



FNS – Cloud

Food Nutrition Security

Food Nutrition Security Cloud

Deliverable 4.5

Mapping tools for existing food intake and consumer behaviour data

Due Date:	30.09.2021
Submission Date:	31.01.2022
Dissemination Level:	Public
Lead beneficiary:	UCD
Main contact:	Eileen Gibney Eileen.gibney@ucd.ie Laura Bardon laura.bardon@ucd.ie

Project acronym: FNS-Cloud **Project Number:** 863059

Start date of project: 01.10.2019 **Project duration:** October 2019 – September 2023



Food Nutrition Security Cloud (FNS-Cloud) has received funding from the European Union's Horizon 2020 Research and Innovation programme (H2020-EU.3.2.2.3. – A sustainable and competitive agri-food industry) under Grant Agreement No. 863059 – www.fns-cloud.eu

Information and views set out across this website are those of the Consortium and do not necessarily reflect the official opinion or position of the European Union. Neither European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use that may be made of the information contained herein.

D4.5 Mapping tool for existing food intake and consumer behaviour data

Document Control Information																																					
Title	<i>D4.5 Mapping tool for existing food intake and consumer behaviour data</i>																																				
Editor	<i>Eileen Gibney, Laura Bardon & Grace Bennett (UCD), Barbara Korousic Seljak, Tome Eftimov, Peter Korosec & Gordana Ispirova (JSI), Maria Glibetic, Mirjana Gurinovic, Agi Kadvan, Jelena Milesevic & Milica Zekovic (CAP)</i>																																				
Reviewers	Paul Finglas (QIB) & Siân Astley (EuroFIR)																																				
Dissemination Level	<input type="checkbox"/> CO - Confidential <input checked="" type="checkbox"/> PU - Public																																				
Approved by	<table border="0"> <tr> <td><input checked="" type="checkbox"/> RTDS (COO)</td> <td><input checked="" type="checkbox"/> UM</td> <td><input checked="" type="checkbox"/> ILSI</td> </tr> <tr> <td><input checked="" type="checkbox"/> QIB (SCO)</td> <td><input checked="" type="checkbox"/> NUTRIS</td> <td><input checked="" type="checkbox"/> BfR</td> </tr> <tr> <td><input checked="" type="checkbox"/> JSI</td> <td><input checked="" type="checkbox"/> RIVM</td> <td><input checked="" type="checkbox"/> AUTH</td> </tr> <tr> <td><input checked="" type="checkbox"/> UCD</td> <td><input checked="" type="checkbox"/> WUR</td> <td><input checked="" type="checkbox"/> FEM</td> </tr> <tr> <td><input checked="" type="checkbox"/> PMT</td> <td><input checked="" type="checkbox"/> UGent</td> <td><input checked="" type="checkbox"/> CNR</td> </tr> <tr> <td><input checked="" type="checkbox"/> JDLC</td> <td><input checked="" type="checkbox"/> IMDEA</td> <td><input checked="" type="checkbox"/> APRE</td> </tr> <tr> <td><input checked="" type="checkbox"/> EuroFIR</td> <td><input checked="" type="checkbox"/> HUA</td> <td><input checked="" type="checkbox"/> CAP</td> </tr> <tr> <td><input checked="" type="checkbox"/> UWTSO</td> <td><input checked="" type="checkbox"/> TUM</td> <td><input checked="" type="checkbox"/> UNIFI</td> </tr> <tr> <td><input checked="" type="checkbox"/> DTU</td> <td><input checked="" type="checkbox"/> GS1</td> <td><input checked="" type="checkbox"/> LIFE</td> </tr> <tr> <td><input checked="" type="checkbox"/> ENEA</td> <td><input checked="" type="checkbox"/> SF</td> <td><input checked="" type="checkbox"/> Nutritics</td> </tr> <tr> <td><input checked="" type="checkbox"/> HYVE</td> <td><input checked="" type="checkbox"/> UoR</td> <td><input checked="" type="checkbox"/> EFF</td> </tr> <tr> <td><input checked="" type="checkbox"/> HYLO</td> <td><input checked="" type="checkbox"/> IFA</td> <td></td> </tr> </table>	<input checked="" type="checkbox"/> RTDS (COO)	<input checked="" type="checkbox"/> UM	<input checked="" type="checkbox"/> ILSI	<input checked="" type="checkbox"/> QIB (SCO)	<input checked="" type="checkbox"/> NUTRIS	<input checked="" type="checkbox"/> BfR	<input checked="" type="checkbox"/> JSI	<input checked="" type="checkbox"/> RIVM	<input checked="" type="checkbox"/> AUTH	<input checked="" type="checkbox"/> UCD	<input checked="" type="checkbox"/> WUR	<input checked="" type="checkbox"/> FEM	<input checked="" type="checkbox"/> PMT	<input checked="" type="checkbox"/> UGent	<input checked="" type="checkbox"/> CNR	<input checked="" type="checkbox"/> JDLC	<input checked="" type="checkbox"/> IMDEA	<input checked="" type="checkbox"/> APRE	<input checked="" type="checkbox"/> EuroFIR	<input checked="" type="checkbox"/> HUA	<input checked="" type="checkbox"/> CAP	<input checked="" type="checkbox"/> UWTSO	<input checked="" type="checkbox"/> TUM	<input checked="" type="checkbox"/> UNIFI	<input checked="" type="checkbox"/> DTU	<input checked="" type="checkbox"/> GS1	<input checked="" type="checkbox"/> LIFE	<input checked="" type="checkbox"/> ENEA	<input checked="" type="checkbox"/> SF	<input checked="" type="checkbox"/> Nutritics	<input checked="" type="checkbox"/> HYVE	<input checked="" type="checkbox"/> UoR	<input checked="" type="checkbox"/> EFF	<input checked="" type="checkbox"/> HYLO	<input checked="" type="checkbox"/> IFA	
<input checked="" type="checkbox"/> RTDS (COO)	<input checked="" type="checkbox"/> UM	<input checked="" type="checkbox"/> ILSI																																			
<input checked="" type="checkbox"/> QIB (SCO)	<input checked="" type="checkbox"/> NUTRIS	<input checked="" type="checkbox"/> BfR																																			
<input checked="" type="checkbox"/> JSI	<input checked="" type="checkbox"/> RIVM	<input checked="" type="checkbox"/> AUTH																																			
<input checked="" type="checkbox"/> UCD	<input checked="" type="checkbox"/> WUR	<input checked="" type="checkbox"/> FEM																																			
<input checked="" type="checkbox"/> PMT	<input checked="" type="checkbox"/> UGent	<input checked="" type="checkbox"/> CNR																																			
<input checked="" type="checkbox"/> JDLC	<input checked="" type="checkbox"/> IMDEA	<input checked="" type="checkbox"/> APRE																																			
<input checked="" type="checkbox"/> EuroFIR	<input checked="" type="checkbox"/> HUA	<input checked="" type="checkbox"/> CAP																																			
<input checked="" type="checkbox"/> UWTSO	<input checked="" type="checkbox"/> TUM	<input checked="" type="checkbox"/> UNIFI																																			
<input checked="" type="checkbox"/> DTU	<input checked="" type="checkbox"/> GS1	<input checked="" type="checkbox"/> LIFE																																			
<input checked="" type="checkbox"/> ENEA	<input checked="" type="checkbox"/> SF	<input checked="" type="checkbox"/> Nutritics																																			
<input checked="" type="checkbox"/> HYVE	<input checked="" type="checkbox"/> UoR	<input checked="" type="checkbox"/> EFF																																			
<input checked="" type="checkbox"/> HYLO	<input checked="" type="checkbox"/> IFA																																				
IPRs underlined	StandFood developed by JSI – additional information in exploitation plans																																				
Datasets underlined	Baseline Food4Me dataset; Serbian national dietary intake dataset 2017; Dutch NEVO dataset from the recipe work.																																				

Version/Date	Change/Comment
26/07/2021	<i>Outline Submitted</i>
30/09/2021	<i>Draft 1 submitted to General Assembly</i>
17/12/2021	<i>Draft 2 submitted to General Assembly</i>
18/12/2021	<i>Draft 3 submitted to reviewers</i>
31/01/2022	<i>Version 4 submitted with responses to reviewers' comments</i>

Table of Contents

1. Publishable Summary	4
2. Introduction	6
2.1 Dietary intake data.....	6
2.2 Mapping across dietary intake datasets	7
2.3 Consumer behaviour data	8
2.4 Deliverable aims and objectives	9
3. Normalisation of data from different sources.....	10
3.1 Tools to support harmonisation of intake data	11
4. Report on mapping activities and their outcome	14
4.1 Dietary intake data collected at food level	14
4.2 Dietary intake data collected at a food group level	25
5. References.....	37
Appendix A - Input data format	39
Appendix B.....	42

1. Publishable Summary

There is a need to improve the infrastructure which supports the reuse of data in the food nutrition and security domain. Dietary intake and consumer behaviour data are widely used to provide information on dietary and meal patterns, food contaminants and residues, the association of dietary intake with health outcomes, and inform determinants of food choice. Existing datasets within these areas are particularly fragmented which limits their reuse thus to overcome these issues and support data reuse there is a need for tools to facilitate the process of aligning or mapping food intake and consumer behaviour data. Development of specific tools in this area are needed due to the complex and diverse nature of both data collection and nomenclature (coding) used.

This deliverable reports on the activities undertaken within task 4.4.1 to advance mapping strategies for food intake data and the relevant learnings from these. Given the variance in levels at which dietary intake data can be collected, the mapping activities for these data took 2 separate approaches: 1) mapping at a food level for data collected using 24-hour recalls or diet diaries 2) mapping across food groups for data collected using food frequency questionnaires (FFQs).

Within the first approach, different food items and recipes from existing datasets, which were not harmonized or coded using FoodEx2, were matched to corresponding FoodEx2 terms and facets. For this task, two methods were applied. The first approach used StandFood to perform matching of uncoded foods, based on the English food names/descriptions. Using a test dataset, it returned an accuracy level of >70% for individual food items, and 32% for recipes. The second approach used a ML-based multi-class classification method to perform mapping based on food nutrient profiles. This method achieved 55% accuracy for individual food items. Whilst good results using these methods were achieved, challenges remain including the impact of human error from manual coding on ML approaches, the variance in names for similar foods across countries and languages, completeness of data within datasets and the lack of standardized formatting across datasets.

Within our second approach to dietary intake data, (data coded at a food group level) a visualization map of previously used FFQs was created. Mapping guidelines were developed

D4.5 Mapping tool for existing food intake and consumer behaviour data

across food groups, food items and frequency of response categories. The feasibility of this mapping system was then tested through a comparison exercise between two FFQs; Food4Me and NHANES. Good similarity was observed between mean food group intakes for some food groups whilst differences were identified across others. The main reasons for these differences arose from: variation between frequency of consumption response categories between FFQs, variations between food items listed within specific food groups across FFQs, presence of non-comparable food items which were listed in 1 FFQ but not captured within the other.

The tools and approaches described within this deliverable will be further used and tested within the WP5 demonstrator analyses. This deliverable advances the mapping strategies for food intake data for future use.

2. Introduction

There is a need to improve the infrastructure which supports the reuse of data in the food nutrition and security domain⁽¹⁾. The Food Nutrition and Security Cloud (FNS-Cloud) consortium was established with the overarching aim to overcome existing fragmentation issues within FNS data by integrating existing data and developing an infrastructure and services to exploit food, nutrition and security data (data, knowledge, tools – resources) for a range of purposes⁽²⁾.

The FAIR principles, act as overarching guidelines for data holders/researchers to put data in a position to be reused⁽¹⁾. Applying the FAIR principles to research data will be mutually beneficial to both scientific research and society. Recognising this, the European Commission (EC) has established an expert group which aims to turn the concept of FAIR data into reality in order to open up science and research⁽³⁾, and through EOSC, to federate existing research data infrastructures in Europe and realise a web of FAIR data and related services for science, making research data interoperable and machine actionable⁽⁴⁾.

Supporting data reuse in the area of FNS is a key aim of FNS-Cloud. It is recommended that instead of collecting new data, datasets already in existence could be reused or used in combination with other datasets to answer ongoing or future research questions. This is further described in FNS-Cloud deliverable 3.1. Dietary intake and consumer behaviour data are widely used to provide information on dietary and meal patterns, food contaminants and residues, the association of dietary intake with health outcomes, and inform determinants of food choice. Existing datasets within these areas are particularly fragmented which limits their reuse. Therefore, to overcome these issues and support data reuse there is a need for tools to facilitate the process of aligning or mapping food intake and consumer behaviour data. Development of specific tools in this area are particularly needed due to the complex and diverse nature of both data collection and nomenclature (coding) used.

2.1 Dietary intake data

Dietary intake data are used to determine food, nutrient and ingredient intakes at both individual and population levels⁽⁵⁾. Dietary data can be assessed in clinical settings to support personalised

D4.5 Mapping tool for existing food intake and consumer behaviour data

care, or within population groups for research purposes and to develop government policy and public awareness campaigns ⁽⁶⁾. The level of detail collected when assessing dietary intake is dependent on the method of collection chosen. The main methods of collection include food diaries, 24-hour recalls, diet histories and food frequency questionnaires (FFQs). Food diaries, 24-hour recalls, and diet histories collect intake data at an individual food level with detailed assessment of portion size. Food diaries and 24-hour recalls collect information on actual food eaten on a specific day(s) and use this information to estimate usual daily intake. FFQs collect intake data at a food or food group level over a longer period of time providing a crude estimate of usual daily intakes.

The complexity in the mapping and aligning of dietary intake data comes from the variety of methods available to collect the data in addition to the numerous sources of portion size data.

Once dietary intake data has been collected, it is linked to a composition dataset which generates a complete dataset containing detailed information on individual foods and their corresponding food groups, portion size (depending on the dietary intake collection method used), and nutrient composition. Harmonisation across dietary composition databases has greatly improved in recent years through use of standardised food description systems ⁽⁷⁾, and will not be the focus of this report, as this is covered elsewhere.

2.2 Mapping across dietary intake datasets

Harmonisation across datasets is aided by following standardised guidelines and practices during data collection ⁽⁸⁾. In order for food data to be easily reused, whilst ensuring the quality of the data is maintained, the description of food items (along with portion size) needs to be accurate, detailed and consistent. Foods should be described as clearly and accurately as possible to ensure the maximum amount of foods can be matched appropriately and the resulting analysis accurate ⁽⁹⁾. If food items aren't well defined even data of otherwise good quality can be compromised ^(10; 11). As dietary intake data can be collected, and then coded at different levels (e.g., individual foods, food groups, and food categories), the mapping of datasets collected using different methods needs to be considered carefully. For example, reported consumption of foods at an individual level, can be subsequently aggregated to food groups, and mapped to data from

D4.5 Mapping tool for existing food intake and consumer behaviour data

existing food groups, if the relevant information for categorisation into food groups is provided. Unless this hierarchical approach is transparent, combining different datasets together will be challenging.

