discussions	5
1.4. Components of the Eurocode system	6
1.5. Food description systems	7
1.5.1. Eurocode Descriptor System	7
1.5.2. LanguaL Thesaurus	7
1.5.3. INFOODS description.....	8
1.5.4. Description of mixed foods	8
1.6. Other food classification systems.....	9
1.6.1. CIAA/GSFA Food Categorization System	9
1.6.2. United Nations Statistics Division systems.....	9
1.7. Core Classification development.....	10
1.8. Overview of terminology	11
2. Food categorisation in data management	12
2.1. Categorisation, classification and grouping	13
2.2. Categorisation, identification and description.....	14
2.3. Advanced category management.....	15
2.3.1. Category identification	15
2.3.2. Category overlap and multiple categories	16
2.3.3. Multilingual categories.....	17
3. Potential applications for the Eurocode system.....	17
3.1. Recording of food consumption data	18
3.2. International comparison of consumption data	18
3.3. Aggregation and disaggregation of consumption data	19
3.4. Harmonisation of household budget survey data	19
3.5. Matching of consumption and composition data.....	19
3.6. Managing food group related food property data.....	20
4. Revision of the Eurocode 2 classification	21
4.1. Summary of Main Group revisions	21

4.1.1.	Main Group 1: Milk and milk products.....	22
4.1.2.	Main Group 2: Egg and egg products.....	22
4.1.3.	Main Group 3: Meat and meat products.....	23
4.1.4.	Main Group 4: Fish and fish products.....	24
4.1.5.	Main Group 5: Fats and oils	24
4.1.6.	Main Group 6: Grains and grain products.....	24
4.1.7.	Main Group 7: Pulses, seeds, kernels, nuts and products.....	25
4.1.8.	Main Group 8: Vegetables and vegetable products.....	25
4.1.9.	Main Group 9: Fruits and fruit products.....	25
4.1.10.	Main Group 10: Sugar, sugar products, chocolate products and confectionery	26
4.1.11.	Main Group 11: Beverages (non-milk)	26
4.1.12.	Main Group 12: Miscellaneous, soups, sauces, snacks and products	26
4.1.13.	Main Group 13: Products for special nutritional use.....	27
4.2.	New Eurocode category codes	27
4.2.1.	Eurocode version 99/1 category codes	28
4.2.2.	Code structuring and special codes	28
4.2.3.	Future enhancement of codes	29
4.3.	Categorisation policies	30
4.3.1.	Categorisation and multiple classification.....	30
4.3.2.	Grouping categories.....	31
4.3.3.	Food source versus product type	31
4.3.4.	Relationship between categories and descriptors	32
4.3.5.	Classification hierarchies.....	33
5.	The overall Eurocode system	34
5.1.	Descriptor System	34
5.1.1.	Requirements.....	34
5.1.2.	Relationship to LanguaL	37
5.1.3.	Relationship to INFOODS system	38
5.2.	Recipe System and recipe coding.....	38
5.3.	Documentation	40
5.3.1.	Website overview	40
5.3.2.	Main Group policies and listings.....	41
5.3.3.	Category documentation.....	42
6.	Recent developments.....	42
6.1.	EFG (Euro Food Groups) proposal	42

6.2.	Eurocode and LanguaL.....	43
6.3.	EuroFIR Food Classification.....	43
7.	Further and related development.....	43
7.1.	Current and short-term tasks	44
7.1.1.	Testing the Eurocode 2 categorisation	44
7.1.2.	Nutrient Loss or Gain applications.....	45
7.1.3.	Eurocode 2 categorisation software tools.....	45
7.2.	Further development.....	46
7.2.1.	Enhancing the classification	47
7.2.2.	Designing other system components	47
7.2.3.	Developing the documentation.....	48
7.2.4.	Development of software tools.....	49
8.	Conclusions and recommendations	50
9.	References	52

Position Paper on the Eurocode 2 Food Coding System

A review of the objectives, current specification, and options for future development and implementations

Summary

The Eurocode 2 Food Coding System for recording and comparing food consumption data was evaluated in 1993 but not further revised until a review and report in 1998 (Unwin & Møller, 1998). This proposed revisions to the higher levels of the hierarchy in the so-called modified Core Classification and was discussed at a COST Action 99 – Eurofoods workshop in October 1998. The workshop concluded that the revision should be extended within the classification, leaving development of the separate descriptor system and recipe management to different forums. As part of the further work, a Website was constructed so that the revisions and other information were available for review and comment as they were produced. Progress was reviewed at a workshop in May 1999 at which it was concluded that it would not be desirable to completely revise all aspects of the system by the end of the COST Action 99. Therefore it was decided that the final deliverable within that programme should be a Position Paper (this document) that reported final conclusions of the review and the further revision of the Core Classification, together with the options for future development. The revisions, version 99/1, and a draft of this paper were discussed at a further workshop in December 1999. Suggested changes have been implemented as version 99/2 of the Core Classification which now covers three levels of the classification hierarchy.

The Position Paper introduction describes these stages of the review project in more detail and summarises the current status of related food classification and description systems. Potential applications of Eurocode are discussed, including the recording of consumption data, the international comparison of consumption data and matching the data with food composition data. It might also be used to provide more detailed standard categories in Household Budget Survey data and as a food grouping in deriving typical values for nutrient loss and gain factors.

The version 99/1 and 99/2 revisions are described for each main group, together with a few issues which remain to be agreed and some general considerations identified in developing the categorisation policies. The latest changes have proposed the use of standard set of subcodes to record information such as the source species in milk products and type of basic ingredient for products in the miscellaneous foods group. The revised specification of codes allows more flexibility in updating the classification and includes reserved series for special purpose codes. Some uses of these are proposed, including their application in structured codes used with standard subcodes to provide flexible options for coding and retrieving food records. They also provide a mechanism for managing alternative grouping categories if these are required when data are recorded

Revision of the Eurocode 2 main groups has aimed to produce a rigorous categorisation system within a basic and broadly acceptable classification hierarchy. It is envisaged that separate, application-specific classification hierarchies will be superimposed on the underlying categorisation system, particularly for the comparison and aggregation of data. A simple grouping system called the Euro Food Groups (EFG) was proposed at the December 1999

workshop, initially to be tested for mapping from, and comparison of, national food consumption data sets but potentially to form an aggregation grouping for Eurocode 2 coded food records.

Food information management is discussed in terms of the roles of categorisation, classification, grouping, food description and food identification. It is concluded that the development of Eurocode 2 as a detailed categorisation system will allow it to make a significant contribution to the 'identification' of recorded food items. In this, it should form part of an overall system that includes support for further food description including recipe information that identifies and quantifies the ingredients. These components must be designed to work effectively together and thus should form parts of a coordinated development effort. A key aspect of this will be the development and dissemination of software tools for the coding of food records. This should be allied with a continuing effort to build effective documentation, and to make this and proposals for Eurocode enhancements and revisions widely available for review and the feedback of comments. The review project has established a basis for this through the development of the Eurocode Website at <http://www.vfd2.dk/eurocode/>. This will be used to post future updates to the Eurocode 2 category listing and additions to the documentation such as a food category index.

It did not prove possible to publish this Position Paper in 2000. However, the commencement of the EuroFIR project in 2005 (see section 6.3) provided the opportunity to update the text, particularly sections 6 and 7, and make the final version available as a EuroFIR deliverable. Although the Eurocode 2 classification is directed at the handling of food consumption data, the underlying categorisation described in this paper has made a significant contribution to the development of the EuroFIR Food Classification system for food composition data. That work has recognised the importance of documentation to the development and widespread implementation of such systems, and the Eurocode documentation provides a useful starting point for an extended resource describing food classification and description categories in Europe.

Introduction

This report was completed on behalf of the **European Food Information Resource (EuroFIR)** Consortium, funded under the EU 6th Framework Food Quality and Safety Programme, project number FP6-513944. EuroFIR, the world-leading European Network of Excellence on Food Composition Databank systems¹ is a partnership between 40 universities, research institutes and small-to-medium sized enterprises (SMEs) from 25 countries. EuroFIR aims to develop and integrate a comprehensive, coherent and validated databank providing a single, authoritative source of food composition data for Europe. The goal of EuroFIR Work Package on Food Identification and Description (IA1.6) is to establish a common standard for the identification and description of foods in European food composition databases that allows for application of state-of-the-art concepts in database linking and management and their comparability as well as the comparison and interchange of food composition data.

1. Background

The Eurocode 2 Food Coding System was developed under the European FLAIR Eurofoods-Enfant Project to provide a standard tool for nutritional surveys and to serve the need for food intake comparisons (Poortvliet *et al.*, 1991; p 4). Version 93/1² was completed in 1993, was described in draft documentation (Poortvliet. & Kohlmeier, 1993), and was the subject of an evaluation exercise (van Kappel, 1993A, 1993B). The evaluation concluded that further work was required to make Eurocode 2 a practical solution for the needs of recording and comparing food consumption data, but no further development was undertaken.

In 1998 the Danish National Food Agency (now the Danish Veterinary and Food Administration) initiated a review of the Eurocode 2 system to establish the options for finalising it as a practical aid to recording and handling food consumption data. The existing system was reviewed and possible changes were prototyped through the drafting of a modified Core Classification covering the subgroup level immediately below the main groups in the Eurocode 2 hierarchy. The work was reported (Unwin & Møller, 1998) and presented at a COST Action 99 – Eurofoods workshop in October 1998.

Following interest shown at the workshop and the decisions it took on the acceptable options for modifying Eurocode 2 (see section 1.3), further work was undertaken to apply these options and produce a revised version of Eurocode 2 as a suitable conclusion of the COST Action 99 work. This Position Paper reports this current version, together with the objectives, requirements and options for implementation and further development.

¹ URL: <http://www.eurofir.net>

² Poortvliet *et al.* (1992) refers to Eurocode 2 version 91/2 but the review project has used the listing incorporated in the 1993 draft manual used in the evaluation exercise. This has been designated 93/1 to match its documentation date and because it is not certain if it is identical to version 91/2.

1.1. Eurocode 2 evaluation, 1993

As reported in the Eurocode 2 Review report (Unwin & Møller, 1998; p 5), the Eurocode 2 food coding system and its documentation were evaluated in an experiment conducted by the International Agency for Research on Cancer in Lyon (van Kappel, 1993A, 1993B). Test material and the documentation were studied by 49 respondents in 16 European countries who coded lists of general foods, local foods (for 13 countries) and a recipe, and completed a questionnaire on the system. It may be useful to repeat here the summary of the findings contained in the Review report:

"The results were good for the coding of the European foods down to the sub-group level. The coding of local, commercially available mixed foods gave problems, as did recipe coding and the assignment of descriptors.

Participants found the classification to be incomplete, particularly in the Meats group, and to fail in recording the type of fat used to prepare foods. It was considered not specific enough, particularly as a basis for nutrient calculation, and to need improvement by the addition of extra generic names and mixed foods, in the principles of recipe coding, in the principle and completeness of the descriptor system, and in the manual. Concern was also expressed about the organisation required to maintain the classification system, for example in allocating new food and recipe codes."

1.2. Brief summary of Review report

The Eurocode 2 Review report (Unwin & Møller, 1998) documented a review of the food coding system and identified possible modifications to be incorporated into a revised 'Core Classification'. The report's introduction noted some source documents used in the review on Eurocode and its 1993 evaluation, briefly described the overall Eurocode system, and summarised alternative approaches to food coding and description. The objectives, the overall structure and content, and the (then) current documentation of the Eurocode 2 food coding system were described. Each of the Eurocode 2 main groups was reviewed and its modified Core Classification presented. This listed categories at the sub-group level, incorporating some changes suggested from the review and providing scope notes and examples of the food item level as appropriate. The final chapter presented conclusions from the review of the version 93/1 classification and the preparation of the core classification, and made some recommendations on the short-term and longer-term development of the overall Eurocode system. Appendices gave a full listing of the Core Classification (later designated Eurocode 2 version 98/1) and described the computer version produced as a Microsoft Access database.

The review considered that although the version 93/1 draft documentation specified policies for coding foods, these were sometimes complex, not always clear and often contradictory in the examples given. Some problems stem from conflicts in the basic principles applied in the classification, a particular example being the alternative criteria of division by source organism and product type used in different main groups. Another source of difficulty was the relative lack of development of the descriptor and recipe systems. As a result, additional and unnecessary roles were assigned to the classification which was viewed in the original design as the central (and, for many implementations, the only) element of the overall system. A clear example is the routine appearance of a *for dietetic use* subgroup in each main group. If this approach is used to record the consumption of a food, all detail on it below the main-group level is discarded. On the other hand, if this information were recorded by including a *Use context* facet in the Descriptor System, the food can be classified to the same detail as any other consumed item. Complications

in the classification and coding of mixed foods and dishes arise in part because information that might be better handled by a listing of ingredients is encoded within a single 'classification' code.

The modified Core Classification concentrated on establishing the nucleus of a classification and on identifying constraints arising from the design of the hierarchy and the codes used to record the classification of foods within it. This initial work retained two specific constraints which could be removed in further revisions, namely that categories were still assigned within the shallow, three-level hierarchy and that foods were not switched from one to another main group. Section 1.7 further discusses the Core Classification development as a prelude to Chapter 4 which reports work on the 99/1 and 99/2 revisions. First, however, other developments since the publication of the Eurocode 2 Review report and other work related to the overall Eurocode food coding system are summarised.

1.3. Workshop discussions

Three papers were presented to a workshop held at the COST Action 99 – Eurofoods workshop held in Wageningen on 25 October 1998, one documenting the Modified Core Classification, another summarising the findings of the report and a third detailing necessary decisions to be taken.

The first decision to be taken was formulated as "existing codes should not be changed when the classification is updated". This would retain compatibility with data already coded. To make this acceptable and practical would need the use of computer coding support that would minimise the significance of storage codes by applying separate mechanisms for the maintaining of a logically ordered classification. This formed the basis for the second decision that "future modifications of Eurocode need not assume that coders will directly use the classification's storage codes". However the workshop felt it unlikely that any key collection of data currently existed that needed compatibility with the version 93/1 codes to be maintained. Therefore it was decided that unless any definite requirements to retain the existing codes could be identified, new codes (and thus a new, more flexible system of codes) could be introduced during the current revision of Eurocode 2, although once this version is finalised, codes should then be fixed. As a third decision point, it was proposed that "the number of hierarchical levels available in the Eurocode 2 classification should be increased". It was generally agreed that this was necessary.

Further discussion covered the use of Eurocode 2 in relation to the handling of household budget survey and food balance sheet data, with the use of 'Commodity trees' and collaboration with EUROSTAT on food balance sheets being suggested. Thus it was suggested that the Eurocode 2 system might move in the future towards being a classification of commodities rather than of foods 'as consumed' which is the current premise. Changes to commodity foods to their state 'as consumed' would be recorded through descriptor and recipe systems.

Future work on the revisions to Eurocode 2 was discussed. It was suggested that this should concentrate on producing an improved classification, leaving development of the descriptor and recipe systems to separate discussions. A final version of the Core Classification should be produced. This would expand and modify version 98/1 using additional hierarchical levels, development of clear policies on categorisation which might involve moving items between main groups, and changing codes as necessary to produce a system that supports logical ordering and the insertion of new codes. The Core Classification should be documented with scope notes to define its categories where necessary. The work would also investigate how they might be

further subdivided when the lower levels of the classification are defined and how the coverage of the classification categories might determine requirements for additional information recorded through associated descriptor and recipe systems. Drafts of the documentation would be made available on the Internet for those actively reviewing the proposals but its availability should not be more widely publicised until a later stage in the work.

As part of this follow-up work, Eurocode 2 revision was included on the agenda of a workshop held by the COST Action 99 – Eurofoods Working Group on Food Description in May 1999. Some general aspects of the classification were discussed, in particular the use of a broad and shallow hierarchy, and the alternative of a narrow and deep one, i.e. with many levels but fewer categories at each level. It was also agreed that a Position Paper (this document) should be prepared as a status report at the end of the COST Action 99 project and as a review of the options for future development. A draft of the paper was presented at a further working group meeting in December 1999. This discussed in some detail the options for developing a system for the harmonised reporting and comparison of food consumption data internationally using roughly 30-40 food groups and proposed the EFG (Euro Food Groups) categorisation for this. This and further comments from the workshop have been incorporated into this final version of the Position Paper.

1.4. Components of the Eurocode system

Notwithstanding the Wageningen workshop comments on the separate development of the descriptor and recipe systems, the Eurocode 2 Review report (Unwin & Møller, 1998; p 39) considered that the integration of the various components of the overall system was the key to the development of an effective, useable and well documented implementation. Further work has concentrated on the specification of categories within the Eurocode 2 classification but this has been considerably helped by the assumption that much of the classification and coding policy for mixed foods should be defined separately as part of recipe support. The 'X' designation for mixed foods is not included in the Core Classification. Before mixed food identification is supported (either through classification coding, the recipe system or as a descriptor), exact requirements for using the information to record, retrieve and aggregate food items need to be defined.

