Food | Consumer | Health
Designing a world-class infrastructure to facilitate research

Horizon 2020
INFRADEV-1-2014 - Design studies

RICHFIELDS Working Package 7
Deliverable D.7.5

Report on gaps and needs

Potentials and limitations for the use of consumer generated food consumption data in nutrition research

Date delivered:
M20

Authors:
Marcus Maringer, Susanne Ekman, Anne Normann, Naomi Klepacz, Monique Raats, Anouk Geelen

Deliverable lead beneficiaries:
WU

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Project	
Project acronym:	RICHFIELDS
Project full title:	Research Infrastructure on Consumer Health and Food Intake for E-science with Linked Data Sharing
Grant agreement no.:	654280
Project start date:	01.10.2015
Document:	
Title:	Report on gaps and needs
Deliverable No.:	D 7.5
Authors:	Marcus Maringer, Susanne Ekman, Anne Normann, Naomi Klepacz, Monique Raats, Anouk Geelen
Reviewer:	Karin Zimmermann – Project Coordinator prof. dr. ir. Pieter van't Veer – Scientific Coordinator
Start date:	1.10.2015
Delivery date:	31.05.2017
Due date of deliverable:	31.03.2017
Dissemination level:	PU
Status:	Final

Change history:		
Version	Notes	Date
001	1st Final Draft	31.03.2017

Karin Zimmermann | **prof. dr. ir. Pieter van 't Veer**

Project Coordinator | **Scientific Coordinator**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Executive summary

Aim of the present deliverable 7.5 was to identify the potentials and limitations of the tools collected in the inventory of deliverable 7.1, for getting a better understanding of the determinants of food consumption behavior. For that purpose, we investigated the data collection process of food consumption data by these tools, including its purpose, the applied dietary assessment methodology, the types of nutrients calculated and the possible contextual influences on users' dietary behavior. In addition, in order to get an overview of the data associated with the collected dietary assessment data, we investigated the types of contextual data collected by the tools and the sources for exchanging and integrating contextual data from external sources such as wearables, partner apps and aggregators. We found that the vast majority of tools in the investigated sample collected consumer generated food consumption data using food diaries allowing for the input of a large variety of food consumption data from various sources. The quality of the compilation process of the underlying pre-compiled as well as user-generated food databases remains undocumented for the vast majority of investigated tools. Contextual data collected by the investigated tool, in addition to food consumption data, seems to bear interesting opportunities for a better understanding of the determinants of food consumption behavior. The type and variability of this data, however, appears to emphasize contextual data related to weight management, which has been identified as purpose for the majority of tools. Similarly, the large amount of potential influences aimed at changing users' food consumption behaviors (e.g., reminders, social support) and the low level of detail regarding food composition estimations might also be a consequence of the numerous tools aiming at weight management in the inventory. Considering the lack of information provided by the investigated dietary assessment tools regarding the procedures and protocols for data access, the emerging networks of consumer generated data might provide a more efficient opportunity for researchers who want to access and integrate food consumption data with relevant contextual data. Further research is needed, however, in order to better understand the nature of this data networks, their access points and the types and structures of data they exchange. Supporting the compilation of food composition

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

databases, the harmonization of consumer generated data, and the reflection on and interpretation of collected users data, might offer potentially important value propositions RICHFIELDS could provide for its various stakeholders. Application vendors, users as well as researchers could benefit from such services.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Content

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

1. Introduction

1.1 BACKGROUND

An important part of the RICHFIELDS project will focus on the evaluation of the scientific, technical, legal and ethical aspects related to integration and governance of consumer-generated food consumption and lifestyle data. For that purpose, a quality framework based on identified areas of quality have been formulated in Deliverable 7.3. In addition to a framework for the evaluation of the data, Deliverable 7.3 also provided structure and guidance for the data collection process of deliverable 7.3. In Deliverable 7.1 this framework has been implemented and its operationalizations have been used to collect the relevant data which resulted in an inventory of consumer generated consumption data and data collection tools. The inventory contains dietary assessment tools (mainly mobile apps) which were collected based on 2 different search strategies: 1) a more systematic search strategy in which the iTunes and Google Play stores was searched with a predefined set of search terms and selected based on popularity of apps. Other tools have been selected in a more open search strategy in which a variety of sources outside the app stores such as app reviews, blogposts or newsletters were included. In addition, the inventory includes tools from work packages 5 and 6 which investigated food purchase and preparation tools. The present report focusses mainly on the systematically selected dietary assessment tools (n = 194), however, in cases where it provides added value to the evaluations of this report, information from these other pools of apps will be used to complement our findings.

1.2 AIM AND OVERVIEW

The focus of the current Deliverable 7.5 was on the identification of the gaps and needs in terms of quality and availability of data for getting a better understanding of the determinants of food consumption behaviour. In order to get an indication about the quality of the collected dietary assessment data, we investigated its data collection process, including its purpose, the applied dietary assessment methodology, the types of nutrients calculated for food consumption data and the possible contextual influences on data collection. In order to get an overview about the availability of data referring to potential determinants we investigated the availability and types of contextual data collected by the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

tools and the possible sources for exchanging and integrating contextual data from external sources such as wearables, partner apps and aggregators. Finally, we also investigated the availability of public information related to the quality criteria formulated in Deliverable 7.3, which we will discuss at the start.

2. Quality assessment

The following assessments of quality and availability of data was based on the 194 dietary assessment tools collected in the inventory which have been selected based on the systematic search strategy.

2.1 AVAILABILITY OF INFORMATION

The availability of public information about a tool was important for the collection of data for Deliverable 7.1. At the same time the availability of information is a critical quality criterion for evaluating online products and services. Specifically, finding general information about a company, details about their offered tools and services, documentation of customers' policies and support are important indicators about perceived product quality (e.g., Aladwani & Palvia, 2002). A lack of available information around consumer generated food consumption can be interpreted as potential limitation for the integration of this type of data in nutrition research.

2.1.1 AVAILABILITY OF INFORMATION SOURCES

The inventory of Deliverable 7.1 used the following sources of information for data collection: 1) App store metadata, 2) Home pages, 3) Terms and conditions documents, and 4) Privacy statements. The inventory of Deliverable 7.1 includes only apps published on the ITunes store and the Google play store. Hence, the availability of these basic sources of information was 100% (see Table 1). In 79% of the cases a home page was available for data extraction. In these cases, the application vendors provided a Uniform Resource Locator (URL) to a working home or support page. For 10% of the apps in the inventory no URL was provided and no app associated home page was found on Google Search. Since providing a support URL is required for publishing apps in the iTunes store, apps for which no URL was provided were only found in the Google Play store. In 8% percent of the cases an URL was provided, however, the website was unavailable and in 3% of the cases the address referred

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

to a social media landing page. In addition, 67% of the apps in the inventory provided contact information, 38% included a terms and conditions document and in 44% of the cases a privacy policy statement was provided.

Table 1: Percentages availability of information sources (n = 194)

App store	Home page	Terms & Conditions	Privacy statement
100%	79%	38%	44%

2.1.2 AVAILABILITY OF QUALITY RELEVANT INFORMATION

Availability of quality relevant information refers to the information that we were able to extract from the publically available sources described under 2.1.1 and which was considered relevant for identifying the potentials and limitations of the tools in the inventory and the data they collect. It is important to note that the absence of information was interpreted differently for criteria related to data connections. That is, only if the absence of information was considered indicative for the quality of the publically accessible information about a tool it was evaluated as “non-available”. Since information about making data connections with other relevant data is not considered essential information which should be provided by default, we interpreted the absence of information relevant to this criterion in terms of “absence of connections” rather than in terms of “non-availability of information”. All other absent information related to data collection, data accessibility, data ownership and data privacy were considered to be indicative of the quality of the publically available information. Table 2 depicts the percentages of available information for the four categories for the associated criteria.