Furthermore, FFQs are frequently designed to capture information on specific types of foods or foods that are not consumed widely within a certain population (e.g., to quantify consumption of seafood) so may not contain a comprehensive list of foods. On the other hand, 24hr recalls and food diaries aim to collect information on all foods and beverages consumed. In addition, differing methods collect data across different timeframes. 24hr recalls, collect information on a preceding day, and when several of these are collected over a given period (e.g., 2 or 3 within 1-2 weeks) one can estimate habitual intake. Whereas FFQs collect data from the preceding month, 6 months or year, assess habitual intake, but over a completely different timeframe. These subtle but important differences mean that the mapping and merging of dietary intake data cannot be a standardised function, and there will be a need for subject matter expert interpretation of both the data and the research question, to determine if the data are appropriate to be merged and which method to use.

The following document outlines the development of these tools in FNS-Cloud.

2.3 Consumer behaviour data

Consumer behaviour research addresses the understanding of human behaviour relating to the purchase and consumption of economic goods ⁽¹²⁾. Research in this area strives to understand associated consumer choices and related cognitive and emotional processes. Within the realm of food and nutrition, the study of consumer behaviour is centred on food choice, motives for food purchasing, and perceptions of risk in relation to certain foods and new food technologies. Knowledge of consumer behaviour, and the factors influencing the same, can inform food product development, dietary health public policy and marketing strategies ⁽¹³⁾.

The majority of methods to collect consumer behaviour data use qualitative approaches, therefore in most cases it is not possible to automatically map across these data. When conducting unstructured interviews or focus groups, researchers typically develop their own topic

D4.5 Mapping tool for existing food intake and consumer behaviour data

guide for use based on their individual research hypothesis. Furthermore, the included population is usually not representative and factors such as the dynamics between the participants and the level of training/experience of the researcher can influence the data arising from the study. The data gathered tends to inform the themes/results derived therefore comparisons are possible in some cases, but this is not consistent.

Conversely, data on the factors influencing food choice is typically collected using questionnaires⁽¹³⁾ which are quantitative in nature; hence, the resulting data is often comparable between different population groups (even across various countries). Whilst the data arising from questionnaires is quantitative, the development of questionnaires arises from qualitative research, therefore consideration should be given to the facets included in questionnaires, the population studied, and whether the tool has been validated⁽¹³⁾ before mapping and merging across different questionnaires.

Consumer behaviour and dietary intake are highly linked fields with a great deal of interaction between factors affecting both. Within the DEDI-PAC project, a framework was developed which visualises the Determinants of Nutrition and Eating (DONE Framework) across different population groups⁽¹⁴⁾. This framework is freely accessible (<http://uni-konstanz.de/DONE>) and is expected to evolve over time as experts add further determinants and ratings.

In 2021, the Communities on Food Consumer Science (COMFOCUS) project was launched (<https://fnhri.eu/projects/comfocus/>). COMFOCUS aims to reduce data fragmentation across food consumer science and to develop a library of meta data and digital service tools to form a knowledge platform within this area. Partners from within the FNS-Cloud consortium are also involved in this initiative. As the project develops further, FNS-Cloud will align with these experts' recommendations on how to harmonise research approaches and methodologies. The mapping of consumer behaviour data will not form the basis of this deliverable.

2.4 Deliverable aims and objectives

As previously mentioned, there is a need for tools/services to facilitate the efficient process of mapping and aligning datasets associated with food intake and consumer behaviour.

D4.5 Mapping tool for existing food intake and consumer behaviour data

The focus of this report is not on the statistical approaches which will be outlined in more detail in WP5, but the tools used to support the initial mapping and then combining of food intake data for subsequent analysis. Therefore, this deliverable will summarise approaches/tools used in the merging of data from different sources and report on the mapping activities undertaken and the relevant learnings from these.

Given the variance in levels at which dietary intake data can be collected, the mapping activities for these data will take 2 separate approaches: 1) mapping at a food level for data collected using 24-hour recalls or diet diaries 2) mapping across food groups for data collected using FFQs.

3. Normalisation of data from different sources

In recent years, advances have been made in data analysis due to the volume of data being generated and available. However, one of the main challenges that remains during data analysis is the process of how to normalize the data, since data from the same domain (e.g., dietary intake data) can be coded using heterogenous standards or systems. Data normalisation is a well-known task that can be defined as a procedure for organising data within a database, which acts to minimize redundancy (duplicate data) and prevents any issues stemming from database modifications such as insertions, deletions, and updates ⁽¹⁶⁾.

In relation to food data, data harmonisation can occur on multiple levels but is achieved initially by presenting data according to reliable classification and description systems ⁽⁸⁾. During the data collection phase harmonisation efforts focus on following standardised protocols and best practice guidelines. Once data has been collected, cleaning the data to generate a dataset focuses on coding the data using a harmonised approach. For example, food consumption data from different countries, may be coded using different systems, with FoodEx2 ^(17; 18) (codes developed by EFSA), and LanguaL ^(19; 20) being 2 of the most widely used to describe food products. Due to the range of methods and ways in which dietary intake data can be collected and produced, data harmonisation between different sources most frequently occurs retrospectively ⁽⁸⁾. As a result, harmonisation approaches must be flexible ⁽⁸⁾. The ability to accurately harmonise dietary intake datasets retrospectively is dependent on heterogeneity of studies and data collection tools and requires time, access to appropriate expertise, and adequate methodologies ⁽⁸⁾. Before

D4.5 Mapping tool for existing food intake and consumer behaviour data

harmonising different datasets, consideration should be given to the collection methods used and the how the data is handled. Domain experts should always consider the appropriateness of the data under consideration for their specific research question and whether the data from different sources are appropriate or possible to be merged.

Whilst Foodex2 and LanguaL are 2 of the most commonly used food classification systems, frequently dietary intake data is coded using nationally or locally developed food codes. This adds further complexity to the mapping process across similar datasets collected in different countries. Combining such datasets or even performing comparisons between the datasets is a time-consuming task as the process of mapping between the datasets must be completed manually. In recent years, much progress has been made in natural language processing (NLP) and machine learning (ML), where the focus has been on developing tools that can help the process of food data normalization. NLP has found applications in areas such as digital health interventions ⁽²¹⁾, extraction of dietary intake information from audio ⁽²²⁾ and extraction of nutrient and food information from recipes ⁽²³⁾. Whilst there are still improvements to be made in some of these techniques it is likely they will be utilised more widely in the near future.

Within the area of nutrition and dietary intake data, ML algorithms can be used to find unknown relations that exist between food entities and disease entities ^(24; 25). These relations are useful for informing hypotheses which can subsequently be tested in studies. The automated extraction of food information can be used to fill in missing values that appear in food databases such as food composition databases. Another application is the use of information extraction to extract food entities from dietary records for individuals (i.e., written as free-form text), and then map them on a nutrient level. This information can be combined and used by recommender systems.

3.1 Tools to support harmonisation of intake data

Several tools / resources exist to facilitate harmonisation of dietary intake data including StandFood, Foodbiz and others described in Table 3.1 below.

StandFood is a semi-automatic system for classifying and describing foods according to FoodEx2 ⁽²⁶⁾. Briefly, this system consists of three approaches:

D4.5 Mapping tool for existing food intake and consumer behaviour data

- 1) a machine learning approach that classifies foods into four FoodEx2 categories, with two for single foods: raw and derivatives, and two for composite foods: simple and aggregated
- 2) a natural language processing approach and probability theory to describe foods
- 3) a combined approach that uses the results from the first and the second part by defining post-processing rules in order to improve the result for the classification part

StandFood is a time-efficient method of finding missing FoodEx2 codes for food composition data and food consumption data that are basic resources that need to be linked and combined for dietary assessment methods ⁽²⁶⁾.

Table 3.1 Existing resources to facilitate harmonisation of dietary intake data

Resource name	Resource type	Description
FoodOntoMap ⁽²⁷⁾	Mapping tool	Data normalisation tool which targets the domain of food and nutrition science, normalizing food concepts across three different ontologies (FoodOn, OntoFood, and SNOMED CT) and one semantic dataset.
FoodViz ⁽²⁸⁾	Visualisation tool for annotated text	A web-based framework used to present food annotation results from existing Natural Language Processing and Machine Learning pipelines in combination with different food semantic data models.
FoodOn ⁽²⁹⁾	Ontology	FoodOn is an open-source, comprehensive ontology resource composed of term hierarchy facets that cover basic raw food source ingredients, process terms for packaging, cooking and preservation, and an upper-level variety of product type schemes under which food products can be categorized.
MESCO ⁽³⁰⁾	Ontology	A novel ontology for traceability in meat supply chains (MESCO, Meat Supply Chain Ontology). MESCO represents

D4.5 Mapping tool for existing food intake and consumer behaviour data

		key concepts of the supply chain such as Activity, Actor, Food Product, Service Product, Lot and Process.
FoodBase ⁽³¹⁾	Annotated corpus of food entities	Annotated corpus of food entities utilising recipe data to detect semantic similarities or differences between food concepts.
StandFood ⁽²⁶⁾	Food classifying and description tool	A Semi-Automatic System for Classifying and Describing Foods According to FoodEx2.

4. Report on mapping activities and their outcome

Food data can be collected using a variety of different methods however the data generated from them can be categorised into two basic levels: individual food and food group. In order to advance the capabilities of mapping across different food intake datasets, work on these two types of food intake data was undertaken simultaneously. The first aspect looked at mapping between food intake data collected at an individual food level. In other words, arising from food diaries or 24 hour recalls. The second approach focused on mapping across food intake data collected at a food group level, i.e., using food frequency questionnaires (FFQs). The tasks undertaken to advance the mapping capabilities between dietary intake data collected at either an individual food or food group level are described below.

4.1 Dietary intake data collected at food level

Overview of activities undertaken

The focus of the first approach was on coding food consumption data from several countries to the FoodEx2 code. The unharmonized food consumption data came from the EU Menu project, where the food products have been manually coded with FoodEx2 codes. Three use cases (UC) were performed on this data:

UC1: Match n= 120 food items using their English names and nutrient profiles with the corresponding FoodEx2 terms and facets

UC2: Match n= 120 recipes with their ingredients using their English names and nutrient profiles with the corresponding FoodEx2 terms and facets

UC3: Match food intake data with corresponding food composition data using only the English food names

Extension of the StandFood Method

In order to perform automatic data normalization for UC1 and UC2, the computational linguistics method, StandFood (developed as a part of the ERA-Chair ISO-FOOD project) has been extended⁽²⁶⁾. In brief, the StandFood method takes a food name description as input data, performs linguistics analysis of the food name, and then links it to the same food product that exists within

D4.5 Mapping tool for existing food intake and consumer behaviour data

the FoodEx2 dataset. The outcome is the list term that is assigned to the food product. The StandFood performance is related to the description level of the food name. In simple terms, the better the food description the better the matching results possible. For the purpose of this task, we extended StandFood with additional features (see Figure 4.1), consisting of three vertical pillars.

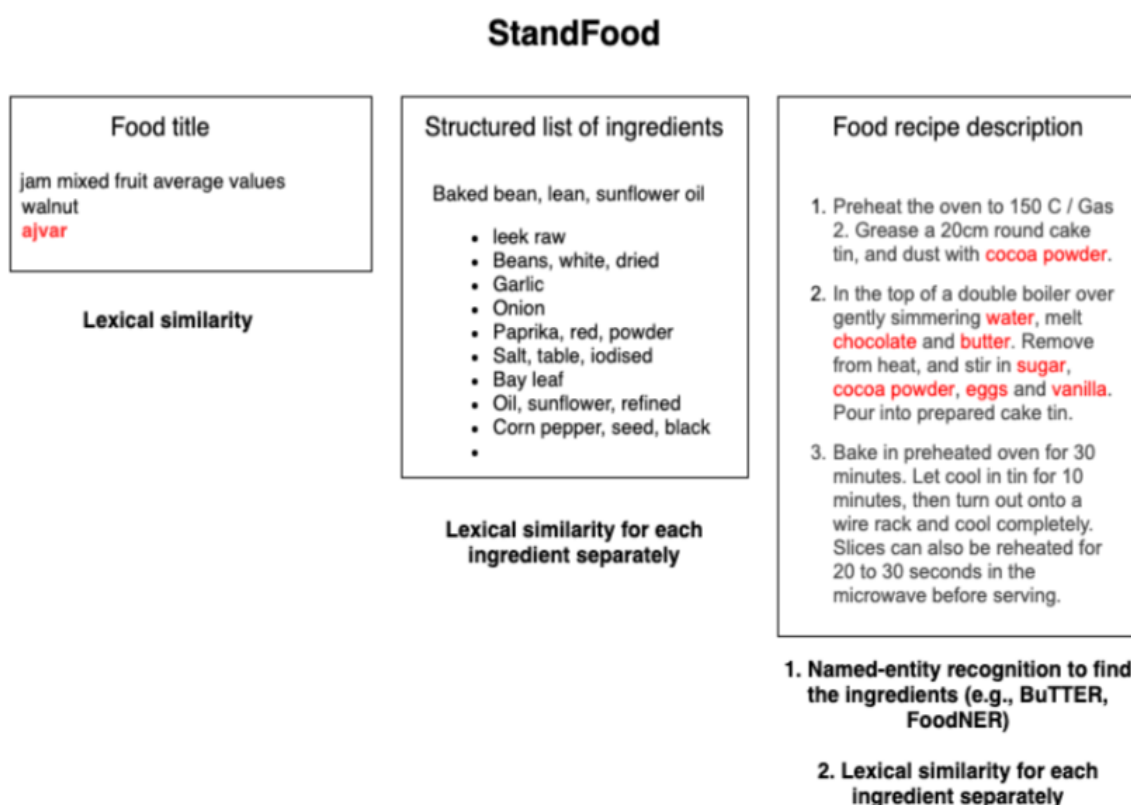


Figure 4.1. The StandFood pipeline

The first pillar is employed in cases whereby there is a good food product name. Using lexical similarity, the StandFood method will find the corresponding food product in FoodEx2 datasets, if it exists. However, there are recipes and traditional food products (e.g., ajvar which is a Serbian roasted red pepper relish) that are within dietary intake datasets collected at a national level across Europe that do not exist in the FoodEx2 database. In order to effectively code these food products/recipes StandFood requires a list of their ingredients to provide all FoodEx2 facets (Pillar 2). In these cases, all ingredients are coded together with the cooking processes associated with them. The second pillar is also used to code branded foods. The third pillar is used when the list

D4.5 Mapping tool for existing food intake and consumer behaviour data

of ingredients is not specified by the domain experts, but instead a description of the recipe's preparation is provided. Here, first named-entity recognition methods (NERs) should be utilized to extract the individual food entities from the text and then to code them in a similar manner as the second pillar. Further description of other NERs previously developed is provided within Deliverable 3.2 *Methodology for FNS data standardisation and interoperability*.