The possibilities for increased use of a descriptor system to code information on industry-processed foods as well as foods as consumed were well illustrated in proposals submitted from Norway (Løken & Borgejordet, 1998) in response to the Wageningen workshop papers. Their classification applied a limited deepening of the classification hierarchy, adding an extra level between the subgroups and the specific food items, but it also sought to reduce the number of classification categories by including information on food constituents³ added or removed as descriptors.

³ In this context, the term food constituent is preferred to food component. It is convenient to use food component for measured or calculated substances, the columns in a food composition table. In contrast, the term food constituents can be used for ingredient foods, the rows of a table. Some substances such as food additives can assume either role. If material is removed from a food, a symmetrical viewpoint would suggest this should be considered 'constituent removed' rather than 'component removed' and that the concepts of addition and removal might be

The third key adjunct to the Eurocode 2 classification is the documentation, both for the classification itself and for the way that classification coding relates to the other components discussed, namely the recipe and descriptor systems. Ideally the documentation should be built in parallel with the development of the system itself so that the description and examples accurately reflect the intentions behind the system structure and content that are implemented.

A further essential consideration for recording and processing food consumption data is the quantitative aspect, how much of a food was consumed. There is no direct documentation of this aspect and how the weight recorded should relate to the classification code and descriptors assigned. However the version 93/1 documentation includes some oblique references to the issue, for example in one of the two (potentially ambiguous) documented uses of the descriptor *Uncooked (raw)*.

1.5. Food description systems

In discussing the options for recording food description that might be used to supplement the information recorded by the Eurocode 2 classification, three systems that have been developed for use internationally and implemented, at least on a test basis, will be considered. These are:

- the Eurocode Descriptor System (Poortvliet *et al.*, 1992; Unwin & Møller, 1998)
- the LanguaL faceted food description thesaurus (Hendricks, 1992; Møller & Ireland, 2000)
- the INFOODS faceted free-text food description system (Truswell *et al.*, 1991).

A comparative review of the use of the LanguaL and INFOODS systems in food composition databases (Burlingame, 1998) has concluded that "no system is adequate for adoption/endorsement as the international standard in its present form". Inclusion of the Eurocode Descriptor System in the review might not have changed this conclusion significantly.

1.5.1. Eurocode Descriptor System

The Eurocode 2 classification was originally intended to classify foods 'as consumed' but it was found necessary in some circumstances to record additional description of their processing or preparation. A simple Descriptor System of approximately 80 terms assigned to 5 different facets was introduced; this is reviewed in Unwin & Møller (1998), pp 11-12. The review of Eurocode has identified various areas where information previously incorporated in the classification would be better handled through extensions to the Descriptor System (see also section 5.1.1.).

1.5.2. LanguaL Thesaurus

The LanguaL thesaurus (Møller & Ireland, 2000) provides a controlled language for describing foods through various facets (sometimes referred to as factors) which cover particular aspects such as the product type, cooking method, etc. Each facet has a vocabulary of terms that are organised hierarchically and are associated with a simple (non-hierarchical) code that is used to record the terms assigned to a food item, and in the retrieval of food items. Each LanguaL term represents a

consolidated in data management within a single listing of ingredients that both reported constituents added and constituents removed.

category to which a food may belong and the assignment of LanguaL codes records detailed description for the food.

LanguaL is relevant to the Eurocode review in several ways. Firstly it provides an option for the design of the Eurocode Descriptor System. The latter might use the LanguaL thesaurus fully, might use a subset of LanguaL, or might simply use LanguaL as a model for its design. Secondly, the information recorded by the Eurocode 2 classification roughly corresponds to that covered by LanguaL facets A (Product Type), B (Food Source) and C (Part of Plant or Animal), although LanguaL terms allow more extended description of plants and animals. These aspects are important in defining the policies used in the Eurocode classification and thus LanguaL provides a useful analogy when discussing these policies. Finally, as noted above, there is as yet no adequate general solution for describing (and, by implication, identifying) foods. Current proposals for LanguaL facet A reformulate this to accommodate alternative classifications, with the Eurocode 2 classification included as one of these. This Position Paper provides a basis for further discussion of options for improved methodologies for recording information on foods using the LanguaL, Eurocode, and indeed INFOODS systems.

1.5.3. INFOODS description

The INFOODS food description system provides a wide range of facets that can be used to describe foods, including details that are relevant to the food sampling, its context of use, etc. The information is recorded in free text, allowing the recorder to use their natural terminology for the concepts involved and to provide the level of detail they find appropriate. This provides for the effective recording of information for local use but does not standardise the language for wider, international interpretation and for efficient retrieval.

1.5.4. Description of mixed foods

All three of these systems propose procedures for describing mixed foods in terms of their recipe, including the identity of the ingredients. These have not, however, for the most part been implemented and have distinct shortcomings, particularly if it is intended to use the information in the calculation of a dish's composition from that of its ingredients.

The systems generally distinguish between single and mixed foods, although the benefits of this approach should be seriously questioned. The main rationales are that the description of a mixed food is based largely on its recipe information and that other description facets differ between single and mixed foods. However the concept of a single-ingredient recipe is very useful in data management since it provides an integrated approach to handling yield and nutrient retention factors as well as the description of cooking method. While it is true that a number of facets are only appropriate to single foods, many are relevant to both types and often other facets may not be used in either case. Any division between single and mixed foods, as in the INFOODS system, should be considered a pragmatic convenience, for example in the design of input screens, rather than a hard-and-fast separation into distinct categories. In practice this is how the INFOODS system can be interpreted, as a choice between sets of questions, but the emphasis on the problem in distinguishing the two types makes this unclear.

The documentation of the INFOODS system provides an example of a recipe. However, because the information is only implicitly structured (as text for the ingredient name followed by the amount), it is not designed for use by a food DBMS in calculating recipe results. To do this requires that an ingredient is identified as another food item in the database that will provide

compositional values, and that the amount is either expressed in grams or that the measure in which it is expressed can be converted to grams. The *General Introduction to the LanguaL Thesaurus*, version 2000, describes so-called Full Ingredient Indexing, which will provide ingredient identification information. Simple LanguaL indexing only provides the facility to record the main ingredient using facets A to C, plus the indexing of additional ingredients in facet H (Treatment applied) as 'Food added' terms.

1.6. Other food classification systems

As noted above, LanguaL facets A, B, and C provide related classifications to the Eurocode 2 system for recording information on foods in food composition tables and as consumed. Classifications of food have been, or are being, developed to serve other purposes such as grouping foods for food additive information and the reporting of trade and other economic statistics. This section supplements a recent more detailed discussion (Ireland & Møller, 2000)

Although the various categorisations may be based on differing criteria and may be elaborated to different levels of specificity, there could be benefits if Eurocode is kept as compatible as possible with other systems. Since Eurocode (and LanguaL) are likely to be more specific than most other systems, this will involve checking whether Eurocode, including its descriptor system, can record the information that differentiates categories in the other systems. Then a judgement can be made in any particular case whether it is useful for Eurocode to do so. The primary objective should remain the ability to map the detailed Eurocode system to closely related aggregations such as the EFG system, described in section 6.1, but the ability to translate Eurocode coding into other systems may allow valuable comparisons to be made with other statistics. Potentially it could also be the route to a wider adoption of Eurocode as an underlying, more detailed classification in a range of fields.

1.6.1. CIAA/GSFA Food Categorization System

The CIAA (Confederation of the Food and Drink Industries of the EEC) developed a food categorisation system to support the authorisation of food additives at the European Union level. The Codex Alimentarius Commission decided that this system should form the basis of that proposed for the Codex General Standard on Food Additives which is referred to as the GSFA food categorization system (Codex Alimentarius Commission, 1999).

In some areas such as dairy products the GSFA categorisation is similar to Eurocode and the compatibility could be further improved. Where the Eurocode classification is based on source organism rather than product type, the classification criteria diverge because the GSFA system, serving the requirements of food additive work, introduces a high level subdivision between fresh and processed products.

1.6.2. United Nations Statistics Division systems

The United Nations Statistics Division supports various classification systems for reporting data on economic statistics, and demographic and social statistics. Within the economics field, the so-called 1993 System of National Accounts (SNA) outlined four functional classifications, of which COICOP (Classification of Individual Consumption by Purpose) is most relevant. The classifications (OECD, 1998) share a common structure in which the top level is termed the

Division, the next level the Group and the final level is called the Class. As well as for the SNA, COICOP is intended for use in the further statistical areas of household budget surveys, consumer price indices and international comparison programmes. Only Division 01, *Food and non-alcoholic beverages*, and Division 02, *Alcoholic beverages, tobacco, etc.*, relate directly to food products.

The SNA system uses the related Central Product Classification, CPC (United Nations, 1998) for the detailed study of goods and services. It is stated that "The main purposes of CPC are to provide a framework for the international comparison of statistics dealing with goods, services and assets and to serve as a guide for developing or revising existing classification systems of products in order to make them compatible with international standards. ... It provides a basis for recompiling basic statistics from their original classifications into a standard classification for analytical use."

CPC Section 2 covers FOOD PRODUCTS, BEVERAGES AND TOBACCO; TEXTILES, APPAREL AND LEATHER PRODUCTS and the relevant Divisions are 21 - *Meat, fish, fruit, vegetables, oils and fats*, 22 - *Dairy products*, 23 - *Grain mill products, starches and starch products; other food products*, and 24 - *Beverages*. Further subdivision uses a restricted number of categories, for example *Dairy products* is broken down into just two Groups, 221 - *Processed liquid milk and cream* and 229 - *Other dairy products*. These are broken down into two and nine Classes, respectively, which correspond very approximately to the Eurocode *Milk and Milk Products* subgroups.

1.7. Core Classification development

Although the 1993 evaluation of the Eurocode 2 system suggested that single foods could be coded fairly accurately at the subgroup level, it also highlighted the need to have classification categories available which matched more closely with the specificity of information commonly reported for a food. The absence of categories for, for example, 'Chicken' (part unspecified) and 'Pasta' (presence or absence of egg unknown) means that these reported foods must be coded as 'Poultry' and 'Wheat, grain/miller products, other use', respectively.

Many such problems could be solved by adding extra levels within the hierarchy, but any radical revision of this nature would involve a large amount of work. It could only be justified if it were based on well-founded principles and if it were likely that the results would be widely accepted and implemented. Therefore the initial work, as reported in the Eurocode 2 Review report, had the much more limited objective of identifying changes to Eurocode at the subgroup level that might improve the information coded within the constraints of a system using only the three levels of main group, subgroup and food item. The resulting proposals were for a so-called Modified Core Classification, now also referred to as Eurocode 2 version 98/1.

As noted earlier in the section on workshop discussions, it was agreed to further revise the Modified Core Classification in the light of the decision that the number of hierarchical levels could be increased and that the classification codes could be changed. Documentation would also be developed for the revised classification and the work would be made available on the Internet for review by interested parties. This stage of the work was completed as Eurocode 2 version 99/1. Final comments on this version have been incorporated as further revisions and the result is Eurocode 2 version 99/2 which represents the final outcome of the work under the COST Action

99 project. Further tasks are defined in section 7, and cover some final work that is required to produce Eurocode 2000, the first fully operational release.

Development of the Core Classification has provided a basis for future improvements to the Eurocode 2 food coding system. Any such further work should be closely allied to related developments in the areas of food description and food data management. It must also be undertaken with clear objectives and thus it is necessary to discuss and agree the applications for which Eurocode should be used and hence the requirements that it must fulfil.

1.8. Overview of terminology

The foregoing review of Eurocode and other methods for describing and classifying foods uses various terms in referring to the systems, often as used by those systems. It may be helpful to formulate some definitions to bring some consistency to the terminology used to describe the systems. These definitions supplement those given in Chapter 2 of Møller & Ireland (2000) and in the following, terms defined in that work are shown underlined. Further definitions relating to food classification are included in Ireland (2005).

Descriptor: In the context of a thesaurus, a **descriptor** is a preferred term in the thesaurus, but in general it can also refer to a free-language term selected as an indexing term. In the context of Eurocode it has a special meaning as a term in an indexing language (the Eurocode Descriptor System) used to supplement the main classification by recording extra description of the food such as its preservation and cooking method.

Category: The difference between descriptors and categories is difficult to define or agree, but might be considered as the property of being adjectival or a noun phrase, or alternatively partial or complete. Thus "whole", as in "whole apple", is an acceptable descriptor but does not encapsulate a complete concept. A **category** is a fully defined, independent concept and thus provides a consistent entity that can be placed in one or more hierarchical classification systems. For example, the category "Meat dishes" might be placed in one hierarchical system under the parent "Meat" and in another under the parent "Prepared dishes".

Classification: **Classification** is the organisation of subjects or objects (in the present context, foods) into a system of defined classes of similar items that can be considered or processed together, for example in information retrieval or data aggregation. Two types are prevalent, hierarchical classification and faceted classification.

Hierarchical classification: The definition of a classification as a set of broad categories, each of which may be further subdivided into more specific categories, repeatedly as a hierarchy of multiple levels, gives a **hierarchical classification**. Each branching position in the hierarchy may be referred to as a **node**. Nodes may be identified by preferred terms (descriptors), by codes that are separately organised into the hierarchy through links to their next-higher (i.e. parent) node (or broader term), or by codes that incorporate the hierarchy, as is the case for Eurocode. A specific node can have only one parent node (broader term), but may have multiple child nodes (narrower terms) and multiple sibling nodes (i.e. nodes that share the same parent node). However, a node with a specific definition can appear in several places in the hierarchy, the so-called poly-hierarchical relationship. For example, "field corn" can be a node under "sugar-producing plant", "starch-producing plant", "oil-producing plant", "grain", etc., according to its use. In a hierarchical classification, the categories used are often defined to be mutually exclusive and unique in that classification, i.e. the same category cannot appear at different nodes

in the hierarchy. Preferably the category name should be complete in itself, not depending on parent categories to complete its definition. For example, the category should be called "Meat products", not just "Products" under the parent "Meat" category.

Faceted classification: In **faceted classification** (Vickery, 1960), the descriptors are organised into facets, these being independent aspects of the subject being classified. Thus the LanguaL food description language, described earlier, includes facets for the independent aspects of "Food source" and "Cooking method". The descriptors defined for each facet can either be a flat list of terms or organised hierarchically. For some LanguaL facets, such as "Cooking method", more than one term may be assigned to a food. The action of assigning multiple terms, both from different facets and the same one, is referred to as indexing and permits retrieval on any of these terms. However, the intended result is, taking LanguaL as the example, to build a composite category that uniquely describes a set of similar foods and thus this remains a classification system.

Ontology: The term *ontology* originated in philosophy, but in computer science **an ontology** is a specification of concepts and the relationships between them applicable to a specified subject area. The concepts consist of individual objects (each sometimes referred as an **instance** of the object and within a database possibly corresponding to a single record) that are usually handled as members of a **class** (within a database possibly corresponding to a database table). Objects have **attributes** (i.e. properties) that may vary between instances or have values shared by all members of the class. Objects also have **relations** that indicate how they relate to each other. The simplest is the IsA relation of membership, i.e. *instance IsA class* as in *carrot IsA vegetable*. Another common relation is IsPartOf, which corresponds to the relation between the LanguaL facets for "Part of Plant or Animal" and "Food Source". More subject specific relations may be implemented, for example HasAFunctionalParent as in *sucrose δ^F -(dihydrogen phosphate) HasAFunctionalParent sucrose*. The construction of an ontology assists in defining terms and relations (as used in a particular scientific community) to facilitate information exchange, as well providing enhanced options for the processing of the information. Upgrading a thesaurus into an ontology is a development that can provide enhanced retrieval possibilities, as exemplified by work on the AGROVOC thesaurus (Soergel *et al.*, 2004).

2. Food categorisation in data management

The objective of the Eurocode 2 Review Project has been to assess the existing system, as documented in 1993, and to propose revisions and other improvements that would enable it to be implemented as a practical system. This Position Paper was prepared at the conclusion of the COST Action 99 – Eurofoods project as a record of the progress and an assessment of the options for further development and implementation of the system. The Eurocode review project has also provided the opportunity to review other existing systems for food classification and food description. It is impractical for any of these, including Eurocode, to provide a completely independent solution other than for specific and limited requirements. Wider harmonisation of information reporting foods would more practically be based on a framework that can use the most appropriate elements from complementary systems for describing, classifying and grouping foods. Thus a key aspect of this review is consideration of the contribution Eurocode 2 can make to such a framework by maximising its ability to fulfil one or more specific roles.

The revision of Eurocode 2 has concentrated on improving the categorisation of foods within a general classification in the expectation that more application-specific classifications would be superimposed on the underlying categorisation as and when required. The EFG proposal can be considered to endorse this approach. The present section considers in more detail the role of food categorisation in food data management and suggests some mechanisms that might form a basis for developing this role.