Data collection criteria: Information relevant for identifying the purpose of the tools was present in 100% of the cases. At this level of app usage and popularity, app vendors obviously follow basic communication strategies by offering solutions for well-defined problems and needs. Similarly, information from which the implemented method for collecting food consumption data could be inferred was available in 96% of the cases and information about the nutrients assessed for 89%. Portion size estimations were less consistently documented. Only 65% of the investigated tools provided additional

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

information about whether the tool allowed users to provide portion size estimations about the amount of food consumed and 61% also provided information about the kind portion size estimation methods implemented.

Data accessibility criteria: Information as to whether data collected by the tools is accessible for extraction by their user was available only for 32% of the tools. 30% provided additional information about the type of data access, and 26% provided information about formats in which the data can be accessed. Information about whether accessing the data requires prior authentication in the form of a login was mentioned for only 9% of the tools.

Data ownership: 32% of the investigated tools included information about the ownership of the collected data and 33% provided information regarding data exploitation licences granted to the tool vendors.

Data privacy: 54% of the investigated tools provided information about whether the tool collects personal identifiable information (PII) and 54% of these tools provided further information about the type of PII collected. Whether and what kind of (mobile) device data gets collected by the tool autonomously (without deliberate input by the user) has been mentioned in 37% of the tools. Information with respect to sharing of PII with affiliated parties, which are parties that by contract need to adhere to the privacy policy of the tool vendor, was offered for 49% of the tools. Information regarding the sharing of data with unaffiliated parties, which are parties that are not bound to the privacy policy of the tool vendor, was offered for 39% of the tools.

In sum, we identified important gaps with respect to the public availability of information relevant to the quality framework developed in Deliverable 7.3. More than half of the investigated tools did not provide a terms and conditions document and nearly half of the tools did not provide a privacy statement. Hence important rules users must accept in order to use a service and the ways a vendor gathers, uses, discloses, and manages their users' data was found to be absent. As a consequence of this lack of legal information, important indicators relevant for data governance such as ownership, usage licences and sharing of PII with unaffiliated parties were relatively under-documented. The methods used for data collection and the type of data collected were in general documented sufficiently. There

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

was a lack, however, of provided information regarding the applied food quantification method(s). Finally, information regarding options and methods for accessing and extracting data from the tools, including protocols for authentication were absent for the vast majority of tools.

Table 2: Percentages availability of information per category and criteria of investigation (n = 194)

Data collection	%	Data accessibility	%	Data ownership	%	Data privacy	%
Tool purpose	100	Access	32	Ownership	32	PII collection	54
Assessment	96	Access type	30	Licence vendor	33	PII type	54
method	89	Format	26			Device data	37
Nutrient	65	Authentication	9			*PII sharing:*	
estimation	61					Affiliates	49
Portions size						Non-Affiliates	39
Method portion							
size							

After describing and evaluating information which was available for the tools, we will next investigate this information with the focus on identifying the potentials and limitations of the tools and their data.

2.2 DATA COLLECTION

The type of information we investigated relevant to the category of food consumption data collection was 1) the purpose of the data collection (or tool), 2) the implemented dietary assessment method, 3) the nutrients estimated from the food consumption data, and 4) possible contextual influences on users' dietary behavior.

2.2.1 DATA COLLECTION PURPOSE

We identified three overarching purposes of tools as they were propagated and marketed through the descriptions of the tools (see Table 3): 1) tools that were aimed at supporting some form of behavioral change (e.g., weight loss), 2) tools that focused on medical support (e.g., diabetes, food intolerance), and 3) tools that focused on efficient and enjoyable ways of recording and memorizing the foods people consume. Note that some tools in the inventory were described as serving more than one purpose and these purposes were usually closely related. For instance, a tool could be described as supporting behavioral

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

change for weight loss as well as providing medical support for diabetes (e.g., MyNetDiary, Noom Coach), chronic stomach issues (e.g., Tummy Tracker) or recovery after bariatric surgery (e.g., Baritastic). In fact, behavioral change tools described in 10% of the cases to also aim at some kind of medical support. In contrast, the group of tools categorized as food logging solely focused on supporting efficient dietary assessments without propagating any purpose for behavior change or medical support.

2.2.1.1 BEHAVIOR CHANGE TOOLS

The group of tools propagating behavioral change represented with 83% the vast majority of tools in the inventory. Within that category of tools, weight management (e.g., by logging energy intakes and expenditures) was with 63% the most prominent aim described by the tool vendors. Tools supporting a healthy diet in general (e.g., eating sufficient amounts of vegetables and fruits) have been identified in 27% of the cases and tools supporting a special diet (e.g., Indian, vegan, low carb, low glycemic index) in 5% of the cases. Weight management co-occurred in 18% of the behavioral change tools with the general aim of supporting a healthy diet.

Table 3: Described purposes for data collection (n = 194)

Behaviour change	**83%**	**Medical support**	**17%**	**Food logging**	**6%**
Weight management	63%	Diabetes	10%	Food diary	6%
Healthy diets	27%	Food intolerance	5%	Food photos	4%
Special diets	5%	Bariatric surgery	2%	Restaurant menus	1%
		Eating disorders	1%	Beverages	1%
		Kidney diseases	1%		

Note: Since tools in the inventory were often described as serving more than one purpose the individual percentages do not add up to the overall percentages

2.2.1.2 MEDICAL SUPPORT TOOLS:

Tools aiming at medical support were represented with 17% in the inventory. Supporting the management of diabetes (e.g., blood glucose logging, diabetes prevention programs), was found to be the most prominent aim described within the group of medical support tools and accounted for 10% of them. The aim of helping with food intolerances (e.g., irritable bowel syndrome, lactose intolerance, food allergies) were described in 5% of the cases. Other targets for medical support were bariatric surgery, eating disorder and kidney

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

disease patients which made up 4% of the tools in the inventory. Some medical support tools were assigned multiple purposes. Apps such as Health Watch 360 supported the management of several food related diseases. Within the group of medical support tools only 1 tool, the mySugr App, was registered as a medical device (FDA class 1 medical device).

Food logging tools: Tools that solely focus on efficient and enjoyable ways of recording and memorizing the foods people consume made up 6% of the inventory. All of these tools implemented a food diary and 4% allowed for visualization of the foods by the means of food images. The logging of restaurant menus, beverages and locations were aimed at in 2% of the cases.

In sum, the vast majority of tools in the inventory have been described as aiming at behavioral change, with support for weight management as the most prominent purpose provided within this category. This implies that these tools will most likely involve some sort of behavioral change technique (e.g., feedback, reminders, advices), which potentially act as (intended) contextual influences on their users' food consumption behavior. The propagated value of tools which were solely aimed at providing food logging services without the emphasize on changing people's behaviors, seem to rest on the efficiency of the implemented dietary assessment method and the capturing of the experience of food consumption (e.g., visualization, localization, sharing). Although several of the tools in the inventory aimed to support the management of medical conditions, only a single app was registered as medical device. This implies that none (except for one) of the tools in the inventory can be held accountable for the medical information and recommendations they provide and should not be used without the supervision of a medical professional.

2.2.2 DIETARY ASSESSMENT

2.2.2.1 ASSESSMENT METHODOLOGY

For each tool in the inventory, data was collected about the type of dietary assessment methodologies applied. The most widely implemented method was a food diary which was identified in 88% of the tools (n = 171). Food diaries allowed for daily records of the foods and drinks people consumed at the individual level, at a certain moment in time (e.g.,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

meals, snacks, date) and over a certain period of time. Although these basic food diary features were implemented by the vast majority of tools, the sources and type of food consumption data collected differed.

2.2.2.2 FOOD CONSUMPTION INPUT SOURCES

The majority (53%) of the food diary methods allowed for inputs from pre-compiled food databases. The size of these food databases varied from 1,900 food items (Calorie counter - Calories!) to 5,000,000 food items (MyFitnessPal). However, we were able to extract information regarding a link to verified sources of the database (e.g., USDA, NEVO) and hence the quality of its compilation process, for only 10% of the tools who offered precompiled database inputs (e.g., My Daily Plate, FitDay, iTrackBites, Virtuagym Food). In none of the tools' public information we investigated we found indications about a link to the UK Composition of Foods Integrated Dataset (COFIDS).