Results from Use Cases

The FoodEx2 dataset version 12 was used for this work, including four different food categories: namely, raw (r), derivative (d), simple (s) and composite (c) food products. All the food terms are structured in hierarchical order (from general terms and its descriptions to more specific ones) as defined by the FoodEx2 standard¹. We have also tried to perform experiments where different hierarchies from the FoodEx2 dataset have been involved in the matching process that belong to the following types of terms (C – core term, E – extended term, and M – generic term) in the basic food list. The individual datasets and type of data used within each use case are listed below.

UC1 – Match n= 120 food items using their English names and nutrient profiles with the corresponding FoodEx2 terms and facets

Data used: To perform the food matching exercise, a dataset from the Serbian dietary survey (collected during the pilot study in 2017) was prepared. This dataset consisted of 120 commonly consumed foods, divided into 12 food groups, which were identified based on the total quantity consumed in the adult population. Foods within the Serbian dataset are described in the Serbian and English languages and have data on 74 nutrients (macronutrients, vitamins and minerals). Foods are coded with FoodEx2 codes and facet descriptors from Exposure hierarchy.

UC1 results: The StandFood method takes the food product name and automatically generates one or more list terms from FoodEx2 that need to be selected/confirmed by a domain expert. This UC utilised the first pillar of the StandFood pipeline which resulted in 71% perfect matches. A 'perfect match' was defined as the code generated by the domain expert corresponding exactly with the code generated by the StandFood method. From these results it was apparent that the accuracy of the automatic matching is dependent on the translated English name of the food product (i.e., better translation equates to better results). Table 4.1 presents some examples of

¹ <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2015.EN-804>

D4.5 Mapping tool for existing food intake and consumer behaviour data

perfect food matches. In the last example, the domain expert provided a code for “Common walnut”, while the StandFood method provided a code for “Walnuts” that already existed in the FoodEx2 dataset.

Table 4.1: Examples of ‘perfect’ matches

Food name	Domain code	StandFood generated code
Apple, whole, raw	A01DJ	A01DJ
Cheese, feta	A02RC	A02RC
Walnut*	A0DXG	A014R

*FoodEx2 name: Common walnut

Table 4.2: Examples of ‘other’ matches

Domain expert code and food name	Foodex2 food name	StandFood code and food name
A02YM Cheese, kachaval	Cheese, emmental	A02QE Cheese
A039E Clotted cream, Kaymak	Blended margarine	A02QR Clotted cream
A033L Honey	Honey, polyfloral	A033J Honey
A00KX Lettuce, green leaf	Lettuces (generic)	A0DKM Baby leaf lettuces

Within table 4.2, examples of ‘other’ matches are displayed where the codes generated by the domain expert and the StandFood method are not perfect matches. However, on examination of the codes provided by StandFood it appears that the provided codes are correct; for example, the domain expert coded the clotted cream as blended margarine, whilst StandFood found a perfect match - A02QR Clotted cream. Furthermore, the cheese, kashkaval (typical cheese consumed within the Balkan region) was coded as Emmental by the domain expert, with the StandFood method suggesting the generic code for ‘cheese’, which is an acceptable code for this product and often used in food coding, when detailed information on the food product is missing. This UC example demonstrates that in many cases the English name of the food product is not correctly

D4.5 Mapping tool for existing food intake and consumer behaviour data

or completely translated from the colloquial or traditional name, which can confuse the automatic matching of food items, resulting in erroneous errors.

As well as exact matching at the food item level, the effect of different hierarchies of FoodEx2 on the automatic coding was examined. For this purpose, we select r, d, s, and c food products that belong to the following combinations of types of terms: E; C and E; or, C, E, and M. One example of this coding is the following:

- Domain expert: Milk cow whole 3/4 (FoodEx2: A02LY)
- StandFood (E) - cow milk whole - A02LY
- StandFood (C, E) - cow milk - A02LV
- StandFood (C, E, M) - cow milk - A02LV

From this example, it appears that the perfect match (A02LY) is only obtained when the extended term (E) hierarchy is used, however in the other two cases, the match found is only one level away in the hierarchy.

In addition to the StandFood approach, a ML approach was also attempted for UC1, where a ML based multi-class classification method for mapping nutrient profiles to FoodEx2 codes was tested. The training of the ML method was performed on European FCDBs, considering energy, macronutrients, sugar, water, vitamins (folate, riboflavin, thiamine, vitamin B12, vitamin B6, vitamin C), minerals (potassium, magnesium, sodium, phosphorous, zinc) as feature variables. The ML method offered three possible FoodEx2 terms as solutions. There are several weaknesses associated with this method. Firstly, the model is learning the 'noise' (poor coding choices) that are present in the manual coding. The ML method providing a poor solution as a result of noise by the manual coding is illustrated in the following example:

- Domain expert: A039E Clotted cream, Kaymak (FoodEx2 name: Blended margarine)
- StandFood: A02QR Clotted cream
- ML: A039E Blended margarine

Therefore, it is evident that the errors made by domain experts will also be adopted when a ML approach is used. Secondly, the ML approach cannot recognize FoodEx2 codes that are not present in the training data set. The percentage of perfect matches when using the ML approach was approximately 55%.

D4.5 Mapping tool for existing food intake and consumer behaviour data

UC2 – Match n= 120 recipes with their ingredients using their English names and nutrient profiles with the corresponding FoodEx2 terms and facets

Data used: Commonly consumed recipes within the Serbian national Food consumption survey were selected for this use case. Foods from these recipes are described with names in Serbian and English languages and with 74 nutrients (macronutrients, vitamins and minerals), as described above. Foods within recipes are coded with Foodex2 codes from the Exposure Hierarchy. Where applicable, the facet F28 Process, is added to some of the foods which reflects the process applied to foods used in the recipe. Some food names suggested the applied process without requiring additional coding.

Results: A list of 120 recipes with their ingredients was prepared for this UC. The second pillar of the StandFood pipeline, where each ingredient is matched separately, alongside their cooking processes (processes from FoodEx2 hierarchy + their past participle form) was used. The StandFood method appears to work relatively well for ingredients (Figure 4.3, where the yellow highlighted rows indicate perfect matches. Rows highlighted in green indicate an example of when the domain expert should select the correct match from the multiple automatically generated possible matches. The red highlighted row indicates an example of when the StandFood method identified a perfect match, but the domain expert missed this).

Food ingredient	StandFood name	StandFood FoodEx2 code	FoodEx2 code	pro_code	process name	RS process
1 leek raw	leeks	A005B	A005B	NA	NA	A005B#F28.A07GX
2 beans white dried	beans dry and similar.mung beans dry.broad beans dry.lablab beans dry.yardlong beans dry.stink beans dry	A012S.A013D.A01	A012T	A07KG	dried	A012T#F01.A05GX\$F28.A07GL\$F28.A07GX\$F28.A07MS
garlic	garlic	A00GZ	A00GZ	NA	NA	A00GZ#F28.A07GX
4 onion	onions	A00HC	A00HC	NA	NA	A00HC#F28.A07GX
5 paprika red powder	paprika powder	A019L	A019L	NA	NA	A019L#F28.A07GX\$F19.A07PH
6 salt table iodised	salt iodised	A042R	A042R	NA	NA	A042R#F19.A07PQ\$F28.A07GX
7 bay leaf	leaves	A0EST	A00VX	NA	NA	A00VX#F19.A07PN\$F28.A07GX
8 oil sunflower refined	olive oil refined	A036R	A037D	NA	NA	A037D#F10.A07XK\$F19.A16RP\$F28.A0C07\$F19.A07PR\$F28.A07GX
9 corn pepper seed black	black pepper	A019C	A018Y	NA	NA	A018Y#F10.A166Z\$F19.A07PL\$F28.A07GX
10 clotted cream kaymak	clotted cream	A02QR	A02ML	NA	NA	A02ML#F19.A16RP\$F10.A077A\$F28.A07GX
11 phyllo wheat flour	wheat flour	A003X	A16FH	NA	NA	A16FH#F19.A165A\$F28.A07GX
12 cheese white fresh unripened	fresh uncured cheese	A02QF	A02QG	A0C6F	ripened	A02QG#F10.A077A\$F10.A07XM\$F10.A166Y\$F19.A16RP\$F28.A07GX
13 egg hen whole raw	hen egg mixed whole dried	A031X	A031R	NA	NA	A031R#F28.A07GX\$F19.A07PQ
14 salt table iodised	salt iodised	A042R	A042R	NA	NA	A042R#F19.A07PQ\$F28.A07GX

Figure 4.3: Matching recipe’s ingredients to FoodEx2

Since recipe’s ingredients can’t be used to describe all of the FoodEx2 facets, the cooking process also needs to be identified. To achieve this, the cooking processes found in the food name

D4.5 Mapping tool for existing food intake and consumer behaviour data

description were also extracted and coded. Figure 4.4 provides an example of the process evaluation results, where the green rows correspond to examples where the cooking process coded by the domain experts is also found by StandFood, while the yellow rows correspond to examples where the cooking process is available in the food description name, extracted by StandFood, but the domain experts have not coded this information.

name	foodex2 code	process	Foodex2 code
macaroni, spaghetti, boiled	A007P#F02.A06CX\$F28.A07GL\$F28.A07KG	boiled	A07GL
oats, rolled	A00DH#F02.A068G\$F27.A000G\$F28.A07LH	rolled	A07LH
pork, cured, bacon	A022X#F02.A06AP\$F27.A01RG\$F28.A07KD	cured	A07KD
tuna, in oil, canned	A0FBT#F02.A0EMS\$F19.A07PH\$F19.A07PJ\$F27.A02DX\$F28.A0BYP	canned	A0BYP
egg, whole, boiled	A032B#F02.A06BX\$F27.A031G\$F28.A07GL\$F28.A07LC	boiled	A07GL
water, carbonated, vrnjci	A03DX#F02.A06CA\$F10.A06HS\$F27.A03DX\$F28.A07JN	carbonated	A07JN
water, knjaz milos, carbonated	A03DS#F02.A06CA\$F10.A06HS\$F27.A03DS\$F28.A07JN	carbonated	A07JN
chicken breast, grilled	A01SP#F01.A057Z\$F02.A0BYZ\$F27.A01SP\$F28.A07GZ	grilled	A07GZ
jam, mixed fruit (average values)	A01NH#F02.A0EMA\$F04.A04RK\$F04.A0BY6\$F08.A032K\$F10.A077J\$F28.A0E	mixed	A0CRL
hazelnut, dried	A014L#F01.A0CFH\$F02.A066R\$F27.A014L	dried	A0C0C
rice, white, long-grain, cooked	A001F#F01.A059Z\$F02.A066Q\$F27.A001F\$F28.A07GL\$F28.A07LK	cooked	A07HF
pepper, pickled	A00ZJ#F02.A0ELK\$F27.A00JA\$F28.A0C0P	pickled	A07KC
potato chips, plain, salted	A011L#F02.A0EPT\$F28.A07GV	salted	A07JP
ham, cooked, in casing	A023K#F02.A06AP\$F19.A16SA\$F27.A01RG\$F28.A07JP\$F28.A0BA1	cooked	A07HF
chicken, cooked in water	A01SP#F01.A057Z\$F02.A0BYZ\$F27.A01SP\$F28.A07GL	cooked	A07HF
egg, fried, salted	A032C#F02.A06BX\$F27.A031F\$F27.A031G\$F28.A07GR\$F28.A07LC	fried	A07GV
egg, fried, salted	A032C#F02.A06BX\$F27.A031F\$F27.A031G\$F28.A07GR\$F28.A07LC	salted	A07JP
bread, wheat, white, toasted	A004Y#F02.A06CN\$F04.A001M\$F10.A07XK	toasted	A07HC
pork, loin fillet, smoked	A022T#F02.A06AP\$F27.A01RG\$F28.A07KD	smoked	A16ET

Figure 4.4: A process evaluation results.

To summarize, out of 120 recipes 113 matches were found by the StandFood tool. For the 133 matches, 39 were correct, while the other 74 matches were incorrect.

D4.5 Mapping tool for existing food intake and consumer behaviour data

UC3 - Match food intake data with corresponding food composition data using only the English food names

Data used: For UC3, a list of 45 Dutch NEVO recipes and their ingredients was prepared. The selected 45 recipes are of different complexity. There are 4 complex recipes, 2 mixed dishes (one with and one without added fat), 6 nested recipes, 2 recipes based on label data, 15 simple recipes, 3 cookbook recipes, 2 recipes used in multiple recipes, and the rest of them were selected randomly. With respect to the number of ingredients, there are 14 recipes with 2 ingredients, 5 recipes with 3 ingredients, 6 recipes with 4 ingredients, 3 recipes with 5 ingredients, 5 recipes with 6 ingredients, while all the other recipes have more than 6 ingredients with the largest one accounting for 19 ingredients.

Results: A tool that uses a linear programming approach to resolve the problem of identifying amounts of ingredients from information of selected nutrients for the recipes was implemented. This tool returned promising results with 43 out of 45 recipes correctly predicted with an average error of less than 1.5 g per ingredient. An example of successful prediction is illustrated in Figure 4.5. From the figure it is evident that the predicted ingredient amounts almost perfectly match the targeted grams with an average error of 0.37g using the selected nutrients.

code recipe	recipe name	average error	maximal error
192	Fruit mixed dried	0.366666667	0.366666667

ingredient	target_gram	predicted_gram	
Apples dried	100	100.3666667	
Prunes	100	100.3666667	
Pear dried	100	99.63333333	
Apricots dried	100	99.63333333	

nutrient code	target_value	calculated_value	predicted_value
CHO	234	234	233.9486667
ENERCJ	4581.1	4581.1	4580.158033
FASAT	0.096	0.098	0.098359333
FAT	0.5	0.5	0.501833333
FIBT	49.7	49.7	49.70623333
NA	55	55	54.945
POLYL	0	0	0
PROT	11	11	10.989
STARCH	31	31	30.989
SUGAR	203	203	202.9596667

Figure 4.5: Results for recipe “Fruit mixed dried”.