2.1. Categorisation, classification and grouping

Systems may be called food categorisation, food classification or food grouping systems and it would be useful to define distinct meanings for these terms (see also section 1.8). Categorisation is based on a set of categories defined either through short descriptors or longer, more descriptive texts. In classification, each category is assigned a unique position in a hierarchy⁴ and is defined to be mutually exclusive except inasmuch as one category wholly contains another in a broad term – narrow term relationship. A grouping is an aggregation of food categories on the basis of some criterion intended to collect the categories together. A grouping system may be used to bring together similar items in a food table or to give meaningful results when data associated with the grouped foods are themselves aggregated. Until now, it has usually been the case that a single grouping system has been applied to a set of food data, but equally a single set of food categories might be organised in several differing ways through the use of different grouping systems.

Eurocode 2 has usually been termed a food coding system which can be considered to mean that it associates defined food categories with codes that may be assigned to food records to indicate that they belong to that category. In that sense it is a food categorisation system. However the hierarchical nature of the codes organised the categories into a classification, and this classification was used for grouping. Much of the difficulty in developing and agreeing a working version of Eurocode 2 arose not from defining categories, but from the requirements of placing categories into an appropriate position in the code hierarchy and of providing a single, useful grouping of foods.

Categorisation and hierarchical organisation cannot be entirely divorced from one another. A hierarchy helps in the management of categories by allowing related ones to be selected and listed. A broad term may be required for coding when there is uncertainty about the specific nature of a food item. Sometimes such a broad term might not form part of the intrinsic hierarchy used to manage the categories, but should be available if required. A commonly quoted example is a category for 'Fatty fish'. A generalised categorisation system that is to be used for various applications involving different grouping systems for aggregation may need to include 'grouping categories' that are appropriate to individual grouping systems, as discussed in section 4.3.2. The categorisation system should have the capability of managing grouping categories that do not fit into its intrinsic hierarchy.

⁴ Unless the classification is designed to include poly-hierarchical relationships, for example as described in the introduction to the LanguaL thesaurus (Møller & Ireland, 2000). However, within the present discussion, a poly-hierarchical relationship implies categorisation rather than classification.

It must be possible to map categories to the required grouping systems. The EFG system was drafted from work on the categories used in the Eurocode and other international coding systems. However the project to test it involves mapping each national food consumption survey to EFG. Development and implementation of Eurocode 2 as an underlying food categorisation system with effective mechanisms for mapping to different grouping systems would provide a harmonised method of recording detailed food consumption, composition and related information. This would enable standard mappings to be defined to provide data aggregations using various international and national grouping systems, not only those designed for nutritional studies but also for the assessment of other intakes such as of additives or contaminants. The usefulness of the approach and the viability of mapping to different grouping systems should be studied on a test file of consumption data coded by the Eurocode 2 system, initially through the implementation of a system mapping Eurocode to EFG and then to further selected grouping systems.

The discussions that led to the proposal of the EFG considered that an optimal grouping system might consist of 30-40 top-level groups, together with approximately 250 subgroups. A revision of Eurocode itself to meet this specification would involve a complete reorganisation, including loss of the current convenient number of main groups. On the other hand, using the model of an underlying categorisation system for coding mapped to one or more grouping systems will allow independent development of a radically different main group plus subgroup system.

2.2. Categorisation, identification and description

The latest revision of the Core Classification defines three levels of the hierarchy, the main groups, their subgroups and the sets of sub-subgroups. Below these levels, either immediately or with further intervening subgroup levels, there is a level of category that equates to food items. The Rationale Document (Poortvliet *et al.*, 1992) states that as an advantage in a simple system "new food items can be assigned to existing categories without requiring revision of those categories". However, it also defines one purpose of Eurocode 2 to be to provide "a code in which no detail in the information identifying [a] food is disregarded or lost in aggregation". This raises various questions about the relationship between the Eurocode food categories, the identification of specific foods, and food description that provides additional information relating to the food item.

As the depth of the hierarchy increases, the categories become more specific until the category name is a food name (which may be further defined through a scope note or a longer documentary note). Depending on the content of the name, this may represent the food in its state as produced, as purchased or as consumed. Additional information about the food may be recorded as food description. When recording foods as consumed, this description should include any relevant modification to the food applied after the state in which the food is named⁵. It should be noted that this is using food description in a distinctly different role from the indexing of description. For example, the food category 'Raisins' would be indexed with the descriptor 'dried' but the descriptor would not be used when recording the consumption of raisins.

⁵ The information may be implicit or explicit in the name. If the basic recording of vegetables is in the raw state, the food 'Broccoli' is implicitly raw, but it may be made explicit as 'Broccoli, raw'.

Alternatively, however, the food categorisation might only define the category 'Grapes', in which case the descriptor would be required to fully specify the food. Thus, although an independently-defined food description system such as LanguaL might be used with the Eurocode categorisation, the rules for its use in this context would be determined by the content of the categorisation system.

In this way, food categorisation and food description should support two separate and well-defined roles in food identification. Do they in combination completely identify a food? The answer depends on the definition of 'a food'. Possibly it is necessary to consider this to be the most precise category that it is possible to specify within the categorisation system⁶. This leaves two further levels of uncertainty, whether the definition of a precise category remains constant⁷ and whether there is a spread of meaning even within the most precisely defined category. This spread may be subdivided into differences that are (or may be) known but which are difficult to incorporate into the categorisation (i.e. differences which are not usefully included in the categorisation such as transient or minor variations, for example in the production process) and, secondly, differences that are unknown, such as biological variation.

2.3. Advanced category management

The above sections have discussed the relationship of food categories to other elements of information on a food such as its classification, grouping and description. A key role in food identification is envisaged for food categorisation, as exemplified by the Eurocode 2 categories, in combination with food description to record further known information. So far it is assumed that categories will be managed using codes of the Eurocode form discussed in section 4.2. This provides a useful basic classification and is suitable for the management of independent categories whose definitions do not change. Changes in codes or category definitions would normally be implemented as a new version of the system, requiring that coded data are accompanied by information on the version used. This simple approach restricts the options for developing the more sophisticated techniques that may be required for the effective recording of information on foods in consumption, composition and other studies. This section introduces a more flexible approach for category identifiers and some of the management requirements that these might address.

2.3.1. Category identification

Although the Eurocode-style codes provide helpful information to the user in terms of recognisability and basic classification, they are not suitable for managing a range of relationships between categories, for example variants of categories that arise over time or apply to use in different contexts. The alternative of a simple numeric identifier without any meaning allows the separate management of an unlimited range of relationships between the categories.

⁶ Similar considerations are likely to apply to a food description system as this is also dependent on its constituent categories.

⁷ Not just over time, but with respect to other variables. For example, local or cultural factors may influence which precise category of a widely used system a food record is assigned to.

A preferred approach to the implementation of this is that the definition associated with a category identifier does not change. Any change will involve the creation of a new category and identifier. Such changes will occur over time as the definition of a food category changes, but they could also occur in other situations. For example, it might be decided that categories used in food composition databases and food consumption studies should never coincide because the difference in context may imply subtle differences in definition. A category defined in legislation will frequently include implications that are not present in common usage of the term and thus the two should be treated as separate categories. Such separately identified categories may be interrelated on any number of linking criteria, including a link to the category representing the current definition of a Eurocode category. The interrelationships might be managed as an ontology, see section 1.8.

Advantages of the approach are that it is unnecessary to record version information when assigning codes or to change the internal identifier of a category that is unchanged in a new version of the categorisation. These considerations make it much more convenient to maintain and update the system through frequent minor amendments, in keeping with the new possibilities for distributing these via the Internet. Identifiers for the obsolete and modified forms of a category differing between versions of Eurocode would be linked to the current Eurocode category. For example, the identifier for the version 99/1 category **6.19** *Basic products of other cereals* would be linked to the version 99/2 categories **6.16** *Barley basic products* and **6.25** *Basic products of other cereals*. The latter category differs in that it excludes barley products. If necessary, all foods assigned the identifier associated with the old **6.19** code could be retrieved and the link list used to provide the user with a choice of the revised codes. If it were required to change a Eurocode code but not the underlying category, only the identifier – code link would need to be changed.

2.3.2. Category overlap and multiple categories

The use of internal numeric identifiers allows a large number of categories to be managed by interrelating two or more categories on the basis of various types of relationship. The commonest of these would probably be a link from a specific category to a grouping category to which it belonged. A further possibility is a relationship indicating that the linked categories overlap, i.e. parts of their scopes overlap as with 'Leaf vegetables' and 'Brassicas' (although these are treated as mutually exclusive in the present version of Eurocode).

The concept of overlapping categories is closely related to the possible assignment of multiple categories to a food record. If these overlap, the food will belong to the more specific category represented by the overlap. For example, if the categories 'Meat pie' and 'Wholemeal pastry pie' are assigned, the food will be a wholemeal meat pie. This is related to faceted classification but instead of defining independent factors overall, these apply to particular broad categories, in this case 'Pies', and are based on more specific categories that may already have been defined. The categories could be completely dissembled, in this case to 'meat(-containing)', 'wholemeal' and 'pie', which would amount to so-called post-coordinate indexing. However post-coordinate indexing discards the overlap relationship between categories which is seen as being a major feature in managing complex category assignment. The overlap relationship is in effect a flexible implementation of pre-coordination applied directly to separate categories rather than by combining them into a composite category.

Multiple categories assigned to a food may relate to the whole food or to separate parts of it as is the case in the wholemeal meat pie example above. It is possible to view the ingredient list of a recipe as an extension of the principle in which a more precise set of multiple categories, together with their quantitative contribution, is used to define a derived category representing the final dish. The full identity of the ingredients would be the combination of the category and description discussed in section 2.2.

2.3.3. Multilingual categories

Multilingual support for a categorisation thesaurus usually consists of direct translation of the category names, and the associated documentation, corresponding to each individual category code. If the multilingual thesaurus is stored in a relational DBMS, the category name in each language would be held as a separate record in a database table of category name translations. However the framework outlined in the preceding sections provides the more flexible option of handling categories in alternative languages as distinct categories associated through appropriate links. This would allow mechanisms to be implemented to handle cases where one-to-one translation of categories cannot be considered exact. Some constraints on the linking of categories are likely to be necessary, for example that any category must at least have a link to another in the same language or to one in a base language of the thesaurus, possibly English.

3. Potential applications for the Eurocode system

Foods are assigned to food groups, food subgroups and more specific categories in order to fulfil a range of requirements for recording, managing and using food-related information. The design of the Eurocode 2 food coding system needs to ensure that it can meet the requirements of the applications for which it is intended to be used. The rationale document (Poortvliet *et al.*, 1992) warns that "General food coding systems, designed to suit 'all purposes' tend to serve their users poorly."

It continues by outlining the purposes of Eurocode 2 as "The general aim of Eurocode 2 is to serve as a practical code for intake assessment and table cross references. Specifically Eurocode 2 is: ..." and then by listing seven features. Some of these are quite demanding, for example "A code in which no detail in the information identifying food is disregarded or lost in aggregation, even if nutrient composition data are not yet available." This implies that Eurocode 2 should support certain functions which are of general utility, for example the recording of detailed consumption data for foods independently of their composition, and the ability to match consumption and composition data⁸. There are various advantages to the wider use of Eurocode, for example economies of scale in support and user experience, and the availability of coded consumption data for use in additional contexts. Thus a range of potential applications, both function-oriented and mission-oriented, for the Eurocode 2 food coding system should be reviewed and the requirements of these should be accommodated within the design where possible.

⁸ As was the practice in Denmark for the earlier version of Eurocode, foods in the composition database can have Eurocode codes added as an aid to matching with consumption records.

3.1. Recording of food consumption data

The implication of the example design criterion cited in the previous paragraph is that a clear distinction should be made between the recording of food consumption data and the aggregation of that data. It requires that the coding system should be able to capture all information defining the food reported as consumed, as well as any information about which details were unknown or not reported. It also requires that foods be recorded to the level of specificity to which they are reported, leaving any grouping or aggregation of the consumption as a manipulation to be performed on the stored records. This has the major advantage that all raw data are retained and thus alternative manipulations can be performed on the same set of data. It may not be possible or desirable to eliminate all elements of grouping from the basic coding since these help identify foods. However it would be an important contribution to the design of the system if the grouping requirements for food identification and for data aggregation were clearly distinguished.

The detailed information available on consumed foods will be highly dependent on the dietary survey or other method that has been used to collect it. This should be documented in association with the recorded consumption, being so-called metadata equivalent to, say, the analytical method used to obtain composition data. Any conversion of the raw survey data to the content specified in the coded file also should be documented. However the preferred scenario is that the coding system should be fully integrated with the survey method used and directly store the reported consumption, although this may have to be a longer term objective.

Thus there are two rather contrasting sets of requirements involving the storage of near-raw data as accurately as possible and the storage of data emanating from current survey methods. At a time of very rapid expansion of information sources and development of information management techniques, accommodation must be made both for current situations and probable future scenarios. The purpose of data coding at input has hitherto been two-fold, to record information in a controlled language suitable for further analysis and to record it in a 'flat file' in which all the stored information for an entry is stored in a single record with a structure common to all the information entries. The coding of a dish as a single mixed-food code illustrates both these features. Increasingly the capture and storage of consumption data in more complex data structures will become practical.

3.2. International comparison of consumption data

One of the main purposes envisaged for the Eurocode 2 food coding system has been to provide a common language for coding food consumption data so that data from different countries can be easily compared. In some cases it would represent the initial storage format for the data but in other cases (and perhaps more often) the coded data would be produced by converting survey data collected or coded using a different local system.

The major task in converting existing data into a format for international comparison will be the mapping between the two codification systems, which hopefully would be implemented as an automatic interconversion rather than through manual recoding. Since it is the international system that fulfils the requirements for comparison, part of its function should be to provide tools and documentation that help implementers to make that interconversion as accurate as possible. These may need to include language translation aids as well as a level of documentation that ensures the mapping of comparable categories defined in differing terminology. In the longer term, the international system may need to support information converted from external

categories that do not correspond one-to-one with its own categorisation, including cases where terms in different languages do not exactly correspond.

The comparison of data from more than one country is likely to involve files that have been subject to different interconversions from their original codings. The international system may need to provide facilities to manage information on the source format and content since factors such as the extent and purpose of the original data collection may have influenced the data translation process. Comparability of data is likely to be better where the data have been collected in the context of a single project or collaboration than where files from different countries were prepared independently.

3.3. Aggregation and disaggregation of consumption data

The first two applications above relate to the storage of consumption data in a form that represents as nearly as possible the data as collected. Applications which use a food coding system for the aggregation of data require it to support one or more hierarchies which allow the data at any level below a high-level category of interest to be retrieved and summated. Except in cases where it is certain that only one basis for aggregation will ever be required, the aggregating classification should be an adjunct to the coding system, not its inherent classification.

One approach to the comparison of international data would be the classification within each country of consumption data to an agreed set of aggregated food groups. The working group meeting held in December 1999 agreed to test this approach on the basis of a set of categories proposed as the EFG system and described in more detail in section 6.1.

Another application is the disaggregation of consumption data for mixed foods and dishes into the intakes of basic foods at the individual or population level.

3.4. Harmonisation of household budget survey data

Some of the most detailed work to date on the Eurocode 2 food coding system and its assignment to data generated in different countries has been undertaken in the context of Household Budget Surveys, HBS, (Lagiou, *et al*, 1992). Instead of using Eurocode, the later DAFNE I project (Trichopoulou & Lagiou, 1997) and DAFNE II project (Trichopoulou & Lagiou, 1998) mapped each national HBS food code directly to one of the 45 DAFNE aggregated food groups. At the time of the DAFNE I project, EUROSTAT was revising the PROCOME food codification system as part of the work to harmonise HBS data. There might be interest in the use of Eurocode as a more detailed food coding system within the broader PROCOME codification but it would be necessary to convince EUROSTAT that Eurocode is an effective tool for recording the information.

3.5. Matching of consumption and composition data

The matching of food consumption data with food composition records as a precursor to the calculation of nutrient (or other food component) intakes from the amounts of foods consumed is commonly seen as a key application of Eurocode. However it would also be one of the most demanding. It would be necessary to match each food consumption record with one, and only

one, food record in the composition database. A common technique to achieve this is the use of the same code to represent a specific food in both the composition and consumption data. This may be the food code of a national food table. However, more foods (that significantly differ from one another) are usually recorded in a consumption survey than in the published table. Thus a survey database is often supplemented with component values obtained by recipe calculation or imputation for extra food items that are reported. However this approach to coding has some significant disadvantages, for example new codes become available during coding. Thus coding is not performed within a consistent environment of an unchanging set of available codes.

A food classification system cannot be used to match a food consumption record with an individual food composition record, i.e. a specific row in a food table. The relationship is at a higher level, whereby a group of food composition items and a set of food consumption records will be classified into the same classification group. Each type of food record must be associated with further information in order to complete the selection of the appropriate composition record corresponding to a consumption record.

The further information needs to be held using food description systems and related facilities such as recipe information management. These must be coordinated to provide an overall system supporting the capability of reporting any aspect of a food that may correlate with its composition, even if the information is not normally available directly in consumption data. The use of categories and descriptors in Eurocode is discussed in section 4.3.4 and the general relationships between information categorising, classifying, describing and identifying foods have been considered further in section 2. An overall Eurocode system is discussed in section 5.