In addition to precompiled food databases in 58% of the tools in the inventory we were able to identify that the implemented food diaries allowed for reusable user generated food database entries (e.g., Fitbit, Yazio, Lifesum, FitDay) of which some could be shared with other users (e.g., MyFitnessPal, Yazio, Lose it!). For users this involves generating new food items and associated nutrition values and adding them to the (empty or precompiled) database. The way food composition values are calculated and assigned to these custom items and how their quality is checked remains undescribed for most of the apps. Combining or aggregating precompiled or added entries into new food items was one identified option (e.g., FitDay, MyPlate Calorie Tracker, Carbs & Cals) and referencing nutrition data from packaged food labels another (e.g., Yazio, SparkPeople, Virtuagym). In some of these tools users were able to report incorrect food items (e.g., Fatsecret, Lifesum). MyNetDiary for instance asked their users to send in images of food labels with the correct values for outdated or incorrect food entries in the database. Other tools introduced a quality label in the form of a "checkmark" which indicates that the entry has been validated as accurate or complete by either users or vendors (e.g., MyFitnessPal, Lose it!).

Allowing for various types of food consumption inputs such as generic foods (e.g., Baked Potatoes, Cooked beef), food images or recipes, which do not come with documented and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

readily accessible nutrient information (such as packaged products), certainly increases the necessary resources and expertise for user compiled food data quality checks. Unfortunately, evidence regarding the validity of this food composition data and the associated dietary assessment tools is nearly absent (for an exception see Carter, Burley, Nykjaer, & Cade, 2013).

We identified that some tools allowed for specific types of customizable or user generated data input sources such as favorites (10%; e.g., FitDay, Fitatu, The Secret of Weight), frequently consumed foods (6%; e.g., SmartFoodTracker, Fitatu, Cals & Macros) or recently consumed foods (5%; e.g., Lifesum, Weight Watchers, Calorie counter - Calories!). These datasets are aimed at enabling a more focused and efficient food item search and selection strategy. One application in the investigated sample offered the option of programming future meals for organizing shopping lists and autonomous? food intake tracking (Mango).

Finally, in the current sample of more popular dietary assessment apps we identified no tool which allowed smart kitchen scales as direct input source for food consumption data. Note, however, we identified several tools which supported smart kitchen scale inputs in our sample of more innovative tools which were not included here due to popularity and number of usage (e.g., Prep Pad for iPhone, Escali SmartConnect, SITU Scale). Within this group of tools, we also identified food consumption inputs from other sensory sources such as spectrometers (e.g., DietSensor) or impedance sensors (e.g., Healbe GoBe). The usability, reliability and validity of these seemingly innovative diet sensors has yet to be proven. There have been several investigations into fraudulent crowdsourcing practices from such systems (e.g., Bioring, BitBite, Proscan).

2.2.2.3 FOOD CONSUMPTION INPUT TYPES

Food diaries applied in the collected tools allowed for the input of various types of food consumption data (see Table 4). Generic food items such as boiled carrots, smoke sausage or salted beef could be logged in 62% of the cases (of which 78% included a precompiled food database; e.g., Argus, Cronometer, Ultimate Food Value). Labeled or packaged food products have been identified as possible input type in 27% of the tools (e.g., CarbsControl, My Macros+ Diet, HealthifyMe) and 22% implemented a barcode scanner for efficient

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

identification and logging of labelled products (e.g., Nutritionist+, SparkPeople, Cals & Macros).

The direct input of the amount of energy or nutrients (fat, carbohydrates, protein) consumed without further specification of the associated food product, was found to be permitted in 19% of the tools (e.g., Calories Calculator, Weight Calorie Watch, Calories Carb Prot Fat Counter). We were able to identify water consumption inputs in 16% of the tools (e.g., My Diet Coach, FoodPrint Diet by Nutrino, WeightWar) of which four tools focused exclusively on water consumption (e.g., Water Log, Daily water, Water drink reminder).

Table 4: Food consumption input types of food diaries (n = 171)

Food consumption input types	%		%
Generic food items	62	Food images	16
Labeled food products	27	Restaurant dishes	12
Nutrients	19	Recipes	8
Water	16	Voice input	3

Note: Since tools in the inventory were often described as allowing more than one input type the individual percentages do not add up to 100%

Food images have been allowed as input in 16% of the cases. In 11 of these tools (total 5%) the collected food images were used to estimate energy and nutrient intakes or evaluate the foods depicted in the images based on nutrient or energy content. These estimations or evaluations were provided by either software, diet coaches or users themselves. That is, we were able to identify 3 tools (Fatsecret, FoodLog, Logameal) which described to use an image recognition software for this task[1]. Other tools asked their users to provide additional ingredient information on which composition calculations were based (Yumget) or sent the images to a diet coach for further analysis (e.g., HapiCoach). The rest of these tools allowed their users to either complement their food images with energy values (e.g., Food Diary, Diet Tracker Lite), with personal evaluations (e.g., Activ8rlives, Pic Healthy) or general notes (e.g., FoodSnap, Careot).

[1] After the data in the inventory had been collected, the tool Lose it! added “Snap It” food image recognition, which is supported by food image analysis models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Other food image diaries aimed at, for instance, encouraging their users to make smarter choices by visualizing their food choices (e.g., See How You Eat, SimpleWeight, The Slow-Carb Diet), emphasizing the efficiency of food photo diaries in order to increase motivation for food consumption tracking (e.g., My Diet Tracker) or visually sharing food consumption experiences with other users, friends or family (e.g., iFood Diary, Meal Planner and Food Manager).

Food diary input in the form of dishes which users were able to select from restaurant menus including fast food restaurants, was identified in 12% of the cases (e.g., MyFitnessPal, Calorie Counter+, FoodPrint Diet by Nutrino). In 8% of the investigated tools we found that users were able to enter recipes into the food diaries which were linked to a set of instructions telling how the food was prepared and cooked, including a list and quantities of ingredients included (e.g., My Diet Diary Calorie Counter, Noom Coach, Health & Weight Loss Coach). In 3% of the investigated tools we identified the possibility for voice input (e.g., Calorie Counter by Calorie Count, Track, HI - Health & Fitness Tracker, HealthifyMe).

In sum, food diaries of the investigated food consumption tools allowed for the input of a multitude of different types of food consumption data, which could either be generated by users or selected from precompiled database entries. On the one hand the support for various types of food consumption inputs supports the investigation of a relatively broad and variable dietary pattern and hence be of potential interest for nutrition researchers. On the other hand, however, the large variability of customized inputs of which no nutrient information is available, renders the procedure for estimating and verifying food composition values of entered food consumption data even more challenging.

2.2.2.4 PORTIONS SIZE ESTIMATIONS

In only 58% of the tools we were able to extract information about whether the implemented food diary included estimations of portion sizes for dietary intake assessments. 36% of the tools mentioned visual portion size estimations by the user based on standard units such as cups, spoons, slices (e.g., Weight Watchers, Noom Coach, Health & Weight Loss Coach). 20% of the investigated tools implemented weight and volume

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

estimations (e.g., Lose Weight - Diet Planner, Simple Calorie Count, Nutrition Tracker). Often people can estimate the size of items more readily when they can visualize their size, shape or weight, in comparison to something else. Only a few tools, however, provided visual aids in the form of images or graphics for supporting the visualization of portion sizes and helping to relate an appropriate serving to the users' consumed foods (e.g., Aqua alert, Health Watch 360, Carbs & Cals). The weighing of food has been supported only in the previously mentioned smart kitchen scales tools found in the group of new and innovative dietary assessment tools (e.g., Escali SmartConnect).

Having discussed the sources and types of food consumption data, we turn next to the types of energy and nutrient values estimated for the entered food consumption data.

2.2.2.5 ENERGY/NUTRIENT INTAKE OUTPUTS

The most commonly estimated food composition value was energy, which was mentioned in 55% of the investigated tool information (see Table 5; e.g., Calorie Counter PLUS, iSkinny, Revive). The most commonly estimated group of nutrients amongst the investigated tools were macronutrients, which were found in 46% of the tools information (e.g., UP by Jawbone, MyPlate Calorie Tracker, Eat This Much). We were able to identify estimations of micronutrients in only 12% of the cases (e.g., MyFitnessPal, Fitbit, Lose it!). Finally, food scores as composition output has been identified in 16% percent of the tools (e.g., Smiley Diet, Logameal, Lark).