D4.5 Mapping tool for existing food intake and consumer behaviour data

One issue identified during UC3 was observed for the data which provided incorrect target nutrient values for the recipe. This was due to the inclusion of some analytical data in the published NEVO dataset instead of values calculated by recipe calculation. An example of this issue is presented in Figure 4.6. The target value represents the nutrient values obtained from the NEVO dataset for the selected recipe, while the calculated value represents the values that were calculated from the nutrients of all ingredients of the recipe. The calculation is done by summing the amounts of individual nutrients over all the ingredients. For each ingredient the nutrient value (per 1 gram) from the NEVO dataset is multiplied by the target gram from the recipe. For this reason, in the experiments, only calculated nutrient values were used. In future, a corrected dataset consisting solely of values created by recipe calculation will be created.

code recipe	recipe name	
284	Yoghurt low fat w fruit	

ingredient	target_gram	
Yoghurt low fat	85	
Sugar granulated	9	
Fruit in own juice tinned	6	

nutrient code	target_value	calculated_value
CHO	14.207	13.0195
ENERCJ	310.68	297.95489
FASAT	0.026	0.123
FAT	0.04	0.18994
FIBT	0.06	0.099
NA	48	36.336
POLYL	0	0
PROT	3.433	3.52715
STARCH	1.067	0
SUGAR	13.14	13.0195

Figure 4.6: Differences for recipe “Yoghurt low fat w fruit” in target and calculated nutrient values

The only recipes for which the tool was not able to correctly predict ingredients’ amounts were the ones that contained ingredients which had the same values for all selected nutrients. In such cases it is impossible to identify the correct amounts (for such ingredients the average value of sum of targeted amounts is set). An example of such a recipe is shown in Figure 4.7. The only solution would be the usage of nutrients which would help differentiate ingredients according to nutrient values.

D4.5 Mapping tool for existing food intake and consumer behaviour data

code recipe	recipe name	average error	maximal error
470	Chinese noodle dish Bami goreng without egg	18.66717288	118.1032627

ingredient	target_gram	predicted_gram	
Pasta white average boiled	750	762.6293495	
Pork stewing meat raw	250	240.443468	
Onions boiled	225	106.8967373	
Leek boiled	105	106.8967373	
Ham shoulder medium fat boiled	100	106.8967373	
Beans French boiled	84	106.8967373	
Bean sprouts boiled	80	106.8967373	
Cabbage white cooked	79	106.8967373	
Cabbage Chinese boiled	60	75.70955278	
Oil soy	45	44.99269114	
Soy sauce sweet	28	31.92381079	
Celery leaves fresh	20	31.92381079	
Juice lemon fresh	6	3.6485069	
Salt not fortified with iodine	3	2.34838657	

nutrient code	target_value	calculated_value	predicted_value
CHO	482.605	248.5132	248.7617132
ENERCJ	13022.13255	8980.07615	8987.884316
FASAT	14.0928	14.09878	14.08468122
FAT	68.7758	68.76825	68.83701825
FIBT	26.057	26.065	26.091065
NA	3580.3052	3580.3055	3583.885806
POLYL	0	0	0
PROT	121.11	117.4341	117.5515341
STARCH	394.158	215.2252	215.4404252
SUGAR	88.447	33.288	33.321288

Figure 4.7: Target and predicted grams are not matched due to inseparability between ingredients' nutrient values.

Summary of results from Use Cases

Within the mapping activities, we tried to match different food items and recipes from existing datasets to corresponding FoodEx2 terms and facets. For this task two approaches were applied. The first approach used StandFood to perform matching based on the English food names. Using a test dataset, it returned an accuracy level of >70% for individual food items, and 32% for recipes. The second approach used a ML-based multi-class classification method to perform mapping based on food nutrient profiles. This method achieved 55% accuracy for individual food items. It was not tested for recipes as the required nutrient profiles were not available within the dataset. For the purposes of this activity, the matching of the ingredient names within recipes and matching nutrients profiles were performed in advance, so a perfect match was provided. A tool based on a linear programming approach was developed that was able to predict the amount of ingredients on the test examples with accuracy of above 95% with an average error of <1.5g per ingredient. All work was completed using the English food names within datasets, this may limit its future use in certain datasets which have been collected in other languages. The datasets used within this task were from Serbia thus the foods described within them were commonly

D4.5 Mapping tool for existing food intake and consumer behaviour data

consumed within this country however these foods and the way in which they were described may not be applicable to datasets from other countries. This mapping system will be further tested within the Demonstrators in WP5. The quality of the identified matches within each of these Use cases could be further evaluated in future work.

Gaps identified

During the activities several gaps were identified:

- Manually coding food items with Foodex2 codes, even by subject matter experts is very subjective, which results in coding variations between different datasets. This calls for a unified and more automatic way of coding.
- Errors made by the experts influence the ability of the advanced ML-based approaches to correctly learn the features to make a correct prediction. Therefore, efforts should be made to minimise the number of errors in existing datasets.
- Translation of names of traditional dishes and colloquial names for specific foods requires a more standardized approach due to the variation in names for similar food items across countries in Europe
- In order to facilitate the use of nutrient profiles for various AI-based activities, this data needs to be available with minimal missing values across various nutrients and acquired for as many foods as possible.
- To improve the quality of automatic processing of complex foods (e.g., recipes and branded foods) a standardized way of describing the ingredients and their amounts should be prescribed and followed.
- In order to apply the StandFood method and obtain the FoodEx2 codes for any type of data, the input data should be in a predefined standardised format. A detailed description of the format for the input data is given in Appendix A.

Future work and use in FNS-Cloud

The approaches presented in this deliverable are to be implemented as web services within WP3 of the FNS-Cloud project. Furthermore, it is planned to use the developed approaches within the Food Labelling Demonstrator in WP5.

4.2 Dietary intake data collected at a food group level

Overview of activities undertaken

The aim of the developed mapping system for FFQs is to facilitate the re-use of previously collected FFQ data by enabling different datasets to be merged together thus creating larger datasets suitable for answering new research questions. In order to create a mapping system, a list of eight existing FFQs previously used was compiled; namely, EuroGene, EPIC (UK based FFQ) ^(32; 33), Scottish Collaborative Group ⁽³⁴⁾ (<https://www.foodfrequency.org/>), Food4Me ⁽³⁵⁾, eNutri (based on Food4Me) ⁽³⁶⁾, Feel4Diabetes, the National Health and Examination survey FFQ (NHANES -American FFQ) (<https://www.cdc.gov/nchs/nhanes/index.htm>), Japanese FFQ.

The overall aim of this activity was to generate a visual map of previously used FFQs alongside guidelines of how to map across these FFQs. This activity was achieved by undertaking the following series of steps:

1. Map and compare the macro and micronutrient compositions per 100g across each food/food group in different FFQs
2. Map and compare the portion sizes across each food/food group in different FFQs
3. Map and compare the response categories for frequency of consumption for each FFQ
4. Test the effectiveness of the above mapping on individual food group intakes by completing a case study using two mapped FFQs.

Development of mapping across FFQs

An overview of each FFQ was compiled containing the list and number of food groups, the list of food items contained within each food group, the timeframe response options for frequency of consumption questions, and, the timeframe each FFQ captured (Figure 4.8). A brief description of each of these parameters is described below.

D4.5 Mapping tool for existing food intake and consumer behaviour data

	A	B	C	D	E	F
1	Food groups	Food items	Questions participants were asked	Answer options	Timeframe of FFQ	Portion sizes
2	10 food groups	130 food items (classification into food groups not available)	Average use last year?	Tick one of the following:	Over a year	No specific portion size question Specified by medium servings of household measures/ food items (e.g. an apple/ a slice of bread)
3	<i>Fruit</i>			<i>Never or less than a month</i>		
4		Apples		<i>1-3 per month</i>		
5		Pears		<i>Once a week</i>		
6		Oranges		<i>2-4 per week</i>		
7		Grapefruit		<i>5-6 per week</i>		
8		Bananas		<i>Once per day</i>		
9		Grapes		<i>2-3 per day</i>		
10		Melons		<i>4-5 per day</i>		
11		Peaches		<i>6+ per day</i>		
12		Strawberries				
13		Tinned fruit				
14		Dried fruit				
15	<i>Vegetables</i>					
16		Carrots				
17		Spinach				
18		Broccoli				
19		Sprouts				
20		Cabbage				
21		Peas				
22		Green beans				
23		Marrow				
24		Cauliflower				
25		Parsnips				
26		Leeks				
27		Onions				
28		Garlic				
29		Mushrooms				
30		Peppers				
31		Beansprouts				
32		Green salad				

Figure 4.8: Overview of food groups, food items, frequency response options and portion sizes in individual FFQs

The list and number of food groups

The total number of food groups within FFQs included ranged from 10 – 19 (Table 4.3). Many similar food groups exist across the FFQs, however variation does exist. For example, in the Japanese FFQ food groups include “Green-Yellow Vegetables”, “Light-coloured Vegetables”, “Pulses”, “Mushroom” and “Seaweed”, conversely, the majority of the other FFQs only include a single food group named “Vegetables”.

A list of food items under each food group

There was much greater variation between FFQs in the individual food items included under each food group than there was between food groups themselves. Many food items listed in one FFQ may be missing from other FFQs. Although many inter-FFQ food items are similar, collapsing these into single food groups to facilitate merging cannot be achieved as easily as the food groups. Comparison of individual food items listed between each FFQ is given in Appendix B.

Timeframe response options for frequency of consumption questions

The phrasing of questions and timeframe response options provided to participants regarding food item consumption was captured to give context to the level of consumption. The manner in which frequency of consumption was asked varied across FFQs with some collecting information on weekly consumption whilst others evaluated daily consumption.

D4.5 Mapping tool for existing food intake and consumer behaviour data**Time Frame considered**

The time frame that is considered must be taken into account when comparing FFQs as seasonal variation of dietary intake does exist. Considerable variation was observed in the timeframe captured across FFQs with n=3 asking participants to consider consumption over the past year, n=4 over the past month, and n=1 over the past 2-3 months.

Assessment of portion sizes

Portion size information can be captured in a number of ways. Allowing participants to specify their portion size through use of defined small, medium and large descriptors or use of food images can facilitate more accurate estimates during retrospective reporting ⁽³⁷⁾. The Food4Me and eNutri FFQs provided participants with 3 food images of varying sizes (representing small, medium and large portions) for each food item chosen. The remaining FFQs used household measures (e.g.,teaspoons) and average portions (e.g.,1 slice of bread).

Visualization of the mapping across FFQs

An excel database to visualize the similarities across each of the individual FFQ food groups was created. Each FFQ was presented in a single column with food group names displayed in red text and individual food items contained within these groups listed underneath. Common food groups across the FFQs were presented in the same row. Food group names which were similar but not exact matches were placed either a row above or below these. For example, the food groups “Fruit” and “Vegetables” were common throughout five of the FFQs and were added to the spreadsheet on the same row (Figure 4.8). The NHANES was the only FFQ which used the combined food group name of “Fruits/Vegetables” thus this food group was added one row above the “Fruit” food group. This process was repeated for all remaining food groups to allow for easy visualization of the common food groups across the different FFQs.

Next, each food item was added underneath their respective food group and the same process repeated to align common foods. Food items which were common within each food group across the FFQs were placed in the same row with blanks left to accommodate food items which may not have been included in all FFQs. Considerably more variation existed between each FFQ food item compared to the food group list (Figure 4.8).

D4.5 Mapping tool for existing food intake and consumer behaviour data
Table 4.3. Mapping of food group categories across FFQs

EPIC	SCG	eNutri	Food4Me	Feel4Diabetes	NHANES	Japanese
(n=10)	(n=19)	(n=11)	(n=11)	(n=16)	(n=10)	(n=14)
Fruit	Fruit	Fruit	Fruit	Fruit	Fruit/Vegetables	Fruits
Vegetables	Vegetables	Vegetables	Vegetables	Vegetables		Yellow-green vegetables
						Light-coloured vegetables
						Pulses
						Mushrooms
						Seaweed
Meat and Fish	Meats	Meat and Fish	Meat and Fish	Meat and Vegetarian dishes	Meat/Fish/Protein	Meats
				Cold cuts		
	Fish					Fish and Shellfish
	Eggs					
Cereals	Breakfast Cereals	Cereals	Cereals	Bread and other cereals	Bread and Cereal	Cereal
Bread and Savoury Biscuits	Bread	Bread and Savoury Biscuits	Bread and Savoury Biscuits			
Potatoes, Rice, Pasta	Potatoes, Rice, Pasta	Potatoes, Rice, Pasta	Potatoes, Rice, Pasta		Pasta and Rice	Potatoes
Dairy Products & Fats	Milk	Dairy Products & Fats	Dairy Products & Fats	Milk	Dairy foods	Eggs and Dairy Products
	Cream and Yoghurt			Cheese		
	Cheese			Cream		
				Meals and snacks	Mixed dishes	
Soups, sauces, spreads	Savoury foods, soups, sauces	Soups, sauces, spreads	Soups, sauces, spreads	Spreads		
	Spreads and sugar	Fats and Spreads	Fats and Spreads	Cooking fats/ oils	Condiments	Miscellaneous
				Salad dressing		

D4.5 Mapping tool for existing food intake and consumer behaviour data

Sweets and Snacks	Chocolate, Sweets, Nuts, Crisps	Sweets and Snacks	Sweets and Snacks	Sugar, honey, sweets	Snack foods	Confectioneries
	Biscuits					
	Puddings			Sweet patisseries, ice cream, puddings, chocolate	Dessert foods	
	Cakes					
				Fast Foods		
Drinks	Beverages and soft drinks	Drinks	Drinks	Beverages	Beverages	Beverages
	Alcoholic Drinks					

D4.5 Mapping tool for existing food intake and consumer behaviour

data

	B	C	D	E	F
	EPIC	SCG	eNutri	Food4me	NHANES
1					
2					
3	<i>Fruit</i>	<i>Fruit</i>	<i>Fruit</i>	<i>Fruit</i>	<i>Fruit/Vegetables</i>
4	Apples	Apples	Apples	Apples	Apples
5	Pears	Pears	Pears	Pears	Pears
6	Oranges	Citrus Fruits	Oranges, satsumas, mandarins	Oranges, satsumas, mandarins	Oranges, tangerines, tangelos
7	Grapefruit		Grapefruit	Grapefruit	Grapefruit
8	Bananas	Banana	Bananas	Bananas	Bananas
9	Grapes		Grapes	Grapes	Grapes
10	Melons		Mango, melon	Mango, melon	Melon
11	Peaches		Peaches/plums/apricots/nectarines	Peaches/plums/apricots/nectarines	Peaches, nectarines, plums
12	Strawberries		Strawberries, raspberries, cherries	Strawberries, raspberries, cherries	Strawberries
13			Kiwis	Kiwis	
14	Tinned fruit	Tinned fruit	Tinned, stewed fruit peaches	Tinned, stewed fruit peaches	
15	Dried fruit		Dried fruit eg raisins, prunes	Dried fruit eg raisins, prunes	Dried fruit
16					Pineapple
17		Fresh fruit salad			
18					
19					
20	<i>Vegetables</i>	<i>Vegetables</i>	<i>Vegetables</i>	<i>Vegetables</i>	
21					
22	Carrots	Carrots	Carrots	Carrots	Carrots
23					
24	Tomatoes		Tomatoes	Tomatoes	Tomatoes
25		Other root vegetables			Sweet potatoes/ yams
26					Summer squash
27			Butternut squash, pumpkin	Butternut squash, pumpkin	
28	Parsnips		Parsnips, turnips, swedes, celeriac, fennel	Parsnips, turnips, swedes, celeriac, fennel	
29	Beetroot		Beetroot	Beetroot	
30	Cauliflower		Cauliflower	Cauliflower	Cauliflower or brussel sprouts
31			Brussels sprouts	Brussels sprouts	

Figure 4.8: Visualization of common foods within each food group across different FFQs

Comparison Exercise

In order to test the effectiveness of the developed FFQ mapping system, a comparison exercise was conducted using 2 example FFQs, namely, Food4Me and NHANES. Baseline Food4Me data from n= 210 participants (n= 30 from each country: Germany, Spain, Ireland, Greece, England, Netherlands, Poland) was re-entered into the NHANES FFQ. Where possible, food groups and food items that directly aligned between the FFQs were entered into the same food groups. Food groups that didn't directly align were either mapped to an alternative food group or merged with other food items to another food group. In some cases, where no appropriate food group was available, the group was dropped and not re-entered into the NHANES FFQ. An overview of this mapping is given in Figure 4.9.