The Eurocode system or other approaches to the systematic identification and description of foods provide the possibility of removing the need for a coding which directly links composition and consumption records. However, the advantage, gained because a systematic approach allows the two types of information to be coded independently but consistently, is counterbalanced by the need to develop algorithms which support the near-matching of records so that the most appropriate composition data can be applied to each individual consumption record. At one level, this near-matching must be able to retrieve on a one-to-one basis the single best composition record to use. Doubtless more sophisticated data imputation procedures will be developed to provide a best estimate of data, for example for a fried vegetable based on known results for related vegetables and cooking methods, to improve on the best near-match data available.

3.6. Managing food group related food property data

For many properties of foods, detailed data may not be available for each individual food, but the reported data may be sufficient to deduce likely property values on the basis of similar foods as defined through some form of food grouping. The main groups of Eurocode were based on a review of the food groups used in various European national food composition tables and thus provide an overall categorisation of foods that may be useful in relating particular types of food to certain properties.

One area where it is generally accepted that property values might be associated with a form of food grouping is in the management of nutrient loss and gain (NLG) data. The general term NLG data includes yield information associated with weight changes in the preparation and cooking of foods and with weight changes ascribed to the loss or gain of specific materials such as water or fat. It also includes nutrient retention factors which give the proportion of specific components

not lost or destroyed in preparation or (more usually) cooking. Many NLG data for specific foods exist (Bergström, 1994; Bognár, 2002) but they require to be organised so that factors may be imputed for related foods.

NLG factors will largely be determined by physical (and also chemical) activity during preparation of the food and its ingredients. The effect of these activities will depend on a range of influences including the type of food, its physical state and the cooking or other process(es) involved. The equipment used will relate to the amount of evaporation and this will be a significant determinant of yield factors for some dishes. However, in other cases the amount of reduction will be determined to a greater or lesser degree by the type of dish, for example for certain sauces. A food grouping system, such as the basic Eurocode classification or a separate grouping of underlying Eurocode categories, is probably required to handle the type of the main starting food or individual ingredients and possibly the type of the resultant dish. Separately the data must be organised in respect of the other aspects, in particular the cooking method, and this requires a suitable descriptor system.

Eurocode may or may not provide a suitable food grouping system for use with NLG data, and this will need to be assessed. It may not be useful to group at the same level for every type of food, but it may be appropriate to use either Eurocode main groups or more probably subgroups as the base grouping level. However, if it is considered that Eurocode does not provide a suitable grouping system that raises the question of how a suitable one should be found or developed. A better option may be to design Eurocode to provide the flexibility so that its basic elements can be adapted for particular applications and implementations. A range of potential applications has been identified and although each may have special requirements there will be much advantage in building solutions on a common foundation.

4. Revision of the Eurocode 2 classification

The Eurocode 2 Core Classification version 98/1 was presented in the Review report and at the Wageningen workshop as described in the Introduction section. It was the basis for the further work that produced version 99/1 and the further revision, version 99/2. This section summarises these revisions, the details of which are available on the Eurocode Website, as described later in section 5.3.1., 'Website overview'. It also presents the revised system of category codes used and reviews some policy considerations relating to the changes made. The Eurocode 2 version 99/2 classification is listed in Annex A.

4.1. Summary of Main Group revisions

This section summarises the major changes made to the classification for each Main Group and records some points which have not been finally resolved. The changes have been made in two phases. The first produced version 99/1 which was discussed at the workshop of the COST Action 99 – Eurofoods Working Group on Food Description held in December 1999. Version 99/2 is the result of the further changes suggested at that workshop.

The Eurocode 2 classification consists of categories defined and organised into a hierarchy on the basis of food source, organism part and product type criteria. Thus there are two aspects to be taken into consideration when reviewing the system, firstly whether the categories are appropriate and well-defined, and secondly whether their hierarchical organisation provides an effective

classification. This section also notes remaining critical comments on version 99/2, for example adverse affects of restricting the number of levels in the hierarchy, resulting in the omission of some accepted categories from the revised classification. In some cases, the missing categories may be considered to be *Grouping categories*, as discussed in section 4.3.2.

Other general comments (Porubska, 2005) have noted that some categories are very general (e.g. those covering 'products'), whereas others are very detailed (e.g. those covering mayonnaise and flours). It has been suggested that categories for *Other* would be useful for cases where the required category is not included, although the option is available to classify such items at the next higher, more general level. The documentation should include photographs, and more names and synonyms of typical national dishes of each country would be useful. This would aid the translation of category names. It is also important to have included the difference between food products, ready meals sold in shops and home prepared meals.

4.1.1. Main Group 1: Milk and milk products

The version 98/1 Core Classification retained the original version 93/1 Main Group 1 classification. For version 99/1 the old category of **Milk** was split into **Cow milks** and **Other milks**, with the latter subdivided by species. Feedback at the workshop indicated that this classified milk products by species at too high a level and that fat content should be given precedence. Further disadvantages were that the main group was then categorised on both source (species) and product type criteria, that source information could not be recorded for products such as yogurt and cheese, and that any products such as condensed milk could only be recorded if they are derived from cow milk.

The workshop noted that further work was required on the categories for cheeses. As a result of this and the aforesaid difficulties with the classification of milk products, a discussion note on the **Classification of milk and cheese** was prepared. This is reproduced as Annex C. As well as describing further work needed to assign specific cheeses to categories, it develops the use of structured codes for milk products and of standard subcodes to be used at a lower level for the source species. The result is an alternative proposal for version 99/2 in which the species would be recorded using standard subcodes at the fourth level, below two levels divided on the basis of product type and a third divided by fat content. The structuring would allow the species to be recorded even where the fat content is unknown.

Additions to the group include a category for higher fat creams and subgroups for **Imitation milk and cream** and **Milk beverage powders**. The *Chocolate-flavoured milk*, *Fruit-flavoured milk*, *Dried milk* and *Acidophilus milk* categories have also been added.

Comments: The question has been raised of whether the categories *Water ice*, *Granita* and *Sorbet* should be moved out of the *Milk and milk products* group as they are more like sugar-based sweets or desserts.

4.1.2. Main Group 2: Egg and egg products

In version 99/1, the subgroup level was treated similarly to Group 1, with a subgroup **Chicken eggs** for the most important species and others assigned to general subgroups. A distinction was made between domestic and wild birds but discussion showed that this was not considered helpful.

For version 99/2, the group was revised so that the species is specified at a consistent level, removing the separate treatment of **Chicken eggs** and the separate subgroup for **Eggs, wild birds** that had been introduced in version 99/1. The primary categorisation is by the egg-producing species, with the next level available for more detailed subdivision, for example by duck species, should this ever be required. At the next lower level, eggs can be coded as whole, white or yolk, using standard codes analogous to the milk source species codes, for example 2.10.00.30 for *Chicken egg white*. Subgroups were added for *Egg products* and *Egg dishes*.

4.1.3. Main Group 3: Meat and meat products

A major change in version 99/1 was the complete separation of sets of subgroups for carcass meats and for offal, with offal divided according to the anatomical organ and then source organism. It should be noted that an **Other offal** subgroup results in the next level of subdivision appearing at a lower level in the hierarchy, for example *Beef liver* is at the sub-subgroup level whereas *Beef heart* is at the food item level. In version 99/2, an additional **Other offal** category of *Intestines* was included and the category *Blood and offal products* was added to the **Meat products** subgroup. Also, a category for *Minced meat products* was also added to this subgroup. The category includes *Sausage meat*, thus limiting the use of the Sausage categories to encased or shaped sausages.

Further work is necessary on several aspects of the Meat group. Carcass meats need to be subdivided on some basis relating to the meat cut but this provides a general problem for classification as every country has its own definitions. Cuts might be classified according to fat content. Meat cuts have not been modified from version 93/1 although categories for the cuts of veal and lamb/mutton are needed to bring these meats into line with beef and pork. The version 99/2 category listing includes these as the undefined codes 3.15.x and 3.25.x.

A suitable subdivision of the **Restructured meat and meat analogues** subgroup is required. The extra subgroup **Meat dishes** has been added as a basis for future discussion. Aspects for consideration include whether there is a useful distinction between **Meat products** and **Meat dishes**, and also the basis for deciding whether a cereal/meat mixed dish such as *Meat pie* should be classified as a meat dish or a cereal dish. An alternative approach would be to create one or more extra main groups for mixed dishes.

Comments: First level subdivisions between *Mammalian meat* and *Poultry*, and also for overall *Offal*, are missing and their insertion into the hierarchy should be considered. The present species-based categorisation excludes the separate classification of game. *Cured meats* is not a separate category, but included in *Preserved meats*.

It has been commented that *Salami* is missing, there being an important difference between salami and sausage. Also, the terms Rohwurst and Bratwurst provided problems and it was suggested that a clear simple division into salami and sausages with simple characterisation such as smoked, hard, soft, etc. would be better. In fact, the definition of *Sausage* used in version 99/2 does include *Salami*, this being a brine-pickled and/or smoked sausage made of raw ingredients. It has been included in the category **Dry, smoked sausages (Rohwurst)**. Three categories are included for sausages and these are given a descriptive name plus the German term in parentheses. This was felt to give the best categorisation, although this may need further review as the classification and definitions for sausages varies considerably between countries.

4.1.4. Main Group 4: Fish and fish products

Eurocode 2 version 93/1 used a taxonomic basis for classifying fish, but used category names such as *Mackerel-type fish* and a large *Fish, miscellaneous* category. The Core Classification version 98/1 changed the category names to the scientific names of individual fish families. In version 99/1 subdivisions for fish species were added to an extended set of family subgroups and to a small subgroup for *Other fish*. No further changes were made for version 99/2. Subgroups for marine mammals and non-aquatic organisms have been brought together at the end of the listing in case it is decided to move them to other appropriate main groups.

There remain strong arguments for and against defining the fish subgroups on taxonomy rather than on more subjective grouping categories such as *Fatty fish*. Possibly the solution lies in an implementation which allows the species categories to be associated with alternative grouping categories. In any reorganisation of the Main Groups, consideration should be given to renaming this group **Fish, crustaceans, molluscs and amphibians** (with molluscs including land snails) and deciding the appropriate main group for marine mammals, reptiles, insects, etc.

Comments: A query on the category for **Frog legs** confirms that some clarification such as the suggested renaming of the main group is necessary, together with clearly stated policies for other types of fauna.

4.1.5. Main Group 5: Fats and oils

Version 99/1 for the **Vegetable fats and oils** subgroup used the version 98/1 list of subgroups, adding the individual oils. The only change made during the version 99/2 revisions was the removal of *Cod liver oil* to the **Vitamin / mineral products / tonics** subgroup in the **Products for special nutritional use** main group.

It might be better to subdivide the overall **Vegetable fats and oils** subgroup to give more detail of the source, for example replacing it with subgroups such as **Seed oil**. An alternative classification doing this was proposed and posted on the Website⁹.

4.1.6. Main Group 6: Grains and grain products

A clear separation has been made between basic products produced by the processing of grains, including their milling to flour, and cooked products including breads. In version 99/2 the Barley categories were removed from **Basic products of other cereals** into a separate subgroup. The new subgroup **Substitute flours and starches** includes a range of products previously located in other main groups.

In version 99/1 bakery goods were divided into subgroups for **Soft bread, Unleaven bread and crispbread, Savoury cereal products** and **Sweet cereal products**. Various further renamings and reorganisation were made for version 99/2. The Soft bread subgroup was renamed **Leavened bread** and within it the category *Potato bread* has been added and the word 'flour' substituted for 'grain' in *Bread, other flour* and *Bread, mixed flour* to include the possible use of *Substitute flours and starches*. A subgroup for **Bread products** has been added. The subgroups for 'Savoury cereal products' and 'Sweet cereal products' have been replaced by **Fine bakery wares, Savoury**

⁹ Currently available at <http://www.ianunwin.demon.co.uk/eurocode/docmn/ec99/vegoil02.htm>

cereal dishes and **Sweet puddings**. Some categories such as *Sponge cake* have been generalised and given sub-categories for more specific food items.

Comments: Cakes and pastries should be identified better or using more generally accepted terminology, since the terms used for categories such as *Danish pastry* may not be familiar. Many such problems could probably be resolved by further work on definitions and documentation.

4.1.7. Main Group 7: Pulses, seeds, kernels, nuts and products

For version 99/1, a separate subgroup, **Underground pulses**, was introduced for *Peanut*, partly to emphasise that its products should be placed under **Pulse products**. In version 99/2, categories for unripe pulses such as *Pea* and *Broad bean* which are consumed as vegetables have been moved to the **Pod and seed vegetables** subgroup of the **Vegetables and vegetable products** main group. The corresponding categories remaining in the Pulses group have had 'Dried' added to their names. The separation of kernels and nuts is somewhat unsatisfactory. These might form a joint subgroup separate from seeds, but ideally further subgroups should be defined to prevent the lists of categories in each becoming too long.

Comments: Soya products should include salted and roasted soya grains, extruded soya products, soya mayonnaise and soya desserts (imitating yogurt). Soya milk, cheese and yogurt are included in the **Imitation milk products** subgroup of the **Milk and milk products** main group, but the comment does indicate that there may be difficulties with implementing a policy on substitute foods. The topic of substitute foods was covered in Eurocode 2 Discussion Note N004 on **Substitute and analogue food products**, which is reproduced as Annex B.

4.1.8. Main Group 8: Vegetables and vegetable products

Some changes were made to the vegetable subgroups to accommodate some extra categories. 'Bulb vegetables' was renamed **Onion-family vegetables** to indicate clearly the inclusion of *Chives*, *Leek*, etc. Categories for immature (unripe) fruits consumed as vegetables were added to the subgroup **Fruit vegetables**, with *Avocado* and *Olive* moved in version 99/2 from the **Fruits and fruit products** main group. The subgroup 'Pod and seed-head vegetables' was introduced in version 99/1 but renamed **Pods and seed vegetables** in version 99/2 when categories such as *Peas* and *Broad beans* were moved to this subgroup from the Pulses group. The subgroup **Seaweeds** was added, removing these from the Fish main group. The subgroup **Vegetable products** was added for puréed, pickled and fermented products.

4.1.9. Main Group 9: Fruits and fruit products

Some changes were made in version 99/1 to reduce the number of categories within each subgroup. These included the separation of 'Stone fruit' into **Prunus species fruit** and **Other stone fruit**, and the use of the extra hierarchical level where possible, for example for *Custard apple*. In version 99/2, *Avocado* and *Olive* were moved to the **Vegetables and vegetable products** main group. The subgroup **Fruit products** was added and includes fruit sauces, previously categorised as 'Sauces, fruit base' in the Miscellaneous foods main group. However fruit jellies such as Apple jelly and Redcurrant jelly remain in the Sugars main group on the basis of their main ingredients. There remains some doubt whether this separation between groups is

the best option. It has been suggested that **Malaceous fruit** is a rather obscure name for the first subgroup, but no better alternative has yet been identified.

Comments: In the **Fruit products** subgroup, the inclusion of categories for specific fruit sauces such as *Apple sauce* excludes the many mixed sauces, such as apple and peach or apple and carrot, where apple is not major constituent. The latter example would remain a problem of fruit and vegetable mixtures even if the specific categories were replaced by a general **Fruit sauces** category.

4.1.10. Main Group 10: Sugar, sugar products, chocolate products and confectionery

In version 99/1, sugars were separated into the subgroups **Sugar (sucrose)** and **Other sugars**. The 98/1 subgroup for 'Confectionery bars' was made more specific as **Chocolate-coated confectionery bars**, with other confectionery bars including cereal bars being under **Non-chocolate confectionery**. No further changes were made for version 99/2.

4.1.11. Main Group 11: Beverages (non-milk)

For version 99/1, the single subgroup **Alcoholic long drinks** was created for beer, cider, etc. However, the December 1999 workshop considered this unclear and for version 99/2 the single subgroup has been divided into two, **Beers and malt beverages** and **Ciders, perries and similar drinks**. Within the latter, the existing category 'Ginger beer, alcoholic' was renamed *Ginger beer*, and the separate version 99/1 category 11.40.36 *Ginger beer* was removed. This brings the classification into line with the policy of categorising alcoholic and non-alcoholic forms together. **Liqueurs** have been subdivided into types to take advantage of the extra hierarchical level and a subgroup has been added for **Alcoholic mixed drinks**. A separate subgroup for **Carbonated soft drinks** was created. Pending further consideration, fruit and vegetable juices and nectars (with nectars merged to a single subgroup) have been kept in the Beverages main group rather than being classified as fruit or vegetable products.

Comments: The group may divide into separate products too quickly, with an initial division into **Alcoholic beverages** and **Non-alcoholic beverages** possibly a better alternative. The **Infusion drinks** category should possibly include *Cocoa*. This is not at present specifically included in Eurocode as a separate category, but the product would probably be included under *Drinking chocolate powder* in the **Milk and milk products** main group. A new category of *Sports and energy drinks* is needed.