The difference in percentage of tools estimating energy versus the percentage of tools estimating micronutrients might be explained by the large percentage of tools aimed at weight management. Micronutrient estimations have been described only in 9% of the apps classified as supporting weight management (e.g., FitDay, My Diet Diary Calorie Counter, MyNetDiary). In contrast, estimations of energy consumption have been mentioned in 72% of the weight management tools. For tools which aimed at supporting a healthy diet, micronutrient estimations have been mentioned in 20% of the tools. Overall this still seems a rather low percentage which is consistent with the notion that the aim of the tool might affect the food composition outputs provided.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Table 5: Food diaries' food composition outputs and potential contextual influences (n = 171)

Energy/Nutrient intake outputs (92%)*	%	Contextual influences (52%)**	%
Energy	55	Nutrition advices	30
Macronutrients	46	Reminders/Notifications/Alerts	24
Food score	16	Connected users/Social support	17
Water volume	16	Coaching	9
Micronutrients	12	Challenges	9
Food groups	4		

*Since tools in the inventory were often described as calculating more than one type of output the individual percentages do not add up to the overall percentage

**Since tools in the inventory were often described as implementing more than one contextual influence the individual percentages do not add up to the overall percentage

Interestingly, none of the tools without precompiled food databases provided estimations of consumed micronutrients. The large majority of the tools (66%) which provided estimations of micronutrients also permitted customized or user generated food consumption inputs (in total 8% of the cases). It is very likely that the nutrient estimations of this user generated food data will not go beyond the nutrition information provided on the labels of packaged food products.

Food photo diaries which applied image recognition technology provided estimations of food compositions on a lower level of detail. Vendors of the FoodLog app claimed for instance that their image recognition software provides energy content. The Logameal food diary application information stated to calculate a food score based on nutrition density. Finally, for Fatsecret, which is the third app which implemented an image recognition software, it remained undocumented what type of composition estimation their software (FoodSnap) is producing. Whereas the FoodLog image recognition technology has been developed and tested at the University of Tokyo (Aizawa et al., 2014), we are not aware of any test or validation regarding the other systems.

Next we will discuss the possible contextual influences on the collected food consumption behaviors.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

2.2.2.6 CONTEXTUAL INFLUENCES:

Overall in the majority of tools investigated we were able to identify some form of potential contextual influence on users' food consumption behavior (52%). Several types of possible contextual influences have been identified such as nutrition advices in the form of opinions or recommendations offered as guidance (27%; e.g., Lifesum, Argus, Fatsecret). Such advices have also been described as personal feedback in the form of short messages conveying reflections, suggestions or criticism on users' performances or consumption behaviors (e.g., MyNetDiary, MyFitnessPal, UP by Jawbone). 24% of the investigated tools mentioned alerts such as eating, drinking or food logging reminders (e.g., Daily Water, Weight Loss & Fitness Program, Easy Fit Calorie Counter), or notifications, badges or rewards for coming close to and reaching predefined weight or nutrition goals or limits (e.g., Fitbit, My Daily Plate, MyFitnessPal).

Connected users refers to users which follow each other's progress, posts, comments and other sorts of shared information and are sources of social support and motivation. Such potential influences on user's dietary consumption behavior have been found in 17% of the tools (e.g., Calorie Counter+, Ultimate Food Value Diary Plus, My Diet Diary Calorie Counter). Personal coaching for the achievement of user specific diet or weight goals have been identified in 9% of the tools (e.g., Noom Healthy Weight Loss Coach, Virtuagym Food, All-in Fitness) and the option for inviting other users to compete or take part in weight loss or exercise challenges was offered in 9% of the tools (e.g., Lose it!, Pic Healthy, Healthy 365, My Diet Coach).

The prevalence and type of potential contextual influences might be explained by the vast number of tools which are aimed at behavioral change. The adherence of these contextual influences to the scientific standards appears to be insufficient. Pagoto et al., for instance reported that only a few of the behavioral strategies found in evidence-based weight-loss interventions have been applied in popular weight loss apps (Pagoto, Schneider, Jojic, DeBiasse, & Mann, 2013). In similar vein, Davis et al. found that most of the calorie counting apps in their sample contained only minimal adherence to health behavior theory (Davis et al., 2016). Whether the presence of these contextual variables can provide further insights into drivers and barriers of food choice and consumption depends also on the availability of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

this data and on the controllability of the influences. If the content and timing of provided alerts, reminders or personal advices are not logged alongside food consumption data, and hence are not associable, the behavioral explanatory value of these contextual influences on the collected food consumption data remains limited. Interestingly, some tools offer the possibility to enable or disable alert messages, for instance, messages prompted due to a user reaching a specific nutrition or weight goal (e.g., My Daily Plate, MyFitnessPal). However, if researchers are not able to control the on or offset, the intensity or content of certain contextual influences, it becomes a great challenge to generate further insights from these possible drivers or barriers for people’s food consumption behavior.

2.2.3 POTENTIAL AND LIMITATIONS OF DATA COLLECTION

In the following section we will sum up the identified potentials and limitations regarding the collection of consumer generated food consumption data in popular dietary assessment apps:

Potentials:

1) The vast majority of tools in our sample collected consumer generated food consumption data at the individual level, on a daily basis, at a certain moment in time and over a certain period of time. Data collected by this methodology has the potential to provide personalized food intake profiles and hence could provide a better insight into habitual food consumption behaviors and its changes over time, at the individual level.

2) The identified food diary methods in the current sample allowed for inputs from various data sources such as pre-compiled databases seeded with often large numbers of various types of food products and related nutrients as well as user-generated and individualized databases including repetitive, favorite or recently consumed foods and recipes. This supports an effective, user-friendly collection with broad and variable dietary consumption patterns, which might to better understand people' habitual food consumption behaviors.

3) Due to the prevalence of tools with the aim on behavioral change we identified various types of potential contextual influences on users’ food consumption behaviors. This data might be able to reveal relevant information about the determinants of food consumption

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

behaviors including the effectiveness of technology driven behavior change interventions in the short and in the long term.

Limitations:

1) The compilation and quality maintenance procedures of precompiled and user generated food composition databases used by these apps remains unknown. Consequently, conclusions with respect to the relationship between the nutrient profile of the foods people consume and the onset and development of nutrition related diseases might be limited. Quality standards and guidelines are needed for food composition data compilation and food consumption data integration workflows.

2) The level of detail of the estimated food composition values is rather low, with the vast majority of apps focusing on energy and macronutrients. This limited depth of food composition profiling might be a barrier for research about the associations between specific nutrients and health outcomes.

3) Although food photo diaries offer new popular and effective ways of logging, visualizing and sharing of food consumption data, the added value for detailed food composition estimations of the foods depicted in the images remains limited.

4) The multitude of potentially non-evidence based, non-registered and uncontrollable contextual influences might pose a barrier towards a better understanding of the determinants of food consumption behaviors as well as on providing an unbiased insight in peoples' habitual food consumption behaviors.

In the previous section we focused on the potentials and limitations of the collection process of consumer generated food consumption data. This data, however, is not collected in isolation of other potentially relevant data. A vital source for a better understanding of the possible drivers and barriers for people's food consumption behavior might come from the various existing associations between food consumption data and other relevant health and lifestyle data. We will discuss these data connections and their potentials and limitations in more detail in the following paragraphs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

2.3 CONTEXTUAL DATA

In addition to food intake data, which was collected for the most part by food diaries, 85% of the tools in the inventory offered additional features for tracking various types of contextual data. The tools with the largest number of contextual data collected was the Nutrition Tracker with 18, Health Watch 360 with 14 and Fitbit, Food Print by Nutrino and Activ8rlives with each 13 different types of additional contextual data collected. An average of M = *3.4* (SD = 3.1) additional variables collected by the investigated tools indicates that data collection in the dietary assessment tools in our sample is not limited to food consumption data, but is in the vast majority of cases enriched with other potentially interesting contextual data points. In the following paragraphs we discuss the identified types of contextual data connected to dietary assessment data. This contextual data will be mapped onto the DONE framework of determinants of nutrition and eating (Stok et al., 2017). The DONE framework describes determinants related to nutrition and eating behaviors as discussed in various disciplines (e.g., Psychology, Public health, Food Sciences). This should provide us with an evidence based indication about the potentials and gaps of this data related to a better understanding of the causes and complexities of food consumption behaviors.