D4.5 Mapping tool for existing food intake and consumer behaviour
data

	A	B	C	D	E
1	Food4me	NHANES	Guidelines		
2		<i>Fruit/Vegetables</i>			
3	<i>Fruit</i>				
4	Apples	Apples			
5	Pears	Pears			
6	Oranges, satsumas, mandarins	Oranges, tangerines, tangelos			
7	Grapefruit	Grapefruit			
8	Bananas	Bananas			
9	Grapes	Grapes			
10	Mango, melon	Melon			
11	Peaches/plums/apricots/nectarines	Peaches, nectarines, plums			
12	Strawberries, raspberries, cherries	Strawberries			
13	Kiwis		not in NHANES drop		
14	Tinned, stewed fruit peaches		not in NHANES drop		
15	Dried fruit eg raisins, prunes	Dried fruit			
16		Pineapple	not in Food4me drop		
17	<i>Vegetables</i>				
18	Carrots	Carrots			
19	Tomatoes	Tomatoes			
20		Sweet potatoes/ yams	not in Food4me drop		
21	Butternut squash, pumpkin	Summer squash			
22	Parsnips, turnips, swedes, celeriac, fen	Mixed vegetables	map with NHANES mixed veg, calculate average from Food4me		
23	Beetroot		map with NHANES mixed veg, calculate average from Food4me		
24	Olives		map with NHANES mixed veg, calculate average from Food4me		
25	Mushrooms		map with NHANES mixed veg, calculate average from Food4me		
26	Cauliflower	Cauliflower or brussel sprouts	calculate average from Food4me		
27	Brussels sprouts		calculate average from Food4me		

Figure 4.9: Overview of mapping across NHANES and Food4Me FFQs
Results

Analysis is being conducted to test the effect of the mapping system that was developed on the subsequent food group and nutrient intakes of the population. To examine differences in food group intakes between the two FFQs, the food items in the Food4Me FFQ and NHANES FFQ were aggregated into 31 broad food groups ⁽³⁵⁾. An example of the comparison of mean daily intakes for Irish participants, estimated from the two FFQs for the 31 food groups are presented in Table 4.4. Statistical analyses were performed using SPSS version 26.0 (IBM Corp, Armonk, NY, USA). Mean food group intakes and standard deviations were calculated for participants for the original Food4Me FFQ entries and re-entered NHANES FFQ. Differences in mean total intakes (g/day) were assessed using paired 2-sample *t* tests. Half of the sample were male (50%) and the mean age was 41.8 years (standard deviation 14.01).

D4.5 Mapping tool for existing food intake and consumer behaviour

data

Table 4.4. Mean food group intakes for Irish participants (N= 30)

Food Groups	NHANES mean intakes		Food4Me mean intakes		p-value*
	Mean (g/day)	SD (g/day)	Mean (g/day)	SD (g/day)	
Breakfast cereals and porridge	96.10	93.41	84.08	85.52	0.211
Bread, rolls, crackers and wraps	113.65	73.71	128.91	95.75	0.056
Potatoes and potato products	96.07	56.68	95.61	61.06	0.889
Rice, pasta, grains and starches	112.06	67.80	83.66	60.78	0.002
Savory dishes	52.41	50.74	42.90	36.88	0.009
Red meat	40.21	42.26	40.57	40.54	0.849
Bacon	3.03	3.63	2.28	2.58	0.002
Meat products	30.66	27.33	31.65	24.73	0.692
Poultry	42.60	29.81	36.82	27.49	<0.001
Fish, fish dishes and fish products	48.43	42.26	62.70	44.80	<0.001
Milk	291.65	309.41	297.67	274.53	0.754
Ice cream, creams and desserts	24.05	22.32	19.85	20.09	<0.001
Yoghurt	47.32	48.92	50.44	55.31	0.200
Cheese	26.93	27.92	26.33	26.17	0.532
Eggs and egg dishes	56.95	73.78	59.66	82.97	0.327
Fats and oils (e.g., butter, low fat spreads, hard cooking fats)	21.11	15.17	25.60	15.93	0.002
Biscuits	68.02	102.04	58.80	101.54	0.037
Cakes, pastries and buns	30.44	36.87	27.05	35.66	<0.001
Confectionary and savory snacks	35.38	83.08	32.89	72.63	0.325
Nuts and seeds, herbs and spices	4.78	8.22	4.48	8.18	0.079
Soups, sauces and miscellaneous foods	63.41	56.10	94.01	70.37	<0.001
Sugars, syrups, preserves and sweeteners	8.33	9.92	12.88	20.72	0.066
Teas and coffees	600.74	330.56	607.29	356.07	0.432
Other beverages (e.g., fruit juices, carbonated beverages, squash)	183.57	248.57	199.41	247.81	0.318
Alcoholic beverages	237.50	302.35	201.00	253.52	0.002
Fruit	315.19	220.44	300.10	213.56	0.002
Carrots	31.22	24.31	27.96	25.16	<0.001
Other vegetables	42.52	23.97	67.72	44.88	<0.001
Green vegetables	118.35	100.38	66.80	82.36	<0.001
Salad vegetables	8.22	12.07	76.42	66.77	<0.001
Peas, beans and lentils	27.23	55.03	27.03	65.28	0.930

P-values derived using 2-samples paired t test. P <0.05 is considered statistically significant. All significant values are indicated in bold text

D4.5 Mapping tool for existing food intake and consumer behaviour

data

Food Groups	NHANES mean intakes		Food4Me mean intakes		p-value*
	Median (g/day)	IQR (g/day)	Median (g/day)	IQR (g/day)	
Breakfast cereals and porridge	49.79	123.00	46.32	96.11	0.211
Bread, rolls, crackers and wraps	89.69	76.39	111.75	90.64	0.056
Potatoes and potato products	95.82	87.71	102.11	90.95	0.889
Rice, pasta, grains and starches	116.06	79.91	78.54	93.79	0.002
Savory dishes	41.94	79.32	41.04	58.22	0.009
Red meat	35.25	50.58	33.57	50.07	0.849
Bacon	1.64	6.86	1.64	4.57	0.002
Meat products	24.87	34.19	28.16	31.80	0.692
Poultry	50.5	53.64	43.29	44.36	<0.001
Fish, fish dishes and fish products	38.29	50.84	51.96	48.96	<0.001
Milk	189.00	263.81	194.00	253.89	0.754
Ice cream, creams and desserts	19.48	23.56	14.04	15.87	<0.001
Yoghurt	26.79	100.97	31.25	99.27	0.200
Cheese	17.29	29.81	18.71	30.50	0.532
Eggs and egg dishes	40.87	53.19	42.79	47.07	0.327
Fats and oils (e.g., butter, low fat spreads, hard cooking fats)	18.71	14.71	23.16	20.29	0.002
Biscuits	37.55	50.74	30.99	35.62	0.037
Cakes, pastries and buns	19.34	36.60	15.59	32.84	<0.001
Confectionary and savory snacks	16.84	20.61	17.61	22.38	0.325
Nuts and seeds, herbs and spices	1.71	4.14	1.14	3.21	0.079
Soups, sauces and miscellaneous foods	43.79	62.17	71.43	76.95	<0.001
Sugars, syrups, preserves and sweeteners	5.75	12.26	7.87	13.69	0.066
Teas and coffees	503.00	409.69	494.25	414.00	0.432
Other beverages (e.g., fruit juices, carbonated beverages, squash)	103.77	173.09	134.79	170.16	0.318
Alcoholic beverages	143.39	260.40	140.00	174.47	0.002
Fruit	256.93	224.50	271.57	245.23	0.002
Carrots	31.50	26.28	27.00	29.57	<0.001
Other vegetables	43.14	38.62	64.64	46.09	<0.001
Green vegetables	98.70	87.80	44.89	66.52	<0.001
Salad vegetables	1.79	12.75	61.11	89.52	<0.001
Peas, beans and lentils	16.73	26.30	13.43	25.86	0.930

D4.5 Mapping tool for existing food intake and consumer behaviour**data**

Fifteen of the food groups showed statistically significant differences across the FFQs. Significant differences were observed between the fruit and vegetable groups (n=5), snack/treat food groups (Ice cream, creams and desserts group; cakes, pastries and buns group), the rice, pasta, grains and starches group, savoury dishes group, certain meat groups (Bacon; poultry; fish, fish dishes and fish products groups), fats and oils group, soups, sauces and miscellaneous foods group and the alcoholic beverages group. These differences can be attributed to a number of factors:

- Differences between frequency of consumption response categories between FFQs
- Variations between foods listed within specific food groups
- Food items which were listed in 1 FFQ which were not captured within the other

Food groups from Food4Me and NHANES were aggregated into broad food groups as described in Forster et al., ⁽³⁵⁾. Certain food groups showed significant differences in the mean amounts consumed between the Food4Me and re-entered NHANES data despite the food groups within both FFQs being highly similar (e.g., Breakfast cereals and porridge group). This may be attributed to our mapping of the frequency of consumption categories. Due to the use of different frequency of consumption categories between both FFQs, new harmonised categories were developed which aligned the differing categories (Table 4.5). This may have affected the overall frequency of consumption reported and subsequently lead to differences in the mean food group intakes reported.

Consumption of certain food items which were captured within the Food4Me FFQ had to be excluded in the re-entry process where there was no similar food or food category available in the NHANES FFQ (e.g., Kiwis, Tinned/stewed fruit peaches). Similarly, there were some foods which were included in the NHANES FFQ where there was no corresponding or similar food to map it to within the Food4Me FFQ. Variations between foods across specific food groups was particularly evident for snack, dessert and treat foods where popular snacks and desserts differed between the populations these FFQs were designed for (Europe vs. America). These factors may have contributed to the differences observed in mean food group consumption between the two FFQs.

data
Table 4.5 Harmonised frequency of consumption categories

Food4Me	NHANES	Mapped
Never or less than a month	Never	Never
1-3 times a month	1 time per month or less	1-3 times/ month
Once a week	2-3 times per month	1-2 times/ week
2-4 times a week	1-2 times per week	3-4 times/ week
5-6 times per week	3-4 times per week	5-6 times/ week
Once a day	5-6 times per week	Once daily
2-3 times per day	1 time per day	2-3 times/ day
5-6 times per day	2-3 times per day	4-5 times/ day
>6 times per day	4-5 times per day	6+ times daily
	6 or more times per day	

Conversely, the remaining 16 food groups showed no significant differences in mean consumption between the Food4Me and NHANES FFQ. Groups showing the highest similarities in mean consumption include the potatoes and potato products; red meat; meat products; milk; yoghurt; cheese; and peas, beans and lentils groups. The food items within each of these categories were highly similar across both FFQs, in particular for the dairy food groups.

Discussion of selected FFQs

The Food4Me and NHANES FFQs were selected as they represented FFQs which were developed and used within different geographical locations of the world and information from either had not been used in the development of the other FFQ. The selected FFQs likely represented a more challenging mapping process than mapping across some of the other FFQs which contain more similarities due to their use to inform the development of another FFQ or which may use the same or similar composition tables. However, these particular FFQs were selected to really test the feasibility of the developed mapping system.

Future work and use in FNS-Cloud

Further work is ongoing to compare the mean nutrient intakes between the different combinations of the FFQs. The full analysis is summarized in Figure 4.10.

D4.5 Mapping tool for existing food intake and consumer behaviour

data

The FFQ mapping system described within this deliverable will be considered for use within DEM03 of the WP5 demonstrators. The development of the mapping system across FFQs and results of the comparison exercise will be included in an academic article which will be submitted for peer-review publication.