The use of separate subgroups for **Fruit juices** and **Vegetable juices** misses the possibility of classifying mixed multivitamin juices made containing both vegetable and fruit juices.

4.1.12. Main Group 12: Miscellaneous, soups, sauces, snacks and products

For version 99/1, the subgroups were reorganised with the aim of providing more logical and comprehensive listings of ingredients, accompaniments, and products. Substantial listings of herbs and spices were added, together with a separate subgroup for **Herb and spice mixtures**. The 'Sauces' subgroup was divided into **Savoury sauces** and **Dessert sauces**, with fruit sauces moved to the Fruit main group. An additional type of product was introduced through the **Prepared salads** subgroup.

In version 99/2, categories in the **Prepared salads** subgroup were renamed to incorporate the word 'Prepared' to distinguish them from homemade salad dishes, and the category *Prepared mayonnaise salad* was added for products containing more than 50% mayonnaise. Further subgroups for **Dips** and **Sandwich fillings** were added, with the version 99/1 'Dressings and dips' category being renamed simply **Dressings**. Standard subcodes have been proposed for the main ingredient in products in the **Dips**, **Savoury sauces**, **Dessert sauces**, **Soups**, **Prepared salads**, **Sandwich fillings** and **Savoury snacks** subgroups. However existing categories in these subgroups have been retained until the proposal has been considered further.

4.1.13. Main Group 13: Products for special nutritional use

No changes were made to this main group in version 99/1 but for version 99/2 *Cod liver oil* was placed in the **Vitamin / mineral products / tonics** subgroup, moved from the **Fats and oils** main group. The coding of baby foods needs to be considered further, possibly with some foods being removed from other groups such as *Rusks* from **Grains and grain products**.

4.2. New Eurocode category codes

The existing Eurocode 2 codes were designed for manual coding of foods through the assigned classification category and to retain some additional information, for example to code mixed foods. A primary design consideration for the codes was their usability by human coders and an important secondary concern was that their computer manipulation should be straightforward. The current assumption is now for computer support of coding, allowing users to select categories rather than directly selecting and assigning codes. Also the concepts and availability of programming have evolved so that the information held in a well-structured code becomes available for a wide range of manipulations, and this becomes the prime consideration in the specification of codes.

As a result of these changes in emphasis, the main task of the codes should be seen as the recording of the information systematically using a controlled language and structure. Earlier versions of Eurocode assumed that in cases of difficulty with recording data for a particular consumed food, the coder would include free-language textual notes. However, as noted in the evaluation report (van Kappel, 1993A; p 12) in the discussion of information loss, "Apparently the candidates [undertaking the evaluation coding] preferred to code information directly instead of making notes." It is impractical to expect the routine appendage of notes during coding (except, perhaps, where the coding support system identifies this as necessary)¹⁰. Therefore the overall code should be capable of recording any information that may need to be systematised for manipulation of the consumption data. As well as the Eurocode 2 classification code, this total code for a food record should include any information that is recorded using the descriptor system, links to recipes, etc. In many cases, the full record will also include the original data as collected for the consumption report.

¹⁰ Also, such notes would mostly be used for later review of records rather than their retrieval or aggregation since they will be in free-language that is unsuitable for automatic processing.

4.2.1. Eurocode version 99/1 category codes

A new specification of code was introduced in Eurocode 2 version 99/1 to increase flexibility while retaining a recognisable similarity to previous codes. In the new codes:

- The Main Group is coded as in the version 93/1 codes. This has the disadvantage that codes will not sort into the correct order of Main Groups (e.g. Group 10 codes will sort between Group 1 and Group 2 codes). This could be corrected by using a leading zero, e.g. Groups 01, 02, etc., but it may be considered better to continue using the familiar Main Group numbering¹¹.
- All lower level classification categories are coded using 2-digit numbers, giving codes such as "3.25.16". Leading zeroes are not used (e.g. the first Meat subgroup is 3.10, not 3.01), leaving these available for special categories such as additional grouping categories or other special purposes.
- Gaps are left between codes, for example by initially assigning codes such as 10, 15, 20, etc.
- Some informal intermediate hierarchy may sometimes be implicit in the numbering. For example, all bird meat are coded 3.3*n*. Following from this, all codes with a trailing '9' might be used for 'other' categories, e.g. 3.39 for *Birds, other*. In practice it was found that it was not always possible to apply this while maintaining a suitable spacing between codes assigned to fully defined categories.
- Codes in the range 10 to 79 are used for specific categories. Numbers in the series 0*n*, 8*n* and 9*n* are reserved for use as special codes (at any level in the hierarchy) for general situations such as 'unknown', 'not specified', etc., or as an escape to alternative or specialist extensions, for example as defined nationally.
- The version 99/2 classification uses up to four hierarchical levels, an increase of one over Eurocode version 93/1, with the Core Classification being defined as the top three levels. Any specific identifiers, for example descriptor or recipe codes, will be handled as identifiably distinct components of the code.

4.2.2. Code structuring and special codes

During the version 99/2 revisions, it was necessary to revise the categorisation of milks so that the source species was classified at a lower level in the hierarchy. The option should be available to record species whenever needed, requiring the ability to append a species code to any milk products category at a lower level than the subdivisions for product type and fat content. Also, there might be cases where information at an intermediate level was unknown, for example it might be necessary to record consumption of *Sheep milk* without knowing its fat content. Therefore a series of species codes has been defined to be used at the fourth level in milk product codes, for example 1.*nn.nn*.30 represents any milk product from sheep and 1.30.10.20 codes *Goat's milk yoghurt, >3% fat*.

In version 99/2, code structuring of this type has been introduced for three areas of information:

- source species in the milk products group
- part of egg in the eggs group

¹¹ However, if there is a major revision of the Main Groups, for example to add one or more for mixed dishes, the Main Group numbering could be revised, perhaps to start at '21'.

- food group of main ingredient of products in the miscellaneous foods group.

The implementation of structured codes using standard sets of subcodes is supplemented by the introduction of special codes based on subcodes from the reserved series *0n*, *8n* and *9n*. These may be used to fill intermediate positions in a code that do not hold information or to flag information about the following subcode(s). The subcode '00' has been defined to indicate 'no information' so that, for example, *Sheep milk* without details of fat content would be coded as 1.10.00.30.

Products such as sauces and soups in the Miscellaneous foods group have always been subdivided on the basis of the food group of their predominant ingredient, for example *Sauces, dairy base* or *Soup, meat / poultry base*. The subcodes defined to record information on the main ingredient could usefully be more detailed. Therefore standard subcodes have been defined for main groups and many subgroups, wherever possible with a similar code to the corresponding group or subgroup, for example a general meat-based soup (12.50.30) and a poultry-based soup (12.50.33). This has necessitated the use of subcodes such as '.83' for *Root vegetable-based* which would normally be a reserved code. Also, conflict during retrieval might arise with codes from subgroups not conforming to the structure, e.g. 12.10.30 *Baking powder* is not a meat product. These two points could be resolved by using a reserved code as an escape to a standard code with a specific meaning. Thus .81 could flag a (milk) species code and .82 a product ingredient code. Then the subcode .81.50 would always indicate a product from *Reindeer* and .82.85 a *Fruit vegetable-based* product.

4.2.3. Future enhancement of codes

The codes specified in version 99/2 serve to record information for the classification categories defined. However they have also been designed to accommodate future requirements for recording additional information and adding new categories.

Codes at each level (except for the Main Groups) use a 2-digit number preceded by a fullstop (.). Thus the string ".nn" identifies a partial classification code (a subcode) in version 99/2. This allows additional codes, for example representing descriptor terms, to be identified, for example through their initial alphabetic character. Thus 3.19.20T4 might be the code for *Boiled hare*. Such a descriptor code would even be recognisable incorporated above the lowest level in a code although this may not be generally useful. If the same approach were used for a recipe reference, however, this might be used to indicate the relationship between an ingredient as reported and as incorporated in a recipe. For example, the reported use of *Chicken leg* and *Chicken wing* within a general recipe, r0248, for *Chicken curry*, might be coded as 3.30r0248.20 plus 3.30r0248.30. This might provide an approach to handling one aspect of recipe variance.

The Eurocode 2 Review report suggested that the logical ordering of categories in listings of the classification should be maintained by a separate mechanism from the codes used to record them. However, the Wageningen workshop concluded that codes could at this stage be changed and that it would be helpful if the codes provided a logical order for the categories. This requires a code specification that allows insertions. This has in the primary instance been met by the leaving of gaps in the code sequence, but at some point the available gaps in particular places will be used up. However the hierarchy categories are enumerated with two-digit numbers (i.e. 00 to 99), without limit on the number of two-digit pairs that make up a subcode. Thus at any given level extended category subcodes such as '1705' are valid, being terminated either by a fullstop and an

alphabetic character signifying a supplementary descriptor. Potentially this makes the system infinitely expandable.

At any level within a code, subcodes from the series $0n$, $8n$, or $9n$ remain available for special purposes, for example to act as escape codes. One use for this might be to allow national (or other) extensions to the classification, for example where local requirements demanded extra or more specific categories. If the code 91, for example, were used for this, a list of various cooked sausages required in Germany could receive codes of the form 3.60.30.91GE.*nn*. In an international context, these would be truncated to the higher level category for *Cooked sausages*, 3.60.30.

Special codes should also be defined to allow details such as 'information not known' or 'information not recorded' to be included in codes, and additional cases such as 'other' and 'mixture of subcategories' should be considered. Thus 8.10.95 might be used for *Leaf vegetable, type unknown* and 8.38.97 for a *Root vegetable mixture* such as carrots and swedes.

The examples in these sections have been suggested by the work on the current version of the Eurocode 2 codes and should be the subject of future discussions. All the proposed special subcodes, including those for 'grouping categories' discussed below, should be reviewed together to ensure that overall they will result in an effective and consistent system for coding food categories.

4.3. Categorisation policies

The modified Core Classification (Eurocode 2 version 98/1) proposed a revised set of subgroup categories for each Main Group within the constraint of only one lower level for more specific terms, including individual food items. The major task of the 99/1 and 99/2 revisions has been to extend the enumerated categories down to the next (sub-subgroup) level with a further level available below it. The primary objective in identifying suitable categories to be included in the classification was not intended to produce an exhaustive enumeration. Rather the intention was to determine the policies for defining (and thus of using) categories in each Main Group and how these should be best organised in the classification hierarchy for the group. In this section, some of the general issues that arose during this review are discussed.

4.3.1. Categorisation and multiple classification

The 99/1 and 99/2 revisions have concentrated on improving the categorisation in Eurocode 2 rather than on developing a classification hierarchy that meets any particular requirements. It is envisaged that differing requirements for grouping and aggregating consumption and other food data will require that the Eurocode categories are organised into various alternative hierarchies designed for different applications. The main function of the hierarchy intrinsic in the Eurocode codes is to provide an effective means to identify, organise, list and help assign categories. This limitation of the purpose of the intrinsic classification should simplify discussion of the categorisation since it removes many of the contentious issues and allows it to concentrate on providing a consistent means of identifying and recording food items. The classification aspect is discussed further in the last of these sub-sections.

4.3.2. Grouping categories

However it is not possible or desirable to separate completely the issues of categorisation and classification. Some alternative classifications will require specialist grouping categories. For example, although version 99/2 categorises fish taxonomically, alternative classifications may require grouping categories such as *Fatty fish*. Even so, the approach should also simplify the definition of such grouping categories, or allow the use of alternative definitions, without adverse implications for the underlying recording system.

Categories of this type can be conveniently referred to as grouping categories and can perform several functions. They can be used for coding when they represent the most specific information available, for example when consumption of *White fish* is reported. They can form the basis of alternative classifications for aggregation and also to provide an alternative route for locating and documenting specific categories.

Grouping categories may be positioned in the Eurocode 2 hierarchy at a level at or above that of the highest level of category linked to it. For example, a *White fish* grouping category might be placed at the subgroup level but include the subgroup *Gadiformes*. Grouping categories that are not within the intrinsic Eurocode classification should have distinctive codes, using one of the available special series *0n*, *8n*, or *9n* to flag that it is a grouping code. For example, *White fish* might be 4.09.10 and *Fatty fish* 4.09.20. The full specification of the new system of codes should include the ability to distinguish between standard assignment of categories to a grouping code and user-specified assignments.

4.3.3. Food source versus product type

As noted in the earlier introduction to LanguaL, the first three aspects it covers are Product Type (facet A), Food Source (facet B) and Part of Plant or Animal (facet C). LanguaL considers the first two to be distinct whereas the Eurocode 2 classification has always aimed to incorporate the two aspects of product type and food source into a single hierarchical breakdown. Rules based on common practice determine which aspect is given precedence in any particular part of the classification. In some main groups such as Vegetables, categories have been principally defined on the basis of the source organism¹² whereas in others such as Milk products they have been defined by product type. The Eurocode 2 Review (Unwin & Møller, 1998; pp 38-39) discussed some difficulties arising from this. It also noted aspects of the specification of Part of Plant or Animal in the Eurocode classification for Meats (*Ibid.*, pp 17-18) and Vegetables (*Ibid.*, p 27).

Some anomalies in the information that can be recorded may be created if similar products are classified according to different policies in respect of product type and food source, particularly if the number of hierarchical levels is restricted as has been the case for earlier versions of Eurocode 2. A major example in version 93/1 is the classification of soya substitutes for mammal milk products. For mammal milks the categories recorded the type of milk by fat content but no information on the species. On the other hand, soya substitutes were recorded as Soya products but without information on the type, for example milk, yogurt or flour. Related

¹² The rationale document (Poortvliet *et al.*, 1992; p 16) states that the classification of fruit and vegetables was to some extent botanical, but was generally based on distinctions made for trade and commercial purposes, this approach being more practical and commonly known to consumers.

difficulties were reviewed in a discussion document on **Substitute and analogue food products**, which is reproduced as Annex B.

Revisions have been made to improve the coding of soya and other substitute products by classifying them according to product type rather than their source. However, more general inconsistencies remain, for example between the coding of dishes (e.g. fish pudding as Fish) and soups (e.g. fish-based soups as Miscellaneous foods). It may be noted that 'Fish-based soups' forms a well-defined category but that it must have been considered more useful to classify these by product type, i.e. as soup, rather than the main food source, i.e. fish. In this case classification by product type may be preferred because of the mixed-food nature of soups and of similar Miscellaneous foods subgroups such as Sauces and Snacks. The version 99/2 proposal for standard codes representing group information on the basic ingredient provides a capability for recording extra detail. The classification might be clearer if such consumed food items were assigned to a separate main group, perhaps called 'Prepared mixed foods', rather than being included in Miscellaneous foods with ingredients, condiments, etc. However such a step should be considered at the same time as the policies for handling mixed foods and recipes are further discussed rather than as a separate classification issue.

In general, main groups initially based on source organism include further categories (for example for mixtures or products) defined by product type whereas main groups divided by product type may have categories further subdivided by source (for example 'Soups, fish base'). It is important for the usability of Eurocode that categories are defined and organised according to clear policies, both generally and for particular main groups. The present proposals, including the introduction of code structuring, move in this direction but further discussion is required before such policies can be finalised.

A consequence which arises from considering both the categorisation-classification and source-type aspects is that where appropriate a category should be defined in terms of both source and type so that they can be placed in both source-based and product type-based classification hierarchies if either alternative is used in data aggregation.

4.3.4. Relationship between categories and descriptors

The Eurocode 2 classification consists of categories defined and organised into a hierarchy on the basis of food source, organism part and product type criteria as discussed above. It has also been noted that it is often required to store additional description for reported food items and that Eurocode version 93/1 included a simple Descriptor System for this. These requirements for food description will be discussed further in section 4.1.1 but first it is useful to consider the relationship between categories that might be defined for the classification and information that should be stored through descriptors.

Comments (Løken & Borgejordet, 1998) on the initial Core Classification proposals suggested a wider use of descriptors in identifying details of food consumption records, including the extensive use of the food Component Added and food Component Removed facets. To take *Whey* as an example, this is included in the present classification as a distinct category, but alternatively might be coded as *Cow milk, < 1% fat* with a descriptor for *Casein removed* and possibly for *Fat removed*. The former approach uses a single category term to encapsulate a package of meaning whereas the second approach processes that meaning into separate elements for classifying the information. By removing some of the information to the Descriptor System, this reduces the number of categories required in the classification but there is also information

loss. For example, it is reasonable to assume that matching consumption coded as *Whey* with composition data will give better nutrient intake results than a match on *Cow milk, < 1% fat, Casein removed, Fat removed*. The effectiveness of this use of descriptors will be highly dependent on their specification and some relevant aspects of the Descriptor System are discussed in section 4.1 covering that component of an overall Eurocode system. The relationship between categorisation and description was further discussed in section 2, *Food categorisation in data management*.