2.3.2 TYPES OF CONTEXTUAL DATA

We identified 5 categories of contextual data connected to dietary consumption data. These categories contain data related to users’ 1) psychological motivations, 2) physical health, 3) physical activity, 4) social interactions, and 5) physical location. We will discuss the data identified and assigned to these categories in the following paragraphs.

Psychological motivations: Motivation is a theoretical construct used to explain behavior on the psychological level. Motivation is comprised of the beliefs, goals, desires and preferences which drive or inhibit people’s actions. In our inventory we identified that 61% of the dietary assessment tools collected some form of data related to psychological motivation (see Table 6).

The majority of the identified motivational data was related to some form of goal setting features (48%). 32% of the tools allowed users to set a desired intake of nutrients or energy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

that the user envisions to achieve (e.g., My Diet Coach, My Daily Plate, Yazio). 16% of the tools allowed users to set a desired body weight (e.g., MyNetDiary, Weighfit, SmartFoodTracker). Other types of goals monitored were goals related to an envisioned state of physical fitness including muscular strength and endurance (5%; e.g., Fitbit, Virtuagym, Weight Loss & Fitness Program) or an anticipated level of daily hydration (4%; Easy Fit Calorie Counter, Health & Weight Loss Coach, Water log). Important psychological predictors for food consumption behavior are health cognitions (e.g, anticipated states) and behaviors (e.g., physical activities) in particular related to weight control (Symmank et al., 2017). Behavioral intentions are important predictors for goal directed behaviors (e.g., Ajzen, 2015). Setting goals and tracking the progression towards these goals might provide potentially interesting information about users' ability to self-regulate. Hence, relating food consumption behaviors to users' behavioral intentions towards changing their physical condition, has the potential to improve our knowledge regarding the determinants of its occurrences.

Another group of motivational data identified in the dietary assessment tools were related to affective states and preferences (17%). 10% of the tools allowed users to save their preferred foods in a list of favorites (see 2.2.2). Food preferences have been implicated in the development and maintenance of overweight or obesity (e.g., Mela, 1996), and hence provide potentially relevant contextual insights in the reasons for people's food choices. Food preferences have also been identified by the DONE framework as important determinants of food consumption behaviors (Stok et al., 2017).

In addition, 5% of the tools allowed their users to record their mood or emotions (e.g., RR Eating Disorder Management, iFoodDiary, Lose weight with Applause). 2% allowed their users to record their experienced stress level (e.g., FoodPrint Diet by Nutrino, HealthWatch 360, Bowelle IBS tracker). People's affective states such as experienced emotions or stress are considered important determinants of food choices. Stress might change the perceived amount of food people consume as well as the type of food actually consumed (Oliver & Wardle, 1999; Oliver, Wardle, & Gibson, 2000). Contextual data about users' experienced affective states and its potential to provide relevant cues regarding the determinants of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

food consumption behavior has also been identified by the DONE framework. However, the low prevalence of data on users' affective states indicates that there is a gap regarding this relevant determinant of food consumption behaviors.

Similar to food preferences, food habits are important determinants of food consumption behaviors (Stok et al., 2017) and could potentially be inferred in tools which allow users to create a list of frequently or repetitively eaten foods (see 2.2.2).

Table 6: Percentages motivational data collected by food consumption apps (n = 171)

Psychological motivation (61%)*	**%**		**%**
Nutrition goals	32	Food habits	5
Weight goals	16	Hydration goals	4
Food preferences	10	Stress level	2
Fitness goals	6	Hunger	1
Mood/Emotions	5		

* Since tools in the inventory were often described as collecting more than one type of contextual data the individual percentages do not add up to the overall percentage

Finally, in a tiny fraction (1%) of the investigated tools we identified the option for tracking users visceral state of hunger (e.g., HealthWatch 360, iFoodDiary, Baritastic). Hunger is primarily linked to food intake by its purpose to maintain an energy balance (Herman, Polivy, Lank, & Heatherton, 1987). The DONE framework has identified hunger as an important situational determinant of food consumption behavior. Similar to affective states, the prevalence and hence availability of this data is limited.

Overall, the collected contextual data we categorized as psychological motivational data, maps to some of the identified determinants of food consumption behavior by the DONE framework. In fact, psychological factors received the greatest interest within the group of individual predictors (Symmank et al., 2017). Hence, data related to users' goals, preferences, habits, emotional and visceral states were considered to provide relevant psychological and situational explanations for the drivers and barriers of people's food consumption behaviors. The availability of this data is inconsistent, however. Contextual data related to users' goals are the majority, followed by food preferences and habits. Data concerning users affective state and hunger are scarce.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Physical health: Data related to physical health has been identified by the DONE framework as relevant determinant for food consumption behavior (Symmank et al., 2017). Hence the following identified physical health data may contain potentially important information for a better contextual understanding of the collected food consumption.

59% of the tools in the inventory allowed their users to monitor indicators about their current health status (see Table 7). Tracking of anthropometric parameters such as body weight (42%), body mass index (12%) or body composition (6%) was found to be supported in total 60% of the investigated tools (e.g., Activ8rlives, Mevo, iSkinny). The storage of photographic images of the users' body has been identified in 4% of the tools (e.g., My Diet Coach, Nutritionist+, Body Tracker). Anthropometric data might be in part explained by food consumption behavior, and hence provides important information regarding food choices and habits. Anthropometrics data has been identified as influencing food consumption behavior on the individual biological level (Stok et al., 2017). Hence this data, and due to its prevalence, especially body weight data has a clear potential in supporting a better understanding of the determinants of users' food choices. Photographic images of users bodies, although not commonly collected, might provide additional interesting information regarding physical health information based on visual health indicators (e.g., skin tone, acne, cold sores).

Monitoring of users' body temperature was allowed in 1% of the dietary assessment tools (e.g., Activ8rlives, Nutrition Tracker). Although there might be no long term relationship between body temperature and food choices (people might change their diet when they have a fever), it can provide an important indication about a person's physical health status. This parameter, however, is available only in a very small fraction of tools.

Physical fitness is a more general state of health and well-being and relates to the ability to perform certain sports or activities. General physical fitness was monitored in 2% of the tools (e.g., Smart Score Calculator Pro, FitWell, Mevo). Muscular strength is monitored in 1% of the tools (e.g., Trainer - Workout & Nutrition, SparkPeople). Again, the prevalence or availability of this data seems to be rather low.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Table 7: Percentages of types of health data collected by food consumption apps (n = 171)

Physical health (59%)*	%		%
Body weight	42	Blood pressure	4
BMI	12	Physical fitness	2
Body composition	6	Allergies	1
Medications	5	Oxygen saturation	1
Symptoms	6	Body temperature	1
Body images	4	Muscular strength	1
Blood sugar	4		

* Since tools in the inventory were often described as collecting more than one type of contextual data the individual percentages do not add up to the overall percentage

The monitoring of drugs or other substances used by users to treat diseases or injuries were identified in 5% of the tools (e.g., mySugr, HealthWatch 360, Diabetes Tracker MyNetDiary). Symptoms in the form of subjective evidences of current (not further specified) diseases were found in 6% of the tools (e.g., Food Diary, mDietGuru, mySymptoms) and the reporting of allergies in 1% (e.g., Nutrition Tracker, FoodPrint Diet by Nutrino). Medical symptoms such as allergies or intolerances are relevant parameters for food-related physiology, and have been identified as important determinants of food choices by the DONE framework on the individual biological level.