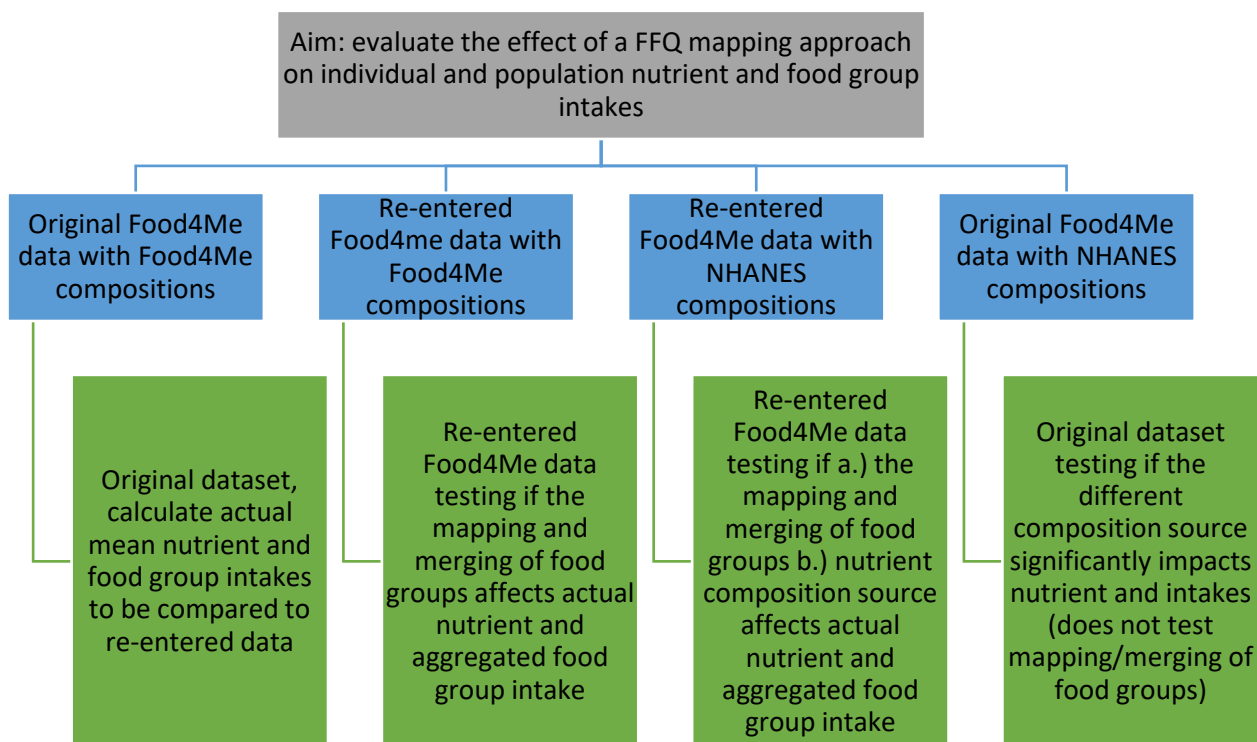


Figure 4.10. Overview of full analyses to assess the feasibility of developed FFQ mapping system

data

5. References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018.
2. FNS-Cloud FNS (2019) What we are doing. <https://www.fns-cloud.eu/>
3. European Commission (2016) Horizon 2020 Commission expert group on Turning FAIR data into reality. <https://ec.europa.eu/research/openscience/index.cfm>
4. European Commission European Open Science Cloud (EOSC). https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en (accessed September 13th 2021)
5. Welch A (2013) *Dietary intake measurement: Methodology*. In *Encyclopedia of Human Nutrition*. Third ed. Waltham: Academic Press
6. Thompson F, Subar A (2013) Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*, 2nd ed.
7. Finglas PM, Berry R, Astley S (2014) Assessing and improving the quality of food composition databases for nutrition and health applications in Europe: the contribution of EuroFIR. *Adv Nutr* **5**, 608s-614s.
8. Zeb A, Soininen J-P, Sozer N (2021) Data harmonisation as a key to enable digitalisation of the food sector: A review. *Food and Bioprocess Technology* **127**, 360-370.
9. Pennington JAT, Butrum RR (1991) Food descriptions using taxonomy and the 'LanguaL' system. *Trends in Food Science & Technology* **2**, 285-288.
10. Ireland JD, Møller A (2000) Review of International Food Classification and Description. *Journal of Food Composition and Analysis* **13**, 529-538.
11. Polacchi W (1986) Standardized food terminology: an essential element for preparing and using food consumption data on an international basis. *Food and nutrition bulletin* **8**, 66 - 68.
12. Kroeber-Riel W, Weinberg P, Gröppel-Klein A (2008) *Konsumentenverhalten*. 9th edition ed: Vahlen.
13. Markovina J, Stewart-Knox BJ, Rankin A *et al.* (2015) Food4Me study: Validity and reliability of Food Choice Questionnaire in 9 European countries. *Food Quality and Preference* **45**, 26-32.
14. Stok FM, Hoffmann S, Volkert D *et al.* (2017) The DONE framework: Creation, evaluation, and updating of an interdisciplinary, dynamic framework 2.0 of determinants of nutrition and eating. *PLoS One* **12**, e0171077.
15. Dekkers ALM, Verkaik-Kloosterman J, van Rossum CTM *et al.* (2014) SPADE, a New Statistical Program to Estimate Habitual Dietary Intake from Multiple Food Sources and Dietary Supplements. *The Journal of Nutrition* **144**, 2083-2091.
16. McMurdo G (1982) Database file normalization as an information science related activity. *Journal of Information Science* **4**, 9-17.
17. European Food Safety A (2015) The food classification and description system FoodEx 2 (revision 2). *EFSA Supporting Publications* **12**, 804E.
18. EFSA (2011) Report on the development of a Food Classification and Description System for exposure assessment and guidance on its implementation and use. *EFSA Journal* **9**.
19. Hendricks TC (1992) LanguaL. An automated method for describing, capturing and retrieving data about food. *World Rev Nutr Diet* **68**, 94-103.
20. Ireland JD, Møller A (2010) LanguaL Food Description: a Learning Process. *European Journal of Clinical Nutrition* **64**, S44-S48.

D4.5 Mapping tool for existing food intake and consumer behaviour
data

21. Funk B, Sadeh-Sharvit S, Fitzsimmons-Craft EE *et al.* (2020) A Framework for Applying Natural Language Processing in Digital Health Interventions. *J Med Internet Res* **22**, e13855.
22. Lacson R, Long W (2006) Natural language processing of spoken diet records (SDRs). *AMIA Annual Symposium proceedings AMIA Symposium* **2006**, 454-458.
23. van Erp M, Reynolds C, Maynard D *et al.* (2021) Using Natural Language Processing and Artificial Intelligence to Explore the Nutrition and Sustainability of Recipes and Food. *Frontiers in artificial intelligence* **3**, 621577-621577.
24. Guasch-Ferré M, Hernández-Alonso P, Drouin-Chartier JP *et al.* (2021) Walnut Consumption, Plasma Metabolomics, and Risk of Type 2 Diabetes and Cardiovascular Disease. *J Nutr* **151**, 303-311.
25. DeGregory KW, Kuiper P, DeSilvio T *et al.* (2018) A review of machine learning in obesity. *Obesity Reviews* **19**, 668-685.
26. Eftimov T, Korošec P, Koroušić Seljak B (2017) StandFood: Standardization of Foods Using a Semi-Automatic System for Classifying and Describing Foods According to FoodEx2. *Nutrients* **9**.
27. Popovski G, Seljak B, Eftimov T (2019) FoodOntoMap: Linking Food Concepts across Different Food Ontologies. *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KEOD*, 195-202.
28. Stojanov R, Popovski G, Jofce N *et al.* (2020) FoodViz: Visualization of Food Entities Linked Across Different Standards. *Machine Learning, Optimization, and Data Science*, 28-38.
29. Dooley DM, Griffiths EJ, Gosal GS *et al.* (2018) FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food* **2**, 23.
30. Pizzuti T, Mirabelli G, Grasso G *et al.* (2017) MESCO (MEat Supply Chain Ontology): An ontology for supporting traceability in the meat supply chain. *Food Control* **72**, 123-133.
31. Popovski G, Seljak BK, Eftimov T (2019) FoodBase corpus: a new resource of annotated food entities. *Database* **2019**, baz121.
32. Day N, Oakes S, Luben R *et al.* (1999) EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1**, 95-103.
33. Bingham SA, Welch AA, McTaggart A *et al.* (2001) Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* **4**, 847-858.
34. Hollis JL, Craig LCA, Whybrow S *et al.* (2017) Assessing the relative validity of the Scottish Collaborative Group FFQ for measuring dietary intake in adults. *Public Health Nutrition* **20**, 449-455.
35. Forster H, Fallaize R, Gallagher C *et al.* (2014) Online dietary intake estimation: the Food4Me food frequency questionnaire. *J Med Internet Res* **16**, e150.
36. Fallaize R, Franco RZ, Hwang F *et al.* (2019) Evaluation of the eNutri automated personalised nutrition advice by users and nutrition professionals in the UK. *PLoS One* **14**, e0214931.
37. Cade JE, Burley VJ, Warm DL *et al.* (2004) Food-frequency questionnaires: a review of their design, validation and utilisation. *Nutrition Research Reviews* **17**, 5-22.

data

Appendix A - Input data format

In order to apply the StandFood method and obtain the FoodEx2 codes for any type of data, the input data should be in a predefined standardised format.

1. Food composition data

Food composition data (FCD) already comes in a well-structured format to begin with. For applying the StandFood method on FCD the names of the foods need to be available in English. In order to obtain the best possible results, the English food names need to:

- be as descriptive as possible
- contain nouns where possible instead of verbs or adjectives
- be consistent in the use of terminology
 - Example: usage of “&” and “and” – “Iceberger/M and M Cookie Ice-Cream Sandwich” vs. “Half & Half sliced pan (Irish) (unfortified)”
- minimise or avoid the usage of adverbs, prepositions, conjunctions and determiners
avoid negations
 - Example: no oil, no fat, no salt, no sugar, no added sugar....
- avoid quantifiers related to food portions or portions per package
 - Example: Sushi (pack of 3)
- avoid using traditional names of foods
 - Example: roasted red pepper relish (ajvar)
 - § if unavoidable include them in brackets after the description in English
- avoid adding unnecessary descriptions
 - Example: “Ribena, all flavours, ready to drink”, “Curry Sauce (Take Away)”
- avoid using abbreviations
 - Example: “Chicken Soup (No Veg)”, “Chocolate Sauce (Choc/Milk/Butter)”, “Turkey and Veg Soup (Carr,Celery,Onion,Leek)”

2. Recipe data

2.1. Recipe name/title/description

D4.5 Mapping tool for existing food intake and consumer behaviour

data

When mapping the English name of a recipe to the FoodEx2 dataset with StandFood, the same rules as for FCD apply, as the mapping is one to one.

2.2. Recipe list of ingredients

If the recipe data is available with the list of ingredients, there should be a uniform standard way of the format in which the list of ingredients comes:

- use a distinct and consistent delimiter for the separate ingredients
 - Example: use “;” as a delimiter as “,” might appear in the description/name of one ingredient.
- use a consistent placement and format of percentages
- use consistent placement and format of brackets
- use consistent format of subgroup ingredients

The proposed format for recipe list of ingredients is given in Table 1 and additional information is given below the same table.

Table 1. Format of list of ingredients

Ingredients
$\text{ingredient}_{1, a\%} (\text{ingredient}_{11, b\%} (\text{ingredient}_{111, c\%} (\text{ingredeint}_{1111, d\%}; \dots; \text{ingredient}_{111q, e\%}); \dots; \text{ingredient}_{11p, f\%}); \dots; \text{ingredient}_{1m, g\%}); \dots; \text{ingredient}_{n, h\%} (\text{ingredient}_{n1, i\%}; \dots; \text{ingredient}_{nr, j\%})$

Where, n is the number of ingredients in the food, i is the number of ingredients in ingredient $_{1,}$ is the number of ingredients in ingredient $_{11,}$ which is the first ingredient in ingredient $_{1,}$ is the number of ingredients in ingredient $_{111,}$ which is the first ingredient in ingredient $_{11,}$ is the number of ingredients in ingredient $_{1111,}$ and $a, b, c, d, e, f, g, h, i, j$ are the percentages of the ingredient before.

Brackets are used to define hierarchy in ingredients. Use only the basic brackets “(“ and “)“.

Brackets shouldn't be used for any other purpose in the ingredient list.

D4.5 Mapping tool for existing food intake and consumer behaviour**data****3. Branded food data**

For branded foods the same applies as for recipe data. For the name of the food, the name in English is required, the mapping is one to one, and following the rules listed for FCD data is advised, and the usage of the format in Table 1 is required for the ingredient list.

4. Dietary intake data

For dietary intake data, the names of the foods are needed, the mapping is one to one, therefore, the same rules listed for FCD data apply.

D4.5 Mapping tool for existing food intake and consumer behaviour data

Appendix B

EPIC	SCG	eNutri	Food4me	NHANES	Japanese
				<i>Fruit/Vegetables</i>	<i>Fruits</i>
<i>Fruit</i>	<i>Fruit</i>	<i>Fruit</i>	<i>Fruit</i>		
Apples	Apples	Apples	Apples	Apples	Apple
Pears	Pears	Pears	Pears	Pears	
Oranges	Oranges, satsumas or grapefruit	Oranges, satsumas, mandarins	Oranges, satsumas, mandarins	Oranges, tangerines, tangelos	Citrus fruits
Grapefruit		Grapefruit	Grapefruit	Grapefruit	
Bananas	Bananas	Bananas	Bananas	Bananas	Banana
Grapes		Grapes	Grapes	Grapes	
Melons	All other fruits (grapes, strawberries, melon)	Mango, melon	Mango, melon	Melon	
Peaches	Peaches or nectarines	Peaches/plums/apricots /nectarines	Peaches/plums/apricots /nectarines	Peaches, nectarines, plums	
Strawberries		Strawberries, raspberries, cherries	Strawberries, raspberries, cherries	Strawberries	Strawberries
	Kiwi fruit	Kiwis	Kiwis		Kiwi fruit
Tinned fruit	Tinned fruit (all types)	Tinned, stewed fruit peaches	Tinned, stewed fruit peaches		
Dried fruit	Dried fruit (e.g., raisins, dates or figs)	Dried fruit e.g., raisins, prunes	Dried fruit e.g. raisins, prunes	Dried fruit	
				Pineapple	
	Fresh fruit salad				
					Persimmon
					Other fruits
<i>Vegetables</i>	<i>Vegetables</i>	<i>Vegetables</i>	<i>Vegetables</i>		

D4.5 Mapping tool for existing food intake and consumer behaviour data

					<i>Green-yellow vegetables</i>
Carrots	Carrots	Carrots	Carrots	Carrots	Carrot
					Tomato juice, vegetable juice
Tomatoes	Tomatoes	Tomatoes	Tomatoes	Tomatoes	Tomato
				Sweet potatoes/ yams	
		Butternut squash, pumpkin	Butternut squash, pumpkin	Summer squash	Pumpkin
Parsnips		Parsnips, turnips, swedes, celeriac, fennel	Parsnips, turnips, swedes, celeriac, fennel		
Beetroot		Beetroot	Beetroot		
Cauliflower	Cauliflower, swede or turnips	Cauliflower	Cauliflower	Cauliflower or brussel sprouts	
	Brussel sprouts	Brussels sprouts	Brussels sprouts		
Leeks	Leeks or courgettes	Leeks	Leeks		
Garlic		Garlic	Garlic	Garlic	
Sprouts		Bean sprouts, radish	Bean sprouts, radish		
Onions	Onions	Onions	Onions	Onions	
				Raw or cooked greens	
					Green peppers
				Cooked greens	
Spinach	Spinach or spring greens	Spinach	Spinach		Spinach
					Greens other than spinach (komatsuna, shungiku, leaves of Japanese radish)
Broccoli	Broccoli	Broccoli, spring greens, kale	Broccoli, spring greens, kale	Broccoli	Broccoli