Whether this approach of processing the information before coding it (or recoding it, manually or automatically, from raw consumption data) is preferred depends on the applications for which Eurocode 2 is to be refined. If its sole purpose is to create a single classification that can be applied to all situations for which it will be used, then this will be appropriate. However, if different classifications are required for different surveys or for different analyses in the same study, then the requirement for the underlying application is the effective recording of data, leaving scope for different analyses of that data, for example by applying alternative aggregation hierarchies.

For this reason the version 99/2 categories generally continue to use terms which encapsulate description with the identity of the basic food where these are generally accepted and well-defined terms for a food. However the category definitions should include a list of descriptors implicit in the category, both to clarify the scope of the category and to assist in automatic aggregations of data based on descriptors. This requirement should be taken into account in designing the Descriptor System.

4.3.5. Classification hierarchies

The above points have discussed issues which mainly relate to the categorisation of foods although (or, in fact, because) the more difficult problem is the organisation of these categories into a generally agreed classification hierarchy. This is difficult for several reasons including:

- differing perceptions (or regulatory specifications) of the relationships between foods
- differing requirements between applications and between studies for grouping foods
- difficulty in defining consistent subdivision policies for categories both between Main Groups and between subgroups within a Main Group
- difficulties with classification-led categorisation, for example cross-classification combinations such as *Fat spread, 45-65% fat, >25% saturates*, and classification-breaking categorisation, for example in food mixtures such as *Beef and pork products*
- selecting, defining and naming grouping categories where there are no clear-cut logical groupings within a large number of lower-level categories.

As discussed in earlier sub-sections, many of the requirements for classification are related to the later manipulation of data rather than the initial recording (coding) of the data. If data are recorded using categories (and other components of the system such as descriptor terms organised into facets) independently of the later grouping classification requirements, this will provide much greater flexibility in analysing the data. The basic Eurocode categories can be organised into alternative classification hierarchies according to the specification requirements of one particular analysis of the data. Although appropriate hierarchies would often be defined during the process of data analysis, some standard classifications could be developed and made

available as resources within the Eurocode system. These could be implemented as a list of parent – child links.

These proposals that application-specific hierarchies should be constructed independently of the assignment of pre-defined coding categories do not necessarily completely remove the need for an integral classification of the Eurocode categories. A simple classification is required to group the categories together in a meaningful way. There are also some applications which might benefit from a generally accepted hierarchy, for example the organisation of foods in a national food tables and databases. Finally, the proposals on structured codes suggest that the use of a more formalised category structure may result in more flexible options for coding and retrieving food records.

5. The overall Eurocode system

The introduction to the 'Components of the Eurocode system', section 1.4, emphasised that although the current revision proposals have concentrated on the Eurocode 2 classification categories, it is necessary to consider that an overall system should include further components such as descriptor and recipe support. Also, the system must be supported by detailed documentation which itself should be considered as one of the components. The present section discusses these components in further detail.

5.1. Descriptor System

Eurocode 2 version 93/1 includes a simple Descriptor System as overviewed in the Eurocode 2 Review report (Unwin & Møller, 1998; pp 11-12). Before deciding how the Descriptor System part of the overall Eurocode system should be developed, it is necessary to specify the requirements that it must meet and consider whether these might be met wholly or in part through the use of the LanguaL and/or INFOODS food description systems.

5.1.1. Requirements

The simple version 93/1 Descriptor System was stated (Poortvliet & Kohlmeier, 1993; p 19) as being designed specifically for more detailed coding of foods as consumed and reported in dietary intake studies. This serves as a general statement of the objectives, although some revision and enhancement is probably required to achieve this. During the Eurocode 2 Review project, more specific requirements have been identified or proposed where particular aspects of reported information should be recorded through the use of a descriptor system. These areas are discussed in the following subsections.

5.1.1.1. Food modification

The existing five descriptor facets include Component Added and Component Removed facets for recording information on food modification. Their intended use is indicated by the documentation of individual terms, and examples in other parts of the Eurocode documentation.

The Component Added facet was not intended to record the constituents of mixed dishes as this information should be held in an ingredients list, as noted in the documentation for the *Alcohol added* term. It may be specific to the reason for addition, for example *Salt added* applies to

flavouring, not to preservation (although whether it is the producer, the consumer or the coder who should decide the purpose for a food as consumed is unclear). This is a difference from LanguaL where facets are coded independently and salt added for preservation would be indexed in both the Treatment Applied and Preservation Method facets. Which is the 'correct' policy depends on the requirements for storing information specified for the Descriptor System.

Descriptors in themselves record qualitative rather than quantitative data. Thus descriptors must be used as indicators, rather than measures, for data aggregation and for data retrieval. The facets available, the thesaurus of descriptor terms, and the definitions of the terms must be designed to allow the necessary qualitative information to be stored and used. The same thesaurus of descriptor terms may be available for use in more than one facet, for example the LanguaL Geographic Places and Regions thesaurus can be used for Place of Origin, Place of Consumption, etc. Similarly, food constituent terms from a common list could be used in separate facets for Component Added and Component Removed. If further detail needs to be recorded, three linked facets could be used for Component, Modification Type and Modification Purpose. Modification Type might distinguish between addition of a new constituent and supplementation of a constituent. Such a system would require a coding support system to control the combinations of terms permitted. However the approach taken should be determined by the requirements rather than any perceived complexity of the system.

If there is a requirement to record quantitative information, this could be stored in association with the descriptor in a quantitative facet (Unwin, 1992). For Component Added this becomes a recipe which records a list of the added constituents (ingredients), their amounts and any additional descriptors, as discussed further in section 5.2, below.

5.1.1.2. Food processing and preparation

A primary requirement of the food descriptor system is to record information about the processes that were applied to a food item before it was consumed and which may influence its composition or other properties that might impact on the health or well-being of its consumer. On this basis, it may be necessary to record, for example, information on the amount of heat treatment applied in terms of the time and temperature as an indicator of appropriate yield and nutrient retention factors to use for calculated compositions. In determining the requirements to be used as design criteria, the most stringent demands should be taken as the basis for the full system with adequate provision for the use of simpler procedures.

A food descriptor system may include suitable terms for a process itself without being able to code significant related details. For example, it may be relevant to record whether a cooking procedure was part of the prepurchase industrial food processing or applied in the home immediately prior to consumption¹³. Similarly home freezing or other preservation might need to be distinguished from the industrial equivalent. Even within one of these scenarios, the order in which procedures are performed might be significant. Some of the extra information might be handled using a data structure for associating descriptors with a food analogous to those in the COST Action 99 – Eurofoods recommendations for relating 'contributing foods' to a food (Schlotke *et al.*, 2000). On this model, the processes applied to a food would be recorded as a sequence of acquired descriptors, allowing that some descriptors could be acquired simultan-

¹³ The Eurocode evaluation report (van Kappel, 1993A; paragraph 5.8) noted a problem with lack of clarity as to whether descriptors referred to food production or preparation.

ously. Data, for example NLG factors, associated with processing and preparation procedures needs to be related to both a starting and final material, each of which must be described. This reinforces the suggestion that the Eurocode classification should record foods as commodities, with further processing recorded through the descriptor system. A starting material is then recorded as classification plus descriptors, with further descriptors added which represent the set of processes for which data are being reported.

Information associated with a single recorded process may need to be organised into several subfacets reflecting different aspects of the process. As well as the main identification of the process, information on the context of processing (industry, catering, home), the conditions of processing (time, temperature, etc.), and similar considerations may need to be recorded. Additional aspects that may influence, for example, NLG yield factors include the type of cooking container used and the scale of the food preparation, which may determine aspects such as water loss by evaporation.

5.1.1.3. Food composition

There remains a continuing debate on if and how the Eurocode 2 classification should take account of the composition of foods. Reasons for not doing so, and some exceptions in Eurocode up to version 93/1, have been documented (Poortvliet *et al.*, 1992; p S11). In some cases subdivision by content, for example of fat, might add useful structure to the classification but two potential problems should be considered. Firstly some food items, as is the case with cheeses, may fall on the borderline and a consistent policy for handling this would need to be developed so that the coding does not become arbitrary. Secondly, the approach leads to a proliferation of categories if more than one component is taken into consideration as in the 'Fats and oils' group categories such as *Fat spread, 45-65% fat, >25% saturates*.

An alternative would be to handle the amount of a component of interest within a descriptor system. This would provide much greater flexibility in the data that could be recorded for consumed foods, including the ability to adapt the nature and level of data recorded to the needs of particular studies, for example ones interested in fat or sugar intake. In the above example of fat spread, separate sub-facets for fat and saturates would allow both aspects to be recorded without the need to predefine a large number of combined categories¹⁴. It would also allow information on the fat or saturates levels in other foods to be recorded without impacting on their basic classification.

5.1.1.4. Food use

Each Main Group in Eurocode 2 version 93/1 concluded with a subgroup for coding foods of the group which are '*for dietetic use*'. This discards any other information about the food below the Main Group level. These subgroups have been discontinued in the Core Classification revisions since it is considered that the information would be better handled in a 'Use' descriptor facet, in many cases with further information included in the 'Component added' or 'Component removed'

¹⁴ The proposal on code structuring (see section 3.2.2) introduces a limited implementation of this approach into the main Eurocode 2 categorisation but its implementation as independent descriptor facets would provide more flexibility and wider applicability. However to take full advantage, a system for mapping coded data to an aggregating grouping must be able to include the descriptors in its mapping.

facets. Other 'Use' subfacets covering the circumstances of food consumption might record aspects such as categorising the place of consumption.

5.1.1.5. Food maturity and storage

On occasions there may be a requirement to record information on the state of maturity of a food at consumption. Differences may arise from its state when harvested or continuing development during distribution or storage. In some cases, these differences may relate to the intended use of the food. Even, in particular for *Fruit vegetables*, they may influence the Main Group to which they are assigned as some items (or at least very closely related items) may be eaten as vegetables when unripe and as fruit when ripe.

5.1.2. Relationship to LanguaL

LanguaL is structured into 14 facets. Each facet consists of a hierarchy of descriptor terms, some such as facet A, Product Type, extensively elaborated and others such as facet F, Extent of Heat Treatment, less so. Some facets allow only one term to be assigned to a reported food, classifying this, while others allow multiple assignments of terms and thus index the food. Some further loose structuring exists through the grouping of these into characteristics such as Processing and the implicit organisation of the Adjunct Characteristics of Food as further facets. Some hierarchies include *de facto* subfacets, a relevant example being the 'Ingredient Added' group of categories within facet H, Treatment Applied.

Generally, it has been assumed that the Eurocode descriptor system needs to be a simpler form of LanguaL, perhaps a subset of it. Whether this is true depends on an assessment of the requirements for the Eurocode descriptor system. However this can be seen as part of a more general question: is there a need for a food description system that records information not currently handled by LanguaL? It is generally agreed that the Eurocode descriptor system should be based on a system with a wider development and application base than would apply to a system designed specifically for use with Eurocode. Thus it is perhaps more useful to identify maximal rather than minimal requirements for the Eurocode descriptor facilities as a contribution to specifying the requirements for a general food description system. In any case, one of the stated purposes of Eurocode is to provide "a code in which no detail in the information identifying food is disregarded or lost in aggregation, even if nutrient composition data are not yet available." (Poortvliet *et al.*, 1992) and this implies the need to record any descriptive detail for a food that may influence the later assessment of its composition¹⁵.

Also, it may be helpful to discuss some areas where the current LanguaL system might not completely fulfil requirements for food description. Possible requirements for handling information on the processing and preparation of food have been outlined above and it is convenient to use some features of the management of such information associated with NLG factors to illustrate how a food description language such as LanguaL may need to be enhanced.

¹⁵ This includes the requirements to associate the amount of consumed food reported with the state it was in when the amount was assessed so that correct composition figures can be applied. This is made difficult by the assumption that foods are reported 'as consumed' and argues that a better approach is to classify the basic food item and supplement this with description to give the state when the amount was assessed. This is one area that will specifically define the Eurocode requirements for detailed food description information.

Although the NLG factors used in recipe calculation may be handled at a general level, for example to be applied for a given cooking method within a given food group, food description is also required for recording observed NLG values when these are reported. Here it is important to retain as precise information as possible and thus, for example, the food description language should be able to record quantitative data for the time and temperature of a cooking process. Food description conventionally applies to a single food item, whereas NLG factors need to be associated with a starting material and a final product. One approach to handling this aspect would be to enumerate separate processing events within the food description. An NLG factor reported for a final product could then be associated with some starting point within the chain of processing events that created the food product. This would also permit intermediate yield or nutrient retention factors to be associated with the final product by associating them with both a start and a finish point. The approach allows the storage of information on cumulative factors, for example vitamin loss measured for a second cooking process in a prepared food that has already been subject to a loss.

5.1.3. Relationship to INFOODS system

The INFOODS system of food description includes more facets than the LanguaL thesaurus. The range of aspects covered overlaps those of LanguaL but each vocabulary extends to areas not covered by the other. For example, LanguaL contains a facet for food contact surface whereas the INFOODS system includes aspects of agricultural production¹⁶ and food sample detail¹⁷. Also, except in a few instances, the information for a facet is stored as free-text rather than as controlled descriptors. This has the advantage that the coder can record any information that is felt appropriate, including quantitative data, but has the major disadvantage that this information is not structured in a way that can be consistently interpreted by computer software.

As noted earlier, coders in the Eurocode evaluation were reluctant to add textual notes in the test environment. Although this finding might not be so clear in a practical situation, it is clear that the free-text aspect of the INFOODS system is unlikely to provide an environment for the recording of systematic description of foods reported in consumption studies. However the ability to record precise information, whether in controlled or free language, may be considered a requirement for a future food description system and thus the current INFOODS procedures should contribute to any such design.

5.2. Recipe System and recipe coding

It was noted in the introductory section that the Recipe System component of the overall Eurocode food coding system would play a very important part in recording information about mixed foods and dishes. Work in this area was undertaken by the COST Action 99 – Eurofoods Working Group on Recipe Calculation and Information, and also by the LanguaL Technical

¹⁶ Some of these aspects are collected together in LanguaL facet Z, Adjunct Characteristics of Food. Facet Z is a set of extra facets which in effect gives LanguaL more than the fourteen enumerated at the top level.

¹⁷ Information on details pertaining to individual food samples is considered outside the scope of LanguaL.

Committee which has incorporated Full Ingredient Indexing procedures in the imminent revision of the system. The group working on recommendations for data interchange and management planned to await the outcome of these efforts before incorporating detailed support for recipe information in a future version of the recommendations¹⁸. It is proposed that similarly the results should be reviewed before decisions are taken on the future of the Eurocode Recipe System are taken. Therefore this section is confined to noting a few key issues that will need to be considered.

The draft manual (Poortvliet & Kohlmeier, 1993; pp 9-10, 12-13) describes proposed procedures for handling recipe information based on recipe codes appended as the final field of the food code. The recipe code would link to a file containing the recipe details. Difficulties with the instructions for coding the recipe are noted in the Eurocode evaluation report (van Kappel, 1993A; paragraph 5.7). However a change to using the Eurocode classification for identifying basic food items, supplemented by a sequential acquisition of processing descriptors, should allow the ingredient information to be based on more logical and easily understood principles. Processing descriptors may need to be associated with recipe intermediates as well as the basic ingredients and the final dish. There may also be a requirement for the recipe information to include additional aspects such as NLG factors and household measures.

The procedures proposed in 1993 envisaged a central coding centre that would register ingredient information and would coordinate national centres assigning national recipe codes. It is unclear whether this was ever a practical framework and now would need to be considered in conjunction with identifying a central coordination for maintaining the Eurocode food coding system overall. On the other hand, the emergence of the Internet as a medium for making documentation and other information quickly and readily accessible to users opens up new possibilities for the coordination, maintenance and dissemination of recipe files.

A fundamental, but perhaps somewhat neglected, aspect of recipe information is the identification and naming of recipes. Many recipe dishes are internationalised and the names of these, for example Spaghetti bolognese, are widely recognised. Although the recipes for such dishes may vary nationally as implied by the version 93/1 recipe procedures, they are just as likely to vary according to other factors such as whether the dish (or part of it, as with Bolognese sauce) is prepared industrially or in the home. Other dishes are local in nature but belong to widely recognised categories of dish, such as a pie or stew. And finally there are local dishes that are distinct from any generally used type of dish. All three categories of recipe need to be associated with ingredient list information but there may be other differences required in their information management, for example in the requirement for handling variations and for registering their details.

The development and use of an effective recipe system is required for various purposes such as the disaggregation of dishes to their basic foods and the accurate conversion of food intake to nutrient intake (or the intake of other food components). The question then remains of the requirements for assigning a classification code to the overall dish. The version 93/1 draft documentation (Poortvliet. & Kohlmeier, 1993; pp 16-17) provides rules for the classification of mixed foods and recipes in which the first consideration for a dish (other than a generic mixed

¹⁸ A similar situation appertains in the EuroFIR project, with one Work Package working on recipe information and calculation and inputting its results into the Work Package compiling updated recommendations.

dish which has a specific code defined) is "its specific cultural identity"¹⁹. This criterion is perhaps not ideal for consistent coding (or for the international comparison of data). Where a recipe is a way of preparing a specific basic food that is the predominant ingredient, this food should be coded, probably with a general 'Component added' descriptor. For dishes that incorporate comparable amounts of ingredients from different food groups, for example pies, stews and ravioli, it is questionable whether there is any advantage in trying to establish a 'main' ingredient so that the dish can be assigned to a regular Main Group. Any back-calculation to food group intakes using a code assigned on this basis is likely to be inaccurate and misleading, not least because it will be biased by the rules defined for selecting the main ingredient.