Other indicators of physical health are available through monitoring of data related to user's blood compositions. In total we found 5% of the tools allowed for monitoring of blood sugar (4%; e.g., Diabetes Tracker MyNetDiary, Health-Tracker, Noom Coach), or oxygen saturation (1%; e.g., Activ8rlives, S Health). Blood pressure monitoring has been identified in 4% of the apps (e.g., HealthWatch 360, Activ8rlives, HI - HAPICoach).

In sum, the majority of data related to physical health seems to be related to users' anthropometrics such as body weight or BMI. There is only a small number of tools which collects data about other physical health indicators such as medical symptoms, medications, and physical fitness.

Physical activity: Physical activity refers to any bodily movement that requires some form of energy expenditure. Physical activity is not only an important indicator of physical health but also has a direct influence on the quantity of the foods consumed (e.g., Mathiassen &

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Hollema, 2014). In general, the corresponding determinant in the DONE framework might be the category of health related behaviors which has been identified on the individual situational level (Stok et al., 2017).

Monitoring of peoples' routines and exercises refer to physical activities that works a body at a greater intensity than the usual level of daily activity. Because energy expended when being active increases the intensity of hunger and hence drives food consumption (e.g., Hopkins, King, & Blundell, 2010), physical activity data has the potential to increase our knowledge regarding the determinants of food consumption. Monitoring of activities was identified in 33% of the tools (e.g., YAZIO, MyNetDiary, Easy Fit Calorie Counter; see Table 8). The logging or recognition of various activity types (e.g., swimming, cycling, running) and the calculation or defining of activity levels have been identified in 14% (e.g., Fitbit, S Health, Argus) and 7% respectively (e.g., MyPlate, SmartFoodTracker, Fitatu Calorie Counter). 9% of the tools in the inventory supported the logging of number of steps taken (e.g., Fitbit, Noom Coach, MyFitnessPal).

Sleep and sleeping patterns, which have been identified as important biological determinant of food consumption behaviors by the DONE framework, were only identified in 6% the of tools (e.g., Jawbone Up, HAPICoach, Lose it!). Sleep and sleep patterns are important indicators due to its influence on peoples' food consumption behaviors (Lundahl & Nelson, 2015). In addition, food consumption has also been found to influence people's sleeping patterns (Crispim et al., 2011). The prevalence of this type of important contextual data measured by dietary assessment apps is relatively low, however. Dietary assessment tools can and do link to partner apps such as fitness trackers for collecting sleep data, though (see 2.3.1).

The majority of physical activity data collected by the tool in the inventory seem to be comprised of data related to users' exercises. Data on sleep and sleep patterns was less commonly identified.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Table 8: Percentages physical activity and location data collected by food consumption apps (n = 171)

Physical activity (48%)*	%	Physical location/Point of sale (12%)*	%
Exercise	33	Venue name	12
Activity type	14	Geo coordinates	1
Steps	9		
Activity level	7		
Sleep pattern	6		

* Since tools in the inventory were often described as collecting more than one type of contextual data the individual percentages do not add up to the overall percentage

Physical location/Point of sale: The physical location of food consumption and preparation has been identified as important factor for food consumption behaviors. For instance, the frequency of eating food which has been prepared away from home has been related to a higher BMI (e.g., Seguin, Aggarwal, Vermeylen, & Drewnowski, 2016). Eating environment has also been identified by the DONE framework as important determinant of food choices (Stok et al., 2017). 12% of the tools in the inventory were identified to allow for inputs of dishes from restaurant menus (see 2.2.2). This implies that food consumption data collected by these tools might contain information regarding the location where the food was purchases. Whether the food was consumed at the same location remains inconclusive though. Geo coordinates provided by a GPS unit which have been identified in 1% of the tools (see Table 8; e.g., Spoonacular Meal Planner and Food Manager, Careot - Nutrition Tracker) could potentially overcome this ambiguity. Careot, for instance, allows their users to tag their meal with geolocation data, which will be used to retrieve a meal for food logging purposes if a person enters that same location again. Whether this can establish certainty about the physical location of food consumption, however, depends on whether a match can be established between the actual time of consumption and the time present at the location. Otherwise this location data should rather be interpreted as referring to point of sale data.

Social interactions: Social interaction refers to the dynamic, changing sequence of social actions between individuals. Such interactions can be important sources for social influences which have a direct effect on the amount of food consumed (e.g., Redd & de

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Castro, 1992). The DONE framework also identified social influences as important determinant for food consumption behaviours (Stok et al., 2017).

17% of the tools offer social media platform infrastructures for exchanging data and information with other connected users (see Table 9). These platforms potentially offer interesting sources for (unstructured) contextual data related to users' food consumption behaviors. The sources of this data are interactions between users such as following each other's progress, posting personal experiences or discussing, evaluating and commenting on each other's contributions and meals. 17% of the tools allowed their users to share data and interact with each other on public forums (see 2.2.2), and 11% allowed their users to share their data and progress updates with popular social media networks.

Table 9: Percentages social interactions, and unstructured data collected by food consumption apps

Social interactions (17%)*	%	Unstructured (15%)	%
Connected user interactions	17	Notes/Comments/Posts	15
Social media interactions	11		

* Since tools in the inventory were often described as collecting more than one type of social interaction data the individual percentages do not add up to the overall percentage

The social interaction data logged in these public platforms, however, is unstructured. This means that the information either does not have a pre-defined data model or is not organized in a pre-defined manner (mainly text format). Hence, although this data potentially provides interesting insights regarding users' social networks, the type of information this unstructured data contains and what it means in terms of its social influences on users' behavior cannot a priori be predicted. In addition, 15% of the tools in the inventory allow their users to complement their dietary data with additional notes or comments (see Table 9; e.g., FitDay, Cronometer, Lose Weight). Similar to the social interaction data this data is unstructured. Both sources of data, however, might potentially provide a deeper insight in users' food beliefs, health cognitions, or self-regulation strategies, all of which are considered important determinants of food consumption behavior. In addition, this unstructured data might also contain relevant cues regarding users' subjective sensory perception (e.g., taste) or their food knowledge and skills.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

In addition to collecting contextual data, dietary assessment tools can also integrate data from other external sources such as wearable devices, partner apps or aggregators which can dramatically increase the type and amount of data available within a tool. We will discuss these external sources for contextual data in the following paragraphs.

2.3.2 SOURCES FOR EXTERNAL CONTEXTUAL DATA

External sensory devices as means to collect contextual data has been identified in only 4% of the tools in the inventory (e.g, Fitbit, Activ8rlives, HAPICoach). Fitbit for instance enriches collected food consumption data with data from its heart rate monitor, accelerometer or GPS unit. The tool Activ8rlives connects to its pulse oximeter for monitoring oxygen saturation, to its smart thermometer to monitor body temperature, and receives blood pressure data from its connected blood pressure monitor.

Even though the vast majority of tools do no connect directly to external sensory devices for collecting contextual data, we identified two other external sources which enables tools to collect contextual data indirectly, which are partner apps and aggregators.

Partner apps: In addition to monitoring nutritional intakes and food consumption behaviors, providers of especially mHealth services show an increased tendency to support the exchange and integration of services and data from other vendors (Research2Guidance, 2016). Hence, partner apps refer to tools of other vendors which are authorized to exchange data. An application that monitors how much calories a user consumes each day, might connect to another dietary assessment tool to complement its food consumption data, to a heart rate monitor of a third party fitness tracker for collecting exercise data, or a sleep monitoring kit from yet another vendor for collecting data on sleep rhythms. The application might also connect to social network applications such as Twitter, Facebook or Instagram in order to provide status updates to the social network about a user's personal goals, progresses and experiences.

In our sample we identified that 24% of the food consumption tools connected to at least one other dietary assessment tool included in the inventory. Tools with the largest percentage of connections with other dietary assessment tools in the inventory were Fitbit

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

connecting with 14%, followed by Jawbone Up connecting with 10%, and MyFitnessPal connecting with 5% of the other dietary assessment tools in the inventory.