D4.5 Mapping tool for existing food intake and consumer behaviour data

		Asparagus, okra	Asparagus, okra		
	Other salad vegetables (lettuce, cucumber etc)				
Green salad		Green salad, lettuce, cucumber, celery, watercress	Green salad, lettuce, cucumber, celery, watercress	Lettuce salads	
Beansprouts					
Watercress					
					<i>Light-coloured vegetables</i>
Cabbage	Cabbage (all kinds)	Cabbage (red, white, savoy)	Cabbage (red, white, savoy)	Sauerkraut or cabbage	Cabbage, lettuce, cucumber
Colelsaw		Coleslaw, sauerkraut	Coleslaw, sauerkraut	Coleslaw	
					Chinese cabbage
					Eggplant
					Japanese radish (excluding grated radish)
					Grated radish
					Burdock
					Lotus root
					Japanese radish
(in soups, sauces, spreads group)	(in soups, sauces, spreads group)	(in soups, sauces, spreads group)	Pickles	Pickles	
				(in pulses group)	String beans
Peas	Peas or green beans	Fresh/frozen peas/sugar snap peas	Fresh/frozen peas/sugar snap peas	Peas	
Green beans		Green beans, broad beans, runner beans	Green beans, broad beans, runner beans	String beans	
Marrow		Marrow, courgettes, aubergine	Marrow, courgettes, aubergine		

D4.5 Mapping tool for existing food intake and consumer behaviour data

Beans		Baked beans	Baked beans		
Lentils		Dried lentils, beans, peas, chickpeas	Dried lentils, beans, peas, chickpeas		
	Olives	Olives	Olives		
Peppers	Sweet peppers	Sweet peppers	Sweet peppers	Sweet peppers	
Sweetcorn	Sweetcorn	Corn (on the cob, sweetcorn)	Corn (on the cob, sweetcorn)	Corn	
Avocado		Avocado	Avocado		
					<i>Pulses</i>
Tofu		Tofu, soya meat, TVP, vegeburger	Tofu, soya meat, TVP, vegeburger	Boiled or chilled tofu (bean curd)	
					Boiled soybeans
					Nattou (fermented soybeans)
					Fried bean curd
					Mapo tofu (bean curd mixed with Chinese meat and chilli sauce)
					Miso (soup)
	Mixed vegetable dishes (e.g., stir-fry, curry or bake)			Mixed vegetables	
	Potato salad				
	Coleslaw, or other veg salads in dressing				
(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	French fries, home fries, hash browns, tater tots	
(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	Potato salad	

D4.5 Mapping tool for existing food intake and consumer behaviour data

(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	Baked/ boiled/ mashed potatoes	(in potatoes group)
		Sauerkraut stew	Sauerkraut stew		
		Stuffed cabbage	Stuffed cabbage		
					<i>Mushroom</i>
Mushrooms		Mushrooms	Mushrooms		Mushrooms (shiitake, shimeji, enokitake)
	Tinned vegetables (all kinds)			<i>Seaweed</i>	
					Hijiki, wakame, kombu (excluding those in miso soup)
<i>Meat and Fish</i>	<i>Meats</i>	<i>Meat and Fish</i>	<i>Meat and Fish</i>	<i>Meats, fish and protein</i>	<i>Meats</i>
Beef	Beef (roast, grilled, casseroled or fried)	Beef, venison (roast, steak, mince)	Beef, venison (roast, steak, mince)		
				Steak	Beef steak, roast steak
				Roast beef/steak in sandwiches	
				Beef stew/ pot pie & veg	Other beef dishes (sukiyaki, vegetables fried with beef, potatoes and beef)
	Mince or meat sauce (e.g., bolognese)		Ground beef in mixtures		
				Other minced meat dishes (hamburg steak, meatball, minced meat cutlet)	
Burger	Burgers (beef, lamb, chicken or turkey)	Burgers e.g., beef, meatballs	Burgers e.g., beef, meatballs	Roast beef/ pot roast	

D4.5 Mapping tool for existing food intake and consumer behaviour data

				Beef hamburgers/ cheeseburgers	
				Pork/beef spareribs	
Sausages	Sausages (pork, beef or frankfurters)	Sausages e.g., pork, jadvurst	Sausages e.g., pork, jadvurst	Sausage	
				Hot dogs	
Pork	Pork or lamb (roast, grilled, casseroled or fried)	Pork (roast, chops)	Pork (roast, chops)	Pork	Roast pork
					Breaded pork cutlet, pork fillet steak, pork cutlet on a skewer
					Sauteed pork
					Other pork dishes (vegetables fried pork, sweet and sour pork)
					Gyoza (jiaozi)
Ham				Baked ham/ ham steak	
Bacon	Bacon or gammon	Bacon	Bacon	Bacon	Ham, sausage, bacon
Lamb		Lamb, goat (roast, chops, souvlaki)	Lamb, goat (roast, chops, souvlaki)		
Chicken	Chicken or turkey (roast, grilled, casseroled or fried)	Chicken or turkey, grilled, roasted	Chicken or turkey, grilled, roasted	Roast turkey	
				Ground chicken/ turkey	
				Turkey/ chicken/ ham/ other cold cuts	
				Chicken including in salads/mixtures	Other chicken dishes (roast chicken,

D4.5 Mapping tool for existing food intake and consumer behaviour data

					chicken boiled with soy sauce)
		Processed chicken or turkey e.g., nuggets, goujons, kiev	Processed chicken or turkey e.g., nuggets, goujons, kiev	Fried chicken/ turkey	Deep-fried chicken
Corned beef	Cold meats (e.g., ham, corned beef, chicken roll)	Cured meats e.g., corned beef, salami, chorizo	Cured meats e.g., corned beef, salami, chorizo		
	Salami or continental sausage				
Savoury pies	Meat or chicken pies, pasties or sausage roll	Savoury pies, meat pie, pasties, sausage rolls	Savoury pies, meat pie, pasties, sausage rolls		
	Haggis or black pudding				
		Sliced cold meats e.g., ham, turkey	Sliced cold meats e.g., ham, turkey		
		Offal e.g., liver, kidney	Offal e.g., liver, kidney		
Liver	Liver, liver sausage or liver pate		Liver/ liverwurst	Liver	
		Pate e.g., meat or liver	Pate e.g., meat or liver		
		Stew and casserole (meat and veg)	Stew and casserole (meat and veg)		
	<i>Fish</i>				<i>Fish and Shellfish</i>
White fish	White fish (e.g., haddock, cod, plaice, scampi) fried or cooked in batter	White fish e.g., cod, haddock, sole	White fish e.g., cod, haddock, sole		
	Grilled, poached or baked white fish				
	Smoked white fish				

D4.5 Mapping tool for existing food intake and consumer behaviour data

Oily fish	Grilled, poached or baked oily fish				
					Salmon, trout
					Horse mackerel (including dried)
					Pacific saury, mackerel, sardine, yellowtail
	Fried oily fish (e.g., salmon, herring, fresh tuna or mackerel)	Non-smoked oily fish, fresh (mackerel, tuna, sardines, salmon)	Non-smoked oily fish, fresh (mackerel, tuna, sardines, salmon)	Tuna	Canned tuna
	Tinned salmon	Non-smoked oily fish, canned (mackerel, tuna, sardines, salmon)	Non-smoked oily fish, canned (mackerel, tuna, sardines, salmon)		
	Tinned tuna				
	Smoked oily fish (kipper, mackerel, salmon)	Smoked fish e.g., salmon, mackerel	Smoked fish e.g., salmon, mackerel	Smoked fish/ seafood	
	Sardines, pilchards or rollmop herrings		Fish, not fried, smoked, raw		
				Fried fish/ fish sticks	
Shellfish	Prawns, crab etc	Shellfish, crab, prawn	Shellfish, crab, prawn	Other shellfish (asari, shijimi, clam, scallop)	
					Cuttlefish, prawn/shrimp, fried prawn
	Mussels, oysters, cockles, scallops		Raw oysters/ clams	Oyster, fried oyster (in season)	
Roe		Fish roe, caviar, taramasalata	Fish roe, caviar, taramasalata	Cod roe, herring roe	

D4.5 Mapping tool for existing food intake and consumer behaviour data

		Sushi	Sushi	Sushi	(in cereals group)
Fried fish					
					Fish with red flesh (tuna, marlin/swordfish, bonito)
					Fish with white flesh (cod, flatfish)
					Eel
					Small fish eaten bone and all (dried sardine, smelt, dried young sardines)
		Fried fish in batter	Fried fish in batter		
Fish fingers	Fish fingers	Fish fingers, fish cakes	Fish fingers, fish cakes		
	Fish cakes, fish pie	Fish dishes e.g., pie, pudding, casserole	Fish dishes e.g., pie, pudding, casserole		
					Fish paste products (kamaboko, chikuwa, hanpen)
(in vegetable group)	(in vegetable group)	(in vegetable group)	(in vegetable group)	Tofu/ soya burgers/ soya meat substitute	(in pulses group)
				Dried cooked beans	
				Eggs, included in salad	
<i>Cereals</i>	<i>Breakfast Cereals</i>	<i>Cereals</i>	<i>Cereals</i>	<i>Breads and cereal</i>	<i>Cereals</i>
Porridge	Porridge or Ready Brek	Porridge, readybrek	Porridge, readybrek	Oatmeal or hot breakfast cereal	
Cereal	Bran flakes, Sultana bran, All bran etc	Breakfast cereals, wholegrain e.g., branflakes, barley flakes	Breakfast cereals, wholegrain e.g., branflakes, barley flakes	Cold cereal	

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Shredded wheat, weetabix				
	Cornflakes, Rice Krispies, Special K etc	Breakfast cereals, non-wholegrain e.g., cornflakes	Breakfast cereals, non-wholegrain e.g., cornflakes		
	Muesli (all types)	Breakfast cereals e.g., muesli, cruesli	Breakfast cereals e.g., muesli, cruesli		
	Coco pops, frosties, sugar puffs, crunchy nut, nut cornflakes				
<i>Bread and Savoury Biscuits</i>	<i>Bread</i>	<i>Bread and Savoury Biscuits</i>	<i>Bread and Savoury Biscuits</i>		
	Bread roll or bun			Bread/ rolls	Breads
White bread	Bread (including toast and sandwiches)	White bread	White bread		
		White rolls	White rolls		
Brown bread		Brown bread and seeded bread	Brown bread and seeded bread		
		Brown and seeded rolls	Brown and seeded rolls		
Wholemeal bread		Dark wholemeal breads e.g., rye or soda	Dark wholemeal breads e.g., rye or soda		
				Pancakes/ waffles/ french toast	
				Bagels/ English Muffin	
				Corn bread/ muffins	
	Other breads (pitta, naan, soft tortillas)	Tortilla, wraps	Tortilla, wraps	Tortillas/ tacos, corn or flour	
Crackers		Cream crackers, cheese biscuits, rusks	Cream crackers, cheese biscuits, rusks		
Crispbread		Crispbread e.g., ryvitta	Crispbread e.g., ryvitta		

D4.5 Mapping tool for existing food intake and consumer behaviour data

(in sweets and snacks group)	(in chocolates, sweets, nuts and crisps group)	(in sweets and snacks group)	(in sweets and snacks group)	Biscuits	
				Stuffing/ dumplings	
	Crossaints, butteries or garlic bread				
<i>Potatoes, Rice & Pasta</i>	<i>Potatoes, Rice & Pasta</i>	<i>Potatoes, Rice & Pasta</i>	<i>Potatoes, Rice & Pasta</i>	<i>Pasta and rice etc</i>	<i>Potatoes</i>
					Potatoes (white potato, taro, sweet potato)
	Mashed potatoes	Potaotes - mashed, instant, roast	Potaotes - mashed, instant, roast	Mashed potatoes	
Boiled potatoes	Boiled or baked potatoes	Potatoes - boiled, jacket	Potatoes - boiled, jacket		
Roast potatoes	Roast or fried potatoes				
					Croquettes
Potato salad					
		Potato dishes e.g., salads, dauphinoise	Potato dishes e.g., salads, dauphinoise		
		Potato or plain dumplings	Potato or plain dumplings		
Chips	Oven chips, potato waffles or croquettes	Chips	Chips		
	Home-cooked chips				
	Chips from chip shop or restaurant				
				Rice/ other cooked grains	Rice
White rice	White rice	White rice	White rice		
Brown rice	Brown rice	Brown rice, buckwheat and barley groats	Brown rice, buckwheat and barley groats		
					Rice ball

D4.5 Mapping tool for existing food intake and consumer behaviour data

					Bowl of rice, topped with chopsuey-like mixture, Japanese pilaf
					Bowl of rice, topped with pork cutlets/chicken and egg/beef
					Pilaf, Chinese fried rice
					Curry and rice, with hashed meat
	Pasta (all types or couscous)				
White pasta		White pasta, noodles and other grains e.g., cous cous, polenta	White pasta, noodles and other grains e.g., cous cous, polenta		
Wholemeal pasta		Wholemeal pasta	Wholemeal pasta		
				Macaroni & Cheese	Spaghetti/ macaroni
				Pasta/ macaroni salad	
				Other salads/ spaghetti	
Lasagne		Lasagne, mousaka, ravioli and tortellini, filled dumplings	Lasagne, mousaka, ravioli and tortellini, filled dumplings	Lasagne, stuffed shells, stuffed manicotti, ravioli, tortellini	
Pizza		Pizza, calzone	Pizza, calzone		
	Noodles (all types)				Japanese noodles (udon, soba, somen, hiyamugi)
					Chinese noodles in soup (lamain)

D4.5 Mapping tool for existing food intake and consumer behaviour data

					Chow mein
		Spring rolls	Spring rolls		
		(in meat and fish group)	(in meat and fish group)	(in meats, fish and protein group)	Sushi
<i>Dairy Products & Fats</i>	<i>Dairy Products & Fats</i>	<i>Dairy Products & Fats</i>	<i>Dairy foods</i>	<i>Eggs and dairy products</i>	
	<i>Milk</i>				
Milk				(in drinks)	Milk
	Full fat milk (fresh/dried)	Full-fat/whole milk, buttermilk	Full-fat/whole milk, buttermilk		
	Semi-skimmed milk	Low-fat or semi-skimmed milk	Low-fat or semi-skimmed milk		
	Fully skimmed milk (fresh/dried)	Zero fat or skimmed milk	Zero fat or skimmed milk		
	Condensed/ evaporated milk				
	Soy milk				
	<i>Cream and Yoghurt</i>				
	Cream (all types)				
Single cream		Single/sour cream	Single/sour cream		
Double cream		Double/clotted cream	Double/clotted cream		
				Yoghurt	Yoghurt
Full fat yogurt	Full fat yoghurt (greek)	Full-fat/greek yoghurt	Full-fat/greek yoghurt		
Low fat yogurt	Low fat yoghurt (plain, fruit)	Low-fat natural yoghurt, fromage frais	Low-fat natural yoghurt, fromage frais		
	Low calorie yoghurt				
	Fromage frais (plain or fruit)	Fruit yoghurt, fruit mousse	Fruit yoghurt, fruit mousse		
	<i>Cheese</i>				
Cheese				Cheese	Cheese