The current revision has concentrated on the categorisation of basic food products and indeed does not specify procedures for coding mixed foods and dishes. These procedures should be based on an overall framework that has defined facilities for supporting descriptor and recipe information, and thus should be developed at a later stage. Since the details of prepared dishes and recipes will be handled by these components of Eurocode, the main purpose of assigning a code to a recipe dish will be to identify its general type. The best approach to this might be the creation of a separate Main Group for recipes with no predominant ingredient. In version 93/1, soups and sauces were assigned to the Miscellaneous Foods group, together with ingredients and unassignable food items. It would be an option to extend this to all recipes. A suitable categorisation within this Recipe Dish group could also help with the coding of unfamiliar dishes if similar dishes are recognised in the category listing.

5.3. Documentation

Effective documentation is seen as an integral part of the Eurocode 2 system. A major objective for the system is the coding of data so that these are comparable internationally and this requires that coding is performed consistently by different groups of coders. This can only be achieved if the documentation is accurate, clear and readily available. The draft manual (Poortvliet & Kohlmeier, 1993) was found by both the evaluation exercise in 1993 and the Eurocode 2 Review report to require considerable improvement. However up until the present work, no changes had been made and indeed it was difficult for potential users of Eurocode to obtain a copy.

Rather than working on a revision of the draft manual, it was decided to build a new set of documentation as an integral part of the revision procedures. Indeed the new twin possibilities of preparing the documentation in a format with hypertext linking and of making it available on the Internet have transformed the options both for designing it and for presenting and discussing proposals for revising the Eurocode 2 system. It was therefore decided that the revision work would be based on the preparation and documentation of proposals made available on the World Wide Web (WWW).

5.3.1. Website overview

The address of the Eurocode 2 Food Coding System website is <http://www.vfd2.dk/eurocode/> although to simplify maintenance during the revision work, the underlying pages are currently located at <http://www.ianunwin.demon.co.uk/eurocode/>. Through the directory structure, the site

¹⁹ Vegetarian Bolognese sauce might be considered to be culturally a meat dish.

is separated into parts for the documentation, for the discussion of proposals, and for reports and similar materials.

The documentation is currently divided into separate sections for versions 93/1, 98/1 (the original 'modified Core Classification'), 99/1 and 99/2 (the current proposals), together with a section that at present reproduces the 93/1 documentation of the Descriptor System. The earlier versions of the classification are currently available in parallel with the latest proposals so they can be conveniently reviewed and compared, but later the obsolete versions will be separated into an archive section.

The 99/2 revision is its documentation. The policies and listings available on the Website are the primary source documents and directly record the current drafts and proposals, as well as providing a forum for discussion through the inclusion of discussion points and feedback received. When the resultant version 99/2 is agreed and finalised, the discussion may be moved from the main documentation but will remain accessible through links from the documentation pages. It is intended that any further comments submitted as a result of using Eurocode 2 will be added to the relevant discussion pages as the basis for formulating future revisions of policy and content.

5.3.2. Main Group policies and listings

The documentation of the Eurocode 2 version 99/2 Main Groups consists of an introductory text plus a list of the Main Groups with links to access their documentation. For each group, there are links to both its 'Classification' and its 'Categories', the latter referring to the category documentation discussed below.

The first section of the classification documentation for a group specifies the policies on which its classification is based and which thus provides guidance on where items should, or should not, be coded. The policies are given in sections covering Inclusions, Exclusions, statements of Policies, notes of Revisions, and a list of points for Discussion. Cross-reference links are provided where appropriate, for example to the place in the classification where Exclusions should be coded. Above the policy documentation is a link to the classification listing for the group which occupies the lower section of the page.

The classification listing section begins with a list of the subgroups which also provides a convenient link to their detailed listing. There is also a link back to the beginning of the policy section. The full listing of the group classification is laid out in three columns, in the same way as it is formatted in Annex A. The first column gives the code and name of the category. The second lists the scientific name or similar information (e.g. the source carcass for types of offal) where this is appropriate. The third column gives synonyms, similar products, or examples of more specific items that are included within the category. Some 'see also' cross-references link to related categories. Some scientific names currently included are taken from the same source as their category and thus are unchecked. It was noted that alternative spellings and errors are common in source material on foods and in some cases have propagated from one source to another. Also confusion between subspecies and variety names is not uncommon.

In Eurocode 2 version 99/2, the Core Classification has been defined at the subgroup level and the level below this. Thus the listing shows the subgroup followed by the next level down of categories indented. In two cases (for the common bean and sucrose) it was felt to be useful to define categories to a further level, potentially the food item level. This was done by providing a linked page that lists the main group, subgroup and sub-subgroup categories aligned with the left

margin, indenting the next, food item level beneath the sub-subgroup. This approach of listing one level and its immediate sublevel on separate pages may limit the view of all levels in the hierarchy but perhaps gives the clearest listings at a given level of specificity. Coding software should allow the user to view all category levels in a selected section of the classification.

5.3.3. Category documentation

Categories listed in the classification are further documented through links to a separate page, one for each Main Group, that provides definitions, explanations and other details about the category. Discussion points are also included for categories where appropriate, and it may be useful to keep these permanently to indicate any doubtful points of information about a category with a view to continually improving its documentation. It is intended that the documentation of a category should provide as much useful information as possible, in the future including images, rather than being confined to a conventional scope note. At present the category documentation consists largely of definitions located while developing the classification and it could usefully be extended to be more comprehensive. As such it could become a useful reference resource beyond its role as documentation of Eurocode. The category documentation of the appropriate group also lists the standard category codes for milk product source species and for the main ingredient group of products listed together as miscellaneous foods.

6. Recent developments

Since the Eurocode Review Project was completed, work in several areas of food classification has progressed. Towards the end of the COST Action 99 – Eurofoods work, a proposal for a simple grouping system for food consumption data was developed, called the Euro Food Groups (EFG). The COST Action 99 work on LanguaL (Møller & Ireland, 2000) proposed the incorporation of existing food classifications such as Eurocode into LanguaL Facet A. More recently within the current EuroFIR project, a EuroFIR Food Classification system has been proposed, updating the existing work. These various developments are reviewed in this section.

6.1. EFG (Euro Food Groups) proposal

The workshop of the COST Action 99 – Eurofoods working group held in December 1999 considered it important that practical progress was made in the international comparison of food consumption surveys. A key question is the level of food grouping at which results should be compared. A comparison of several food classification systems used for food consumption (Volatier, 1999) had shown that although much of the categorisation was compatible, a significant problem was their divergence in the assignment of categories between the top and second levels in the classifications. By extracting all the groups that appeared at the top level in at least one of the systems, a list of 27 possible first-level groups was produced. This list was discussed and refined at the workshop, resulting in a set of 33 food groups, which were named the Euro Food Groups and form the categories of the EFG classification system. The EFG groups are listed in Annex D.

The EU **EFCOSUM** project (Ireland *et al.*, 2002) compared the EFG food grouping to food classification systems used in food consumption surveys, showing that in general the groups compare well. Most countries participating in the EFCOSUM project found that they could use

the EFG system to classify foods in food consumption surveys, but also indicated the need for further instructions to enable foods to be assigned to the different EFG groups in a comparable way (Ireland, 2005). The EFCOSUM project recommended using the EFG system for food classification in food consumption surveys, using Eurocode definitions for the categories.

6.2. Eurocode and LanguaL

The faceted food description language LanguaL was briefly introduced in section 1.5.2. LanguaL facet A indexes the *Product type* of the food, but it is difficult to provide a single structure for this as the categorisation depends on a user's requirements and perceptions. These differences result in the various differing systems used for food classification, which fulfil particular purposes. Therefore, in LanguaL 2000 (Møller & Ireland, 2000), facet A was upgraded to contain a range of widely used food classification systems, with the Eurocode 2 system contained in the list. Although Eurocode categories were not in that version of the thesaurus, the Eurocode main groups are included in the current version²⁰. It needs to be decided if one or more further levels of Eurocode 2 should be incorporated into LanguaL facet A.

6.3. EuroFIR Food Classification

The European Food Information Resource (EuroFIR)²¹ is a five-year Network of Excellence EU-funded project with the objective of improving the availability, harmonisation and quality of food composition data in Europe. Most European national food composition data compilers are participating, operating within the EuroFIR Compiler Network. The national databases generally use country specific food classification systems, but a major element of harmonisation would be the implementation of a single classification system (supplementing or replacing the national system, as decided by each national compiler). Thus the EuroFIR project is developing the EuroFIR Food Classification system (Ireland, 2005; Ireland, 2006) to be used in European food composition databases, and also to be incorporated in LanguaL facet A as a separate food classification. The EuroFIR Food Classification makes extensive use of the Eurocode documentation for its category definitions.

7. Further and related development

This Position Paper was written as a statement of the status of the Eurocode 2 Food Coding System at the end of the COST Action 99 – Eurofoods project, which terminated at the end of 1999. It has been updated within the initial phases of the EuroFIR project and the main developments during the intervening period have been noted in section 6, above.

At the end of 1999, there seemed to be considerable interest in further work on the Eurocode 2 system, as demonstrated by the work planned on the Euro Food Groups proposal. Various aspects required further work in order to make Eurocode 2 an operational system for recording and managing information on foods. The first subsection below discusses the short-term tasks

²⁰ http://www.langual.org/langual_thesauri.asp

²¹ <http://www.eurofir.net>

that are needed for this, largely as envisaged in 2000, but also acknowledging related work since then, as noted in the introduction to section 7.1.

The second subsection reviews those tasks that should be seen as significant for the further development of effective solutions for the recording of information categorising, describing and otherwise identifying foods, and for grouping and aggregating the resultant food data.

7.1. Current and short-term tasks

As discussed in previous sections, the Eurocode 2 version 99/2 revisions and the EFG system provide the basis for a detailed categorisation and a harmonised aggregation, respectively, that can contribute to an overall solution. Although the systems are intended for use with consumption data, the linking of foods in consumption and composition data is a key task that has to be supported through food classification and description systems. This must be achieved through mapping between the categories of the two systems, which requires that these categories are consistently defined between the systems. The Eurocode review project paid particular attention to improving the definition and documentation of the categories used in the Eurocode 2 system, and thus these should be suitable as the basis for linking between consumption and composition data. Indeed, due account of this has been taken in the development of the EuroFIR Food Classification and much of that work can be seen as contributing to the testing of the EuroFIR categorisation.

There are several tasks that should be carried out in the short-term in order to validate the work to date. The Eurocode Review project has been a paper exercise in which the previous work has been carefully considered, proposals for improvements have been made, and these proposals have been discussed. The essential next step is that the proposals are subjected to practical testing. Tasks to achieve this are outlined here.

7.1.1. Testing the Eurocode 2 categorisation

The proposed Eurocode 2 categorisation needs to be tested to evaluate its suitability for recording reported data, its effectiveness in the retrieving relevant food records in response to realistic queries, and its ability to map to grouping categories that can be used for data aggregation. Ideally, testing should involve the coding of a test file of consumption data that can also be coded using alternative systems, one of which would be the grouping system (such as a national system) that would normally be used on the data. Since the normal coding for a data set might be expected to be to some extent optimised for it, the testing should cover at least two independent test files. This would also allow the performance of Eurocode 2 as a standard coding to be tested for the retrieval, comparison and aggregation of data from different sources.

The testing should investigate specific aspects of the performance of the Eurocode 2 categorisation. The experimental design should specify these fully but some aspects to be covered might include:

- a separate assessment of the effectiveness of the categorisation in each Main Group and an analysis of the characteristics of the Group that may determine this
- a review of the performance in the reporting of mixed products and dishes
- an assessment of the use of the currently proposed standard subcodes and their potential for use in other parts of the categorisation
- an assessment of the requirements for food description that complements the categorisation.

Generally food composition databases have a more limited, controlled and well-defined set of food items than may be encountered in food consumption data records. A common coding of the two would allow the degree of matching that might be achievable between food records in each to be assessed. Thus a simplified but useful initial approach to testing might be the coding of one or more food composition databases. This would provide the opportunity to assess the discrimination that the version 99/2 categorisation affords between food items and also the relationship between the categorisation and the additional requirements for food description that are needed to identify a food item.

7.1.2. Nutrient Loss or Gain applications

The applicability of a food grouping based on Eurocode 2 to the management of Nutrient Loss or Gain (NLG) factors was discussed in section 3.6. Within the COST Action 99 – Eurofoods project, a Short-Term Scientific Mission was undertaken at the Swedish National Food Administration to work on *Yields for Recipe Calculation*. This was an initial investigation of the data management of NLG factors, including the derivation of typical values associated with the food group and the process involved. It is intended that this work will indicate how well NLG factors correlate with the type of food involved, and the extent to which such a correlation corresponds to the grouping system(s) generally applied to the foods. Part of this work included the development of a facility for coding the foods in an NLG database according to a two-level grouping system, including the Swedish PC-kost food groups and the version 99/1 Eurocode 2 groups and subgroups. The software also supports the coding of data with food descriptors.

The EuroFIR project includes work on recipe calculation, together with the collection and management of NLG data. The work acknowledges the relation of nutrient retention factors to food group, so far without addressing harmonisation of the food group representation for NLG data. Nutrient retention factors are increasingly applied at the ingredient level in recipe calculation, rather than to the overall recipe (to which yield factors are usually applied). Since the information for ingredients in recipe calculation normally is taken from a food composition database, it would be most convenient to use the composition rather than consumption food grouping system in the management of nutrient retention factors. Thus the EuroFIR Food Classification system might provide the basis for the food group association with factors, although since it shares the same main-group structure with Eurocode 2, the difference may not be significant. However, the suitability of the proposed food grouping systems should be reviewed and, where other considerations fail to resolve some details of the grouping, the NLG aspect might provide a convenient answer.

7.1.3. Eurocode 2 categorisation software tools

The assignment of Eurocode 2 category codes to food data records requires software support to ensure that the process can be performed efficiently and accurately. A further requirement for software is for the centralised maintenance of the thesaurus, together with the dissemination of the resulting updates.

Eurocode 2 category codes may need to be assigned to either food consumption or composition records. Although the facility may be incorporated into the overall data management system in use, this may not always be practical, and differing implementations and user interfaces may give variable results. Also, for many applications, the production of a separate list associating food records with their food classification (and/or food description) may be a convenient alternative.

Thus a software tool supporting the assignment of food classification and description is required, and its availability should encourage the uptake of Eurocode.

Two software packages provide some support for Eurocode assignment. The Food Product Indexer (FPI) is being used in the EuroFIR project to assign LanguaL food description to food composition records in European national databases. Since LanguaL incorporates Eurocode main groups in its facet A, the top-level Eurocode groups can be assigned as LanguaL codes. Extended Eurocode support could be provided by incorporating further levels of Eurocode into LanguaL facet A or by producing a separate FPI implementation for Eurocode. The Food Table Input (FTI) system for the management of food composition databases (Unwin & Becker, 2002) also has a facility enabling the assignment of Eurocode codes to food items.

As these tools provide for the assignment of food classification and description that may be separate from the main data management, the software needs to import and export the food data to be classified and indexed. The exchange format should be compatible with the general specification for data interchange within the EuroFIR project, simplifying the export of food information from the national database management system and the import of the results. Thus an XML format should be defined and supported, although the software might additionally offer support for other formats, such as text and spreadsheet files.

7.2. Further development

The short-term tasks described above would complete the Eurocode review project by testing the proposed revisions and finalising a specification for the completed Eurocode 2 categorisation system, for example by defining how extensively it should implement structured codes and standard sets of subcodes. Also they would form a basis for deciding how the categorisation system should manage information on mixed products and dishes both internally and through separate support for recipe information.

The Eurocode 2 Review project has indicated that food coding should use a revised categorisation system supported by a descriptor system with a clearly defined role in specifying modifications to a basic food item²². It is important that the descriptor system is designed in conjunction with the categorisation and this argues for a Eurocode descriptor system. However it may be impractical to undertake such a development and it would be undesirable to create an independent system partially overlapping the existing ones, and potentially overburdening and confusing potential users of food description systems. Even if the descriptor system used with the Eurocode classification must fulfil different requirements to existing systems, its design and specification should be coordinated with theirs so that the capabilities are complementary.