In addition to dietary assessment apps the inventory contains a group of tools categorized as popular health and fitness trackers (n = 13; see Deliverable 7.1). The purpose for adding this group of tools was to investigate its possible connections and data integrations with the group of dietary assessment tools. 14% of the investigated dietary assessment tools connected to at least one of the 13 popular health and fitness tracker tools in the inventory. Connecting to such partner tools, potentially enriches dietary assessment data with addition contextual data such as sleep (identified in 84% of the trackers), steps (62%) and exercise (62%). Fitbit and Jawbone Up are categorized as both, dietary assessment apps as well as popular health and fitness trackers and connected to 14% and 10% of the dietary assessment tools repsectively. Withings Health Mate had with 7% the third highest percentage of connected dietary assessment tools. Misfit, Garmin Connect and Samsung Gear had connections with 4%, 2% and 1% of the tools respectively.

Finally, in 10% the tools in the inventory we identified the option to share (at least part of) the collected food consumption or contextual data with social media platforms such as Twitter or Facebook (e.g., MyNetDiary, FoodSnap!, Lose it!). Again, these platforms potentially offer relevant sources of (unstructured) contextual data related to users' food consumption behaviors.

Aggregators: This connectedness of the investigated food consumption tools is accompanied by an emerging new type of tools, which are the data aggregators or central data hubs such as Apple's HealthKit or Google-Fit (Curtis, 2014; Mandl, Mandel, & Kohane, 2015; Menaspa, 2015; Williams, 2015). HealthKit for instance is a framework designed to integrate healthcare and fitness apps, allowing them to work together and collate their data. For instance, an exercise monitoring app and dietary tracking app that do not offer the option of exchanging data and services through their own infrastructure could alternatively exchange and integrate data via a data aggregator platform. Similar to popular health and fitness trackers, the purpose for adding this group of tools was to investigate its possible connections and data integrations with the group of dietary assessment tools. In our sample

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

of popular food consumption tools 24% of the tools were exchanging data with at least one data aggregator. The tools with the largest number of data aggregators connected were MyfitnessPal and Fatsecret with each connecting to 5 aggregators (e.g., Healthkit, Google Fit, Human Api, Validic). The three aggregators which integrated to the highest numbers of food consumption tools in our inventory were Apple's HealthKit, which integrated with 23% of the tools, followed by Google Fit, which integrated with 10% of the tools, and Samsung's S Health integrating with 3% of the tools in our inventory.

To summarize, in addition to monitoring food consumption data, users are given the opportunity to collect relatively large amounts of various types of contextual data and exchange and integrate it with data and services from various third party systems. There seems to be a clear tendency towards integrating and enriching collected food consumption data with contextual data collected from third party health and fitness trackers, other dietary assessment tools and data aggregators. The strive towards integration and interconnectedness with other services and applications and the sharing of information with connected users and prominent social media platforms opens new potentially interesting sources for a better understanding of the determinants of food consumption behaviors.

2.3.4 POTENTIAL AND LIMITATIONS OF CONTEXTUAL DATA

Potentials:

1) In addition to food consumption data, the dietary assessment tools investigated for the current report also collect various types of relevant contextual data such as physical activity data, physical health data, or social interaction data. The DONE framework identified this types of data as important determinants for food consumption behaviours (Stok et al., 2017).

2) The interconnectedness of tools and platforms opens new opportunities to further enrich the collected dietary assessment data from external sources.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Limitations:

1) There are potentially important gaps of relevant determinants of food consumption such as data related to users' affective states, sleep patterns and the physical location of food consumption.

2) The emphasis of physical health data is mainly on anthropometrics, especially on body weight and BMI. Data on food related physiologies such as intolerances and allergies, or the intake of medications are less prominent.

3) Data about social interactions, food related notes and comments are unstructured data formats. Hence the type and availability of information from these potentially relevant data sources cannot a priori be determined.

4) Since the focus of the investigated tools is mainly on the individual, relevant determinants of food consumption related to the interpersonal, environmental and cultural domain are missing.

3. Discussion and conclusions

Aim of the present deliverable 7.5 was to identify the potentials and limitations of the tools collected in the inventory of deliverable 7.1, to get a better understanding of the determinants of food consumption. For that purpose, we investigated the data collection process of food consumption data by these tools, including its purpose, the applied dietary assessment methodology, the types of nutrients calculated and the possible contextual influences on users' dietary behavior. In addition, in order to get an overview of the contextual data associated with the collected dietary assessment data, we investigated the types of contextual data collected by the tools and the sources for exchanging and integrating contextual data from external sources such as wearables, partner apps and aggregators.

We identified a gap with respect to the availability of publicly accessible data about the tools. Specifically, due to the lack of available legal documents related to the terms and conditions and privacy statements, there is insufficient public information available about the rules users must accept in order to use a service and the ways a vendor gathers, uses, discloses, and manages their users' data. Hence, the legal limitations, organizational

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

restrictions, confidentiality and privacy concerns related to collection, integration and dissemination of this consumer generated data remains relatively under-documented. In addition, we identified a lack of documentation about the procedures for data access. Data accessibility refers to how easy it is to access collected data and metadata (e.g., Dufty, Bérard, Lefranc, & Signore, 2014). Hence there is insufficient documentation about the interactions with the technical infrastructures for data access, as well as about the format of the data and whether the data is retrievable using an open, free, and universally implementable communications protocol. Finally, we also identified a lack of information regarding the procedures for estimating portion sizes in the implemented dietary assessment methods. Since portion size estimations are an important source of error in dietary assessments (Jonnalagadda et al., 1995), a lack of information about the underlying procedures, limits the evaluation of the quality of the collected food consumption data.

The vast majority of tools in the investigated sample collected consumer generated food consumption data at the individual level, on a daily basis, at a certain moment in time and over a certain period of time. The investigated food diaries allowed for inputs from various data sources such as pre-compiled food databases as well as user-generated and individualized databases. This supports an effective, user-friendly collection of food consumption data. It might also provide important insight in the prevalence and variability of individual dietary consumption patterns and habitual food consumption and how they change over time. The quality and usability of modern technology driven dietary assessment methods depends on the completeness and accuracy of the underlying food databases (e.g., Arens-Volland, Spassova, & Bohn, 2015). The compilation and quality maintenance procedures of precompiled and user generated food composition databases, however, remains unknown. Scientific studies investigating the validity of these databases are scarce. Consequently, the quality (and completeness) of the underlying databases remains inconclusive. Hence, conclusions with respect to the relationship between consumed nutrients and energy and the development of nutrition related diseases might be limited. Quality standards and guidelines are needed for food composition data compilation and food consumption data integration into the supporting databases. In addition, research investigating the current state and quality of these databases and their entries are needed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

The large occurrence of tools with the aim on behavioral change, however, is accompanied by various types of potential contextual influences on users' food consumption behaviors. This data might be able to reveal relevant information about the determinants of food consumption behaviors. Controllability and linkeability of this contextual influences seems crucial for a better understanding of influences on food consumption. Another potential consequence of this emphasis on behavioral change, and in particular weight-management, might be reflected in the type of food composition data estimated for the collected food consumption data. The vast majority of tools focused on the estimation of energy and macronutrients. This limited level of detail might be a barrier for research about the associations between specific nutrients (e.g., vitamins) and health outcomes. Similarly, although emerging food photo diaries offer popular and effective ways of logging, visualizing and sharing food consumption data, the added value for detailed food composition estimations of the foods depicted in the images remains limited.

In addition to food consumption data, dietary assessment tools also collect various types of important contextual data. By mapping this contextual data to the DONE framework, we identified several important determinants of food consumption behaviour, such as data related to physical activity, physical health, or social interactions. While the focus of physical health data is mainly on parameters such as body weight or BMI, data on food related physiologies such as intolerances and allergies, or the intake of medications are less prominent. There are also potentially important gaps with respect to relevant determinants of food consumption, such as data related to users' affective states, sleep patterns and the physical location of food consumption.