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Full hard cheese (e.g., cheddar, gruyere, wensleydale, gouda)	High fat cheeses e.g., stilton, cheddar, brie, gouda	High fat cheeses e.g., stilton, cheddar, brie, gouda		
	Medium fat cheese (e.g., edam, brie, camembert, feta, cheese spreads)	Medium fat cheeses e.g., edam, goats, camembert, feta, emmental	Medium fat cheeses e.g., edam, goats, camembert, feta, emmental		
	Low fat cheese (e.g., low fat cream cheese, low fat hard cheese)	Low fat cheeses - fresh mozzarella, cream cheese, Katiki	Low fat cheeses - fresh mozzarella, cream cheese, Katiki	Cream cheese (in condiments group)	
		Very low fat cheese e.g., cottage cheese, quark	Very low fat cheese e.g., cottage cheese, quark		
	Full fat cream cheese (Philadelphia, Boursin, Danish Blue)				
Cottage Cheese	Cottage cheese (all types)		Cottage cheese		
Dairy dessert					
	(in beverages group)	Milkshakes, fruit smoothies	Milkshakes, fruit smoothies		
				Mayonnaise (in condiments group)	Mayonnaise (in miscellaneous)
Salad cream		Salad cream, mayonnaise	Salad cream, mayonnaise	Sour cream (in condiments group)	
Low cal salad cream		Low fat salad cream, mayonnaise	Low fat salad cream, mayonnaise		
				Salad dressing (in condiments group)	Dressing (in miscellaneous)
French dressing		French dressing, vinaigrette	French dressing, vinaigrette		
Other dressing		Other salad dressing	Other salad dressing		
	<i>Eggs</i>				

D4.5 Mapping tool for existing food intake and consumer behaviour data

Eggs	Boiled or poached eggs	Egg - boiled, scrambled, omelette etc.	Egg - boiled, scrambled, omelette etc.	Egg, egg dishes (in egg and dairy products group)	
	Fried eggs				
	Scrambled eggs or omelette				
Quiche		Quiche, savoury pancakes	Quiche, savoury pancakes		
	<i>Spreads and Sugar</i>	<i>Fats and Spreads</i>	<i>Fats and Spreads</i>	<i>Condiments</i>	<i>Miscellaneous</i>
	Butter, margarine, fat spread or oil (on bread)				
Butter		Butter	Butter	Butter on breads/pancakes or potatoes/vegetables	Butter for breads
Hard margarine		Block/hard margarine e.g., stork/krona	Block/hard margarine e.g., stork/krona	Margarine on breads/pancakes or potatoes/vegetables	Margarine for breads
Polyunsaturated margarine	Polyunsaturated marg flora/sunflower/soya	Polyunsaturated marg flora/sunflower/soya			
Other margarine					
		Soft marg olive oil based, bertolli/blue band	Soft marg olive oil based, bertolli/blue band		
Low fat spread		Low fat spreads (<60%)	Low fat spreads (<60%)		
Very low fat spread					
	Fat or oil (while cooking)		Oils for cooking		
		Olive Oil	Olive Oil		
	Liquid vegetable oil	Other vegetable oils	Other vegetable oils		

D4.5 Mapping tool for existing food intake and consumer behaviour data

				Salsa	
				Gravy	
					Soy sauce
					Worcester sauce
					Sesame
					Deep-fried foods, tempura
(in soups, sauces and spreads)	Jam, honey or marmalade	(in soups, sauces and spreads)	(in soups, sauces and spreads)		
(in soups, sauces and spreads)	Peanut butter or chocolate spread	(in soups, sauces and spreads)	(in soups, sauces and spreads)		
	Table sugar				
(in soups, sauces and spreads)	Yeast or meat extract (marmite, bovril etc)	(in soups, sauces and spreads)	(in soups, sauces and spreads)		
<i>Soups, sauces and spreads</i>	<i>Savoury foods, Soups and Sauces</i>	<i>Soups, sauces and spreads</i>	<i>Soups, sauces and spreads</i>		
	Soup (homemade)				
	Soup (tins/ cartons)				
	Soup (dried/ instant)				
Vegetable soup		Non-creamy soups e.g.,minestrone, vegetable	Non-creamy soups e.g.,minestrone, vegetable		
		Creamy soups e.g.,chowder, cream of mushroom	Creamy soups e.g.,chowder, cream of mushroom		
Meat soup					
Sauces	Other sauces (white, cheese curry, sweet & sour)				

D4.5 Mapping tool for existing food intake and consumer behaviour data

		Creamy sauces e.g.,carbonara/cheese	Creamy sauces e.g.,carbonara/cheese		
	Gravy	Dark sauces e.g.,gravy, stir-fry sauces	Dark sauces e.g.,gravy, stir-fry sauces		
	Tomato-based sauces	Tomato sauces e.g.,in Bolognese, on meatballs or pasta	Tomato sauces e.g.,in Bolognese, on meatballs or pasta		
Ketchup	Bottled sauces (e.g.,ketchup)	Tomato ketchup	Tomato ketchup	Ketchup (in condiments group)	
Pickles		Pickles, chutney, satesaus (sate)	Pickles, chutney, satesaus (sate)	Ketchup (in condiments group)	
Marmite		Marmite/bovril	Marmite/bovril		
Jam		Jam/marmalade/honey	Jam/marmalade/honey		
Peanut butter		Nut or chocolate spreads e.g.,peanut butter, Nutella	Nut or chocolate spreads e.g.,peanut butter, Nutella	Jam/jelly/honey (in condiments group)	
	Pesto			Peanut butter/ other nut butter (in condiments group)	Jam, honey for breads (in miscellaneous)
	Mayonnaise or salad cream		Apple sauce		
	Oil & vinegar dressing				
	Pickled vegetables or chutney				
	Hummus				
				<i>Mixed dishes</i>	
(in soups, sauces and spreads group)	(in savoury foods, soups and sauces group)	(in soups, sauces and spreads group)	(in soups, sauces and spreads group)	Chilli	
(in potatoes, rice & pasta group)	Pizza	(in potatoes, rice & pasta group)	(in potatoes, rice & pasta group)	Soup	

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Quorn products (all types)			Pizza	(in meats group)
	Soya beans, TVP, tofu or soya meat substitute				
	Other beans (kidney, beans, chickpeas)				
	Lentils (excluding soup)				
	Nut roast, nut burgers or vegetable burgers				
(in dairy products and fats group)	Quiche or savoury flan	(in dairy products and fats group)	(in dairy products and fats group)		
	Baked beans				
<i>Sweets and Snacks</i>	<i>Chocolates, Sweets, Nuts and Crisps</i>	<i>Sweets and Snacks</i>	<i>Sweets and Snacks</i>		
Sweets	Boiled sweets, mints	Sweets, toffees, mints, liquorice	Sweets, toffees, mints, liquorice		
	Fruit gums, pastilles, jellies, chewy sweets				
Ice cream		Ice-cream, choc ices	Ice-cream, choc ices	<i>Snack foods</i>	
		Sorbets and jellies	Sorbets and jellies		
Sugar		Sugar, added to tea, coffee, cereal	Sugar, added to tea, coffee, cereal		
Crisps	Crisps	Crisps or other packet snacks e.g., wotsits	Crisps or other packet snacks e.g., wotsits		<i>Confectioneries</i>
	Reduced fat crisps				
	Other savoury snacks (quavers, tortilla chips, popcorn etc)	Tortilla/corn chips			
				Potato chips	

D4.5 Mapping tool for existing food intake and consumer behaviour data

Nuts	Unsalted nuts	Nuts and seeds e.g., almonds, peanuts, pumpkin seeds	Nuts and seeds e.g., almonds, peanuts, pumpkin seeds		
	Salted nuts (peanuts, cashews etc)				
Chocolate	Chocolate sweets, toffees or fudge	Chocolates, single or squares	Chocolates, single or squares		
Chocolate bars	Chocolate (e.g., Mars, Dairy Milk)	Chocolate snack bars eg mars/crunchie	Chocolate snack bars eg mars/crunchie	Popcorn	
		Flapjacks, muesli bars, oatmeal cookies	Flapjacks, muesli bars, oatmeal cookies	Pretzels	
		Snackbarproducts eg frikandel, kroket, bamibal, kaassouflé...	Snackbarproducts eg frikandel, kroket, bamibal, kaassouflé...	Peanuts/walnuts/seeds / other nuts	
	<i>Biscuits</i>				
Plain biscuit	Plain (e.g., rich tea, digestive)	Sweet biscuits,chocolate eg digestive, cookies	Sweet biscuits,chocolate eg digestive, cookies		Peanuts
	Sweet (e.g., ginger, custard creams)	Sweet biscuits, plain eg Nice, ginger	Sweet biscuits, plain eg Nice, ginger	Granola bars	
Chocolate biscuit	Chocolate coated biscuits				
	Shortbread				
	Oatcakes				
	Cereal bars, flapjacks				
(in bread and savoury biscuits group)	Savoury biscuits (crackers, crispbreads)	(in bread and savoury biscuits group)	(in bread and savoury biscuits group)		Rice cakes
					Rice cakes
	<i>Cakes</i>				

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Plain cakes (sponge, madeira, ginger etc)	Plain cakes eg fruit, sponge, scones, gingerbread, raisinbread	Plain cakes eg fruit, sponge, scones, gingerbread, raisinbread		
	Cakes with jam, cream or icing (victoria sponge, carrot cake etc)	Rich cakes e.g., chocolate, cheesecake	Rich cakes e.g., chocolate, cheesecake		
Home baked cake					
Readymade cake					
Home baked sponge				Crackers	
Readymade sponge				<i>Dessert foods</i>	
				Cake	
	Pastries, doughnuts or muffins	Buns, muffins, pastries e.g., croissants, doughnuts	Buns, muffins, pastries e.g., croissants, doughnuts	Cake, sponge cake, Japanese cakes	Crackers, cookies
Homebaked buns					
Readymade buns					
	Fruit cakes (all types)	Fruit pies, tarts, crumbles	Fruit pies, tarts, crumbles		
Homebaked fruit pies					
Readymade fruit pies					
				Brownies/cookies	
		Stroopwafle	Stroopwafle	Doughnuts/ sweet rolls/ Danish/ pop tarts	
	Pancakes or scones	Waffles, pancakes, crepes	Waffles, pancakes, crepes		
		Baklava, kantaifi	Baklava, kantaifi		
	<i>Puddings</i>			Fruit crisp/cobbler/pies	
Milk puddings	Milk-based puddings (e.g rice, semolina)	Milk puddings, eg rice, custard, trifle	Milk puddings, eg rice, custard, trifle		

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Sponge puddings (e.g., steamed, syrup, jam)	Sponge puddings	Sponge puddings		
	Custard or other sweet sauces			Sweet muffins/ dessert breads	
	Wrapped ice creams (cornetto, solero, magnum)	Frozen yoghurt, ices/sorbets, ice/ice-cream bars/sherbet			
	Other ice creams (all types)				
(in sweets and snacks group)	(in sweets and snacks group)	(in sweets and snacks group)			
			Chocolate candy		
	Gateau or cheesecake			Pudding/custard	
	Fruit-based puddings (e.g., pie, tart, crumble)				
	Mousse, trifle, meringue				
<i>Drinks</i>	<i>Beverages and Soft Drinks</i>	<i>Drinks</i>			
			<i>Drinks</i>		
Instant coffee	Instant coffee (regular)				
Decaff coffee	Decaffeinate coffee				
	Filter, espresso, or cappuccino coffee	Coffee, milky, latte, cappuccino		Other candy	
		Coffee, Americano, black	Coffee, milky, latte, cappuccino		
		Coffee whitener	Coffee, Americano, black		
Tea	Tea (regular)	Tea (black, green, fruit, herbal)	Coffee whitener		

D4.5 Mapping tool for existing food intake and consumer behaviour data

	Herbal, fruit or decaffeinated tea		Tea (black, green, fruit, herbal)	<i>Beverages</i>	
				Coffee	
Cocoa					
	Hot chocolate	Hot chocolate, Ovaltine, Horlicks made with milk			
Horlicks	Horlicks or Ovaltine	Hot chocolate, Ovaltine, Horlicks, made with water	Hot chocolate, Ovaltine, Horlicks made with milk		
			Hot chocolate, Ovaltine, Horlicks, made with water		
Fruit juice	Pure fruit juice (orange, apple)	Pure fruit juice e.g., orange		Hot tea	<i>Beverages</i>
			Pure fruit juice e.g., orange		Coffee
				Iced tea	
	Tomato juice	Tomato and vegetable juices			
Fruit squash	Blackcurrant squash	Fruit squash/cordial/nectar	Tomato and vegetable juices		
	Other fruit squash		Fruit squash/cordial/nectar		
Fizzy drinks	Regular fizzy drinks	Fizzy soft drinks eg coca cola/lemonade			
			Fizzy soft drinks e.g., coca cola/lemonade		

D4.5 Mapping tool for existing food intake and consumer behaviour data

Low cal fizzy drinks	Diet fizzy drinks	low calorie/ diet fizzy soft drinks			
		Mineral water	low calorie/ diet fizzy soft drinks	Orange juice or grapefruit juice	
		Tap water (not in other drinks)		Apple juice	
				Grape juice	
(in dairy group)	(in dairy group)	(in dairy group)		Other fruit juices	
	<i>Alcoholic Drinks</i>		(in dairy group)	Tomato juice, vegetable juice	
Wine		Wine			Wine
	Red wine		Wine		Sake (rice wine)
	White wine			Soda, regular or diet, with or without caffeine	
				Fruit drinks, Hi-C, lemonade	
Beer		Beer, larger, cider			Beer
	Cider		Beer, larger, cider		Shochu (liquor distilled from sweet potatoes, rice, buckwheat)
	Low alcohol larger or beer				Whiskey
	Dark beer (export, bitter or stout)			Meal replacement, energy/ high-protein beverage	Other alcoholic beverages
	Light beer (lager or continental beers)			Milk, not including milk in cereal	
Port	Sherry, port etc	Port, sherry, vermouth, liqueurs			

D4.5 Mapping tool for existing food intake and consumer behaviour data

Spirits	Spirits or liqueurs	Spirits e.g., Gin, brandy, whiskey, vodka	Port, sherry, vermouth, liqueurs	Wine, wine coolers	
	Alcopops (e.g., Bacardi, breezer)	Sweet alcoholic drinks e.g., alchopops, cocktails	Spirits e.g., Gin, brandy, whiskey, vodka		
			Sweet alcoholic drinks e.g., alchopops, cocktails		
				Beer	
				Liquor, mixed drinks	