As discussed earlier, the LanguaL and INFOODS systems take very different approaches to recording food description information and it has been suggested that neither yet provides a full solution (Burlingame, 1998). Both also outline proposals for recipe support but both have yet to finalise specifications. A possible reason that this has not progressed more rapidly is that both systems are essentially homogeneous, built around the concept of facets which are independent in

²² This contrasts with the earlier view that Eurocode should directly record foods in their consumed state. Using that approach, it is difficult to know what can be assumed about a food (and the reported amount consumed) and what must be explicitly coded.

terms of the aspect they record, but which seek to operate as far as possible in a consistent way across their facets. In contrast, the Eurocode concept is more heterogeneous, providing a single basic classification which can be supported by supplementary and possibly diverse systems for descriptor and recipe support. This provides the opportunity to be flexible in developing prototype systems that could be tested prior to their incorporation as upgrades to the current versions of the LanguaL and INFOODS systems. Together with the contribution of the categorisation system to the overall goal of food identification, the further development of Eurocode can play a key role in evolution of food information handling techniques.

The following sections outline more specific areas and activities for further development.

7.2.1. Enhancing the classification

It has been proposed that any application-specific organisation of Eurocode 2 categories into classification hierarchies should be separate from the enumeration of the categories and their placement within a simple, multi-purpose grouping. Areas of the categorisation system and associated simple grouping that may require further work as a result of the testing described above include:

- the definition of general principles for placement now that the requirements have been simplified, which will, for example, provide a basis for deciding whether fruit and vegetable juices are best classified by their source or as beverages
- policy for classifying parts of organisms (for example, whether the leaves and root of a plant should be in separate subgroups, adjacent in the same subgroup or distinguished by descriptors) and meat cuts
- wider application of structured category codes and standard lists of subcodes
- determination of the geographical coverage of the classification (e.g. Europe or wider) and provision of a level of classification coverage consistent throughout that area
- assessment of the requirements for handling mixed foods, of how the classification, descriptor system and recipe system should be used to meet these requirements, and of whether one or more main groups for mixed dishes should be introduced
- definition of policies for, and enumeration of, grouping headings.

7.2.2. Designing other system components

The best and most realistic option for developing the descriptor and recipe components of the system is the development of facilities in collaboration with further work on the LanguaL and INFOODS systems, and on recipe information management. The Eurocode 2 categorisation system provides a basic component of an overall system to identify food items. The overall system requires additional description and recipe facilities and the Eurocode system could provide a suitable environment for prototyping their further development.

As discussed in section 2.2, within an overall Eurocode system food description is used to add to the information represented by the assigned category. Food description that is implicit for a category, e.g. as *dried* for Raisins, is not separately recorded when coding a food item. However, enhanced categorisation support may include any implicit descriptors such as *dried* as part of the category 'package' so that these can be used to generate a full indexing of the coded food items.

Section 5.1 of this paper gives an extended (but not necessarily comprehensive) discussion of the information that should be supported within the Eurocode system by its descriptor component. Further work is required on this but some probable requirements for the handling of food description can be listed. In addition to a faceted design, with capability for the hierarchical organisation of terms within a facet, facilities might include:

- separate management of thesauri applicable to more than one facet²³
- parallel recording of controlled and free language representations of food description
- management of related facets and/or of sub-facets
- management of information on sequential processes²⁴
- association of quantitative or semi-quantitative information with descriptors.

Eurocode requirements for recipe support need to be defined in conjunction with further work on the procedures for the categorisation of mixed products and dishes. The recipe system will need to identify ingredients and their amounts must be recorded. The processes applied both to individual ingredients and the overall dish must be described. It may be a requirement that yield and nutrient retention factors (NLG data) should be recorded for processes and dishes when this information is available.

7.2.3. Developing the documentation

Although printed versions of the documentation will have their uses, the primary source should be materials available over the WWW. The resources developed so far within the Eurocode 2 Review project consist of a listing of the classification with additional documentation of categories, arranged as two sets of separate Web pages for each Main Group.

Additional aids need to be developed. The provision of examples would be straightforward²⁵. However the preparation of alternative presentations of the classification, for example alphabetical listings or listings of synonyms, requires convenient reformatting procedures so that such secondary outputs can easily be updated when the primary classification document (or database) is revised. Such alternative output formats should be generated by software used to manage the classification. The outputs should include HTML or XML formats that can be incorporated into the documentation pages on the WWW.

The management of the documentation as a database with facilities to select and reformat output can provide, as described in the previous paragraph, a range of semi-dynamic materials. Alternatively, the database could be dynamically accessed through a Web server, allowing the

²³ Including equivalent sub-facets of different facets. Location is an obvious case, as demonstrated in the use of LanguaL facet R in several food description fields of the recommended Food record data structure (Schlotke *et al.*, 2000.; Part II, section 1.10).

²⁴ The order in which processes are applied in food preparation might have an effect on the composition. A separate, but analogous, requirement for sequential documentation arises in data documentation, for example for derived and standardised values. A common approach to the two issues might be beneficial.

²⁵ At least in a 'system' sense. However the examples included in the 93/1 draft documentation indicate that there may be difficulties in creating and maintaining a consistent set of accurate and helpful examples.

database to be searched, and retrieved information to be reformatted in response to user requests. It might be difficult to justify this approach for documentation only relevant to Eurocode users, at least until the system were widely used, but possibly the documentation could be developed as a useful WWW resource for any user interested in information describing and identifying foods.

An intermediate option would be to incorporate client-side scripting that implemented some capabilities for filtering and displaying the data contained within the downloaded Web page. Another alternative would be the distribution to registered users of access software that would collect and manage information from the Web site²⁶. This could be down-loadable but alternatively might be distributed on CD or DVD disk with additional resources such as images to avoid delays while these download. Any developments using HTML should take account of the likelihood that it will be superseded by XML which may provide a unifying data format for management of the documentation outside a database environment. Development related to the EuroFIR project would be XML based. Since LanguaL is a priority for EuroFIR, it should be considered whether the existing Eurocode documentation might be developed as a joint resource to support the complementary systems for food classification and description.

7.2.4. Development of software tools

The development of basic coding facilities for assigning Eurocode 2 categories to food records was discussed as a short-term task in section 7.1.3. However any design for an overall Eurocode system (or any other food identification and description system) that supports the suggested facilities will involve relatively complex data structures rather than a list of independent descriptor codes. For example, if a mixed dish is described using a recipe, the ingredients and their amounts must be recorded, possibly with description of processes applied both to individual ingredients and the overall dish.

It is increasingly becoming possible for a software tool to interact directly with data in a range of environments, including multi-user server and local single-user databases, spreadsheets, and HTML, XML or text files. However the software must be able to access the correct part of the data structure and this would be greatly facilitated by the use of standard names for database tables and fields (and their implicit equivalents in other types of software). The data management and interchange recommendations (Schlotke *et al.*, 2000) are being extended in the EuroFIR project and should cover the full requirements of food description and recipe information. It is also suggested that the database table names should include a specific prefix to identify them as conforming to the guidelines, rather than using general names such as 'Foods' and 'Values'²⁷. In this way a software tool would only attempt to access correctly structured data.

The basic coding facility described in section 7.1.3 should be enhanced to give coders effective access to thesauri and other documentation. They will need to use a software tool that provides an effective coding environment for the accurate and efficient assignment of classification and

²⁶ This might be a candidate as a "Web service". Web services are described in the EuroFIR document *Report on specifications and plan for development of the EuroFIR databank system*, which is under preparation

²⁷ Other problems can arise with simple names. For example, 'Values' is a reserved word in some programming environments and can cause problems as a table name in SQL statements unless it is protected as "[Values]".

descriptor terms, and also of free-text input and recipe information. The software should also be able to interact with online documentation and latest version releases available over the WWW. Annex E lists some thesaurus lookup and coding assignment features that should be included in the specification of an enhanced coding program.

8. Conclusions and recommendations

The Eurocode 2 Review project has produced revisions to the Eurocode classification that should prove useful in the implementation of the system for recording information on foods in consumption and other data. Some points remain to be resolved but this should be done on the basis of practical experience with the system.

Review project discussions and Eurocode revisions have identified new possibilities for the further development of Eurocode into a general facility for recording and processing information on foods. The EFG proposal of food groups for the international comparison of food consumption data is a grouping system that can be superimposed on the basic Eurocode 2 classification that provides the underlying categorisation. Different, application-specific groupings can be similarly superimposed without affecting the categorisation system, allowing the latter to concentrate on identifying as precisely as possible the basic food product being reported, which would allow consistent category definitions to be used in consumption and composition data. Data grouping and aggregation can then be implemented as a mapping between the grouping system and the recorded categorisation information, although a basic, generally agreed classification in the categorisation system may simplify the mapping procedures.

The categorisation of the basic food product should be supported by additional facilities that further identify the food by recording as food description the processes applied to it and by detailing its recipe in the case of a mixed food. This overall system also requires the convenient availability of documentation and software tools to provide a comprehensive and effective solution to the requirements of recording information on foods. The continuing development of the overall system should be undertaken in collaboration with further work on the related LanguaL and INFOODS food description systems.

This paper contains a range of proposals and suggestions for enhancing the present approaches to handling various aspects of information identifying foods. These are mostly covered by a number of general recommendations:

- Eurocode 2 version 99/2 should be subjected to a modest practical test, supported by simple coding software. Further revisions based on the results and a study of the benefits of a general structuring of codes should be applied and the resulting categorisation should be launched as Eurocode 2000.
- If the evaluation of the EFG proves positive, a mapping from Eurocode 2000 to the EFG should be produced and implemented in a software tool. In developing the mapping, a second-level set of EFG might be proposed.
- An overall system for recording information identifying food items should be developed based on a combination of food categorisation, food description and recipe information, and in close collaboration with any further development of LanguaL and the INFOODS food description system.

- The resources developed for recording information on food items should contribute to the extension of recommendations on data management and interchange, with the proposed data structures being prototyped in software tools that are developed in parallel to support the thesaurus management and data recording requirements of the information resources.
- Descriptions and updated results of further development should be immediately disseminated over the Internet and available for comment and other input, with the submission of additional documentary information on foods being particularly encouraged. This documentation should be considered one of the primary assets of the work.

The methods available to develop information management systems, to collaborate internationally in this development, and to disseminate proposals, results and documentation have progressed enormously over the last fifteen years. The Eurocode 2 Review project has provided a basic but detailed food categorisation system and a range of proposals and recommendations. It is hoped that these results will make a significant contribution to the development of an overall system for recording information on reported food items and interfacing this with application-specific systems for the grouping and aggregation of food consumption data.

9. References

- Bergström, L. (1994). *Nutrient Losses and Gains in the Preparation of Foods*, Report No. 32, National Food Administration, Uppsala, Sweden.
- Bognár, L. (2002). *Tables on weight yield of food and retention factors of food constituents for the calculation of nutrient composition of cooked foods (dishes)*, Report No. 02-03, Bundesforschungsanstalt für Ernährung, Karlsruhe, Germany. Available at: <http://www.bfa-ernaehrung.de/Bfe-Deutsch/Information/e-docs/Bognara2002.pdf>.
- Burlingame, B.A. (1998). *Food Nomenclature and Terminology: Standards and Harmonisation for Food Composition Databases and Food Trade*.
- Codex Alimentarius Commission (1999). *Codex General Standard for Food Additives (GSFA)*. Joint FAO/WHO Food Standards Programme. Document no. CL 1999/15-FAC, Annex B.
- Hendricks, T.C. (1992). *LanguaL: An Automated Method for Describing, Capturing and Retrieving Data about foods*. In: *International Food Data Bases and Information Exchange*, World Rev. Nutr. Diet. **68**, eds A.P. Simopoulos & R.R. Butrum. Karger, Basel, Switzerland, pp 94-103.
- Ireland, J. (2005). *Food identification in European food composition databases*. Report of EuroFIR Work Package 1.6.
- Ireland, J. (2006). *EuroFIR food classification*. Proposal of EuroFIR Work Package 1.6, dated 13 March 2006.
- Ireland, J., van Erp-Baart, AMJ, Charrondiere, UR., Møller, A., Smithers, G. & Trichopoulou, A. (2002). *Selection of a food classification system and a food composition database for future food consumption surveys*. *European Journal of Clinical Nutrition*, 56 Supplement 2, S33-S45
- Ireland, J. & Møller, A. (2000). *Review of International Food Classification and Description*. *J. Food Comp. Anal.* **13** (4), 529-538.
- van Kappel, A.-L. (1993A). *Draft Report: Evaluation of the Eurocode-2 Food Coding System*. FLAIR Eurofoods-Enfant Concerted Action and International Agency for Research on Cancer (IARC), World Health Organization, Lyon, France.
- van Kappel, A.-L. (1993B). *Report on the Evaluation of the Eurocode-2 Food Coding System*. In: *Report of the Third Annual FLAIR Eurofoods-Enfant Project Meeting*, eds J. Castenmiller & C.E. West. FLAIR Eurofoods-Enfant Project, Wageningen, The Netherlands, pp 173-174.
- Lagiou, A., Valaora, A., Vassilakou, T. & Trichopoulou, A. (1992). *Comparability of food availability data in household budget surveys in European Community Countries*. *Eur. J. Clin. Nutr.* **46** (Suppl. 5), S35-S136.
- Løken, E.B. & Borgejordet, A. (1998). *EUROCODE 2 – Some comments to the modified Core Classification presented at the COST Action 99 meeting in Wageningen 24-26 October 1998*. Working paper of the COST Action 99 – EUROFOODS Research Action on Food Consumption and Composition Data, Working Group on Food Description.

- Møller, A. & Ireland, J. (2000). *LanguaL 2000: Introduction to the LanguaL Thesaurus*. European Commission, COST report EUR 19542.
- OECD (1998). Three functional classifications of the 1993 SNA: COICOP, COPNI and COFOG. Joint OECD/ESCAP Meeting on National Accounts, 1993 System of National Accounts: Five Years On, Bangkok, 4-8 May 1998. Agenda item 8.
- Poortvliet, E.J., Klensin, J.C. & Kohlmeier, L. (1991). *Rationale Document for the Eurocode Food Coding System*. Institute for Social Medicine and Epidemiology, Berlin.
- Poortvliet, E.J., Klensin, J.C. & Kohlmeier, L. (1992). *Rationale Document for the Eurocode 2 food coding system (version 91/2)*. Eur. J. Clin. Nutr. **46** (Suppl. 5), S9-S24.
- Poortvliet, E.J. & Kohlmeier, L. (1993). *Manual for using the Eurocode 2 Food Coding System, Draft version, March 1993*. FLAIR Eurofoods-Enfant Project, Wageningen.
- Porubská, J. (2005). *FRI-28 vs. EUROCODE*. Working paper of EuroFIR Work Package 1.6, dated 11 July 2005.
- Schlotke, F., Becker, W., Ireland, J., Møller, A., Ovaskainen, M., Monspart, J. & Unwin, I.D. (2000). *Eurofoods recommendations for food composition database management and data interchange*. European Commission: COST report EUR 19538.
- Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. & Katz, S. (2004). *Reengineering Thesauri for New Applications: the AGROVOC Example*. J. Digital Inform. **4** (4), Article No. 257, 2004-03-17.
- Trichopoulou, A. & Lagiou, P. (1997). *Methodology for the exploitation of HBS food data and results on food availability in 5 European countries*. Report No. EUR 17909 EN, Office for Official Publications of the European Communities, Luxembourg.
- Trichopoulou, A. & Lagiou, P. (1998). *Methodology for the exploitation of HBS food data and results on food availability in six European countries*. Report No. EUR 18357 EN, Office for Official Publications of the European Communities, Luxembourg.
- Truswell, A.S., Bateson, D.J., Madafiglio, K.C., Pennington, J.A.T., Rand, W.M. & Klensin, J.C. (1991). *INFOODS Guidelines for Describing Foods: a systematic approach to describing foods to facilitate international exchange of food composition data*. J. Food Comp. Anal. **4** (1), 18-38.
- United Nations (1998). *Central Product Classification (CPC), Version 1.0*. Statistical Papers, Series M, No. 77, Ver. 1.0, Sales No. E.98.XVII.5.
- Unwin, I.D. (1992). *Food Naming and Description using Faceted Descriptors*. In: Report of the Second Annual FLAIR Eurofoods-Enfant Project Meeting, eds J. Castenmiller & C.E. West. FLAIR Eurofoods-Enfant Project, Wageningen, The Netherlands, pp 77-84.
- Unwin, I.D. & Becker, W. (2002). *Software management of documented food composition data*. J. Food Comp. Anal. **15**: 491-497.
- Unwin, I.D. & Møller, A. (1996). *Data Interchange Formats as import/export formats for food database management systems*. National Food Agency of Denmark, Søborg, Denmark.
- Unwin, I.D. & Møller, A. (1998). *Eurocode 2 Review and modified Core Classification*. Danish Veterinary and Food Administration, Søborg, Denmark.
- Vickery, B.C. (1960). *Faceted Classification*. Aslib, London, U.K.

Volatier, J.-L. (1999). Comparison of Eurocode 2 with the WHO GEMS food classification used for regional diets. Presentation to the COST Action 99 – EUROFOODS Research Action on Food Consumption and Composition Data, Working Group on Food Description, Paris, 3-5 December 1999.