The lack of contextual data as well as the lack of detail with respect to estimated food composition profiles might be compensated, however, by exchanging data with other tools, such as partner tools and aggregators. On the one hand, this interconnectedness of tools and platforms might open new ways to further enrich the collected dietary assessment data from external sources, and hence enable researchers to study the determinants of food consumption behaviours. On the other hand, however, the integration of this various types of data might be challenging. For instance, in their endeavour to integrate data from Apple's

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

data aggregator tool HealthKit, the technicians at the Open mHealth Platform experienced several problems with data integration. One problem was related to the lack of external data representation for information stored in the Health app. Hence data objects exported from the Health app were not easily portable outside of Apple's devices (Open mHealth, 2015) . This problem might be a general problem related to the various (potentially undocumented) structures and formats of the data collected and exchanged by these tools.

Due to this increasing data network of interconnected dietary assessment and contextual data, there is a need for standards regarding data structures and data formats. Supporting the harmonization of consumer generated data is key for efficient data exchanges within and outside these networks. The emerging networks of consumer generated data provide an interesting opportunity for researchers who want to integrate food consumption data with relevant contextual data. Further research is needed, however, in order to better understand the nature of this data networks, their access points and the types and structures of data they exchange.

We believe our current investigation has several implications for the services RICHFIELDS could provide. The most important role RICHFIELDS could fulfil is in the development and maintenance of standards regarding food consumption data compilation. One purpose of European Food Information Resource (EuroFIR), for instance, is to promote international harmonisation of standards to improve the quality of food composition databases. The adherence to such standards should be promoted towards organizations collecting and exchanging consumer generated data. Not only food composition data requires standardization, but also the consumer generated dietary assessment and contextual data needs to be harmonized. This could enhance the quality and utility of consumer generated data, and the efficiency of data integration, transformation, data analyses and visualizations. In addition, creating quality standards or providing standardized databases for public access, will enable vendors of dietary assessment tools to develop higher quality applications, and will increase the quality of data that users are able to collect about themselves.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Generating valid inferences from the vast amount of diverse data collected, might be challenging for users as well as organizations without a scientific background in behavioural and nutritional sciences. Since the user is involved in both collection and reflection on the data (Li, Dey, & Forlizzi, 2010) another important service RICHFIELDS could provide, is supporting the interpretation and reflection of the collected data. Hence, RICHFIELDS should not only support and improve consumer's collection of data, but should also support insightful reflections on the data, contributing to the pursued gain in self-knowledge.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Literature

Aizawa, K., Maeda, K., Ogawa, M., Sato, Y., Kasamatsu, M., Waki, K., & Takimoto, H. (2014). Comparative Study of the Routine Daily Usability of FoodLog: A Smartphone-based Food Recording Tool Assisted by Image Retrieval. *J Diabetes Sci Technol, 8*(2), 203-208. doi:10.1177/1932296814522745

Ajzen, I. (2015). Consumer attitudes and behavior: the theory of planned behavior applied to food consumption decisions. *Rivista di Economia Agraria. Firenze University Press, 2*, 21-138. doi:10.13128/REA-18003

Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & Management, 39*(6), 467-476. doi:http://dx.doi.org/10.1016/S0378-7206(01)00113-6

Arens-Volland, A. G., Spassova, L., & Bohn, T. (2015). Promising approaches of computer-supported dietary assessment and management-Current research status and available applications. *Int J Med Inform, 84*(12), 997-1008. doi:10.1016/j.ijmedinf.2015.08.006

Carter, M. C., Burley, V. J., Nykjaer, C., & Cade, J. E. (2013). 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. *Br J Nutr, 109*(3), 539-546. doi:10.1017/s0007114512001353

Crispim, C. A., Zimberg, I. Z., dos Reis, B. G., Diniz, R. M., Tufik, S., & de Mello, M. T. (2011). Relationship between food intake and sleep pattern in healthy individuals. *J Clin Sleep Med, 7*(6), 659-664. doi:10.5664/jcsm.1476

Davis, S. F., Ellsworth, M. A., Payne, H. E., Hall, S. M., West, J. H., & Nordhagen, A. L. (2016). Health Behavior Theory in Popular Calorie Counting Apps: A Content Analysis. *JMIR Mhealth Uhealth, 4*(1), e19. doi:10.2196/mhealth.4177

Dufty, D., Bérard, H., Lefranc, S., & Signore, M. (2014). *A Suggested Framework for the Quality of Big Data*. Retrieved from http://www1.unece.org/stat/platform/download/attachments/108102944/Big Data Quality Framework - final- Jan08-2015.pdf?version=1&modificationDate=1420725063663&api=v2

Herman, C. P., Polivy, J., Lank, C. N., & Heatherton, T. F. (1987). Anxiety, hunger, and eating behavior. *J Abnorm Psychol, 96*(3), 264-269.

Hopkins, M., King, N. A., & Blundell, J. E. (2010). Acute and long-term effects of exercise on appetite control: is there any benefit for weight control? *Curr Opin Clin Nutr Metab Care, 13*(6), 635-640. doi:10.1097/MCO.0b013e32833e343b

Jonnalagadda, S. S., Mitchell, D. C., Smiciklas-Wright, H., Kris-Etherton, P. M., Karmally, W., & VanHeel, N. (1995). Portion Size Estimation:: A Source of Error in Diet Assessment Studies. *Journal of the American Dietetic Association, 95*(9, Supplement), A21. doi:http://dx.doi.org/10.1016/S0002-8223(95)00422-X

Li, I., Dey, A., & Forlizzi, J. A. (2010). *A Stage - Based Model of Personal Informatics Systems.* Paper presented at the CHI '10.

Lundahl, A., & Nelson, T. D. (2015). Sleep and food intake: A multisystem review of mechanisms in children and adults. *J Health Psychol, 20*(6), 794-805. doi:10.1177/1359105315573427

Mathiassen, A., & Hollema, S. (2014). What is the effect of physical activity level on food consumption, energy deficiency, and dietary diversity? *Food Nutr Bull* (2015/04/24 ed., Vol. 35, pp. 351-360).

Mela, D. J. (1996). Eating behaviour, food preferences and dietary intake in relation to obesity and body-weight status. *Proc Nutr Soc, 55*(3), 803-816.

Oliver, G., & Wardle, J. (1999). Perceived effects of stress on food choice. *Physiol Behav, 66*(3), 511-515.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Oliver, G., Wardle, J., & Gibson, E. L. (2000). Stress and food choice: a laboratory study. *Psychosom Med, 62*(6), 853-865.

Open mHealth. (2015). 3 painful lessons learned building with HealthKit. Retrieved from http://www.openmhealth.org/3-painful-lessons-learned-building-with-healthkit/

Pagoto, S., Schneider, K., Jojic, M., DeBiasse, M., & Mann, D. (2013). Evidence-based strategies in weight-loss mobile apps. *Am J Prev Med, 45*(5), 576-582. doi:10.1016/j.amepre.2013.04.025

Redd, M., & de Castro, J. M. (1992). Social facilitation of eating: effects of social instruction on food intake. *Physiol Behav, 52*(4), 749-754.

Seguin, R. A., Aggarwal, A., Vermeylen, F., & Drewnowski, A. (2016). Consumption Frequency of Foods Away from Home Linked with Higher Body Mass Index and Lower Fruit and Vegetable Intake among Adults: A Cross-Sectional Study. *J Environ Public Health, 2016*, 3074241. doi:10.1155/2016/3074241

Stok, F. M., Hoffmann, S., Volkert, D., Boeing, H., Ensenauer, R., Stelmach-Mardas, M., . . . Renner, B. (2017). The DONE framework: Creation, evaluation, and updating of an interdisciplinary, dynamic framework 2.0 of determinants of nutrition and eating. *PLoS One, 12*(2), e0171077. doi:10.1371/journal.pone.0171077

Symmank, C., Mai, R., Hoffmann, S., Stok, F. M., Renner, B., Lien, N., & Rohm, H. (2017). Predictors of food decision making: A systematic interdisciplinary mapping (SIM) review. *Appetite, 110*, 25-35. doi:10.1016/j.appet.2016.11.023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.