Horizon 2020
INFRADEV-1-2014 - Design studies

RICHFIELDS Working Package 6
Deliverable D6.4

Paper on quality criteria and overview of criteria applied to available data/methods - WP6

Date delivered:
M38

Authors:
Monique Raats, Sebastian Bartos, Lada Timotijevic, Matthew Peacock, Charo Hodgkins, Morro Touray, Naomi Klepacz, Indira Carr, Haris Hondo, Erik Kaunisto, Anouk Geelen, Marcus Maringer, Anne Norman, Muriel Verain

Deliverable lead beneficiaries:
University of Surrey

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 user practices – WP5-7
Deliverable No.:	D6.4
Authors:	Monique Raats, Sebastian Bartos, Lada Timotijevic, Matthew Peacock, Charo Hodgkins, Morro Touray, Anouk Geelen, Marcus Maringer, Anne Norman, Muriel Verain
Reviewer:	Karin Zimmermann – Project Coordinator Pieter van 't Veer – Scientific Coordinator Bent Engberg Mikkelson – Phase 2 Leader
Start date:	Month 22
Delivery date:	28.11.2018
Due date of deliverable:	30.09.2018
Dissemination level:	Public
Status:	Final

Karin Zimmermann
Project Coordinator

Prof Pieter van 't Veer
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

This deliverable contains two draft manuscripts (PAPERS 1 and 2) for submission to peer-reviewed journals and once accepted be made available as open access. The papers deal with the extent to which there is food-related data that would be of interest to the research community (PAPER 1), and if available whether the public would be willing to share this data with researchers (PAPER 2) and what the key predictors of this would be.

PAPER 1: Understanding the availability of personal food preparation data

Introduction: This paper describes a study that aims to identify food-preparation related data consumers generate through everyday food- related activities, either actively or passively, through the use of tools such as apps and sensors, "outside the research environment". A further goal of the study was to gain insight into the extent of public is currently generating food preparation data generated by any electronic system, including data related to the purchasing and consumption of food.

Method: Based on an agreed definition of domestic food preparation, a protocol was devised for the identification of relevant apps. An internet-based search of the following sources was conducted of the online mobile application stores iTunes Store and Google Play. An additional search was made using the search engines Google and fnd.io. The aim was to identify a range of apps that captures the breadth of apps available. Apps were categorized on an ongoing basis to develop a typology that captured all types of available apps. Apps with similar functions were grouped together. As data collection continued and more apps were logged, new app functions and types data were identified and the typology was adapted, i.e. a dynamic search process was used.

In the survey regarding data generation, we aimed to recruit 1000 participants from each of the following countries: France, Italy, the Netherlands, Slovenia, Spain, Sweden, the UK, and Germany. Stratified sampling was employed in each country, so that an equal number of participants was recruited from the following age groups: 18-29 years; 30-39 years; 40-49 years; 60-59 years; and 60+ years. In each age group, we aimed to recruit an equal number of men and women. All participants were recruited through Lightspeed Research (www.lightspeedresearch.com). Participants filled in an online questionnaire that took about 35 minutes to complete. The instrument contained questions on participants' willingness to share with commercial, academic and government organisations the data that food-related apps generate on their daily habits. Questions also covered participants' reasons for (not) sharing such information, their health, as well as their relevant attitudes and values. Participants were asked what types of food related activities they performed on their computers and smartphones. They could select any number of 24 options, including "search for places to eat and drink", "share views on recipes", and "keep records of what you eat and drink".

Results: The 54 prototypical apps identified were classified. For 48 (89%) of the apps, the motivation for use was classified as 'Knowledge and understanding' with 33 (61%) allowing the user to 'search for information', and 15 (28%) for the user to 'share knowledge and experience'. A further 53 (98%) were classified as having the 'Planning and organisation' as their primary motivation for use, of these 18 (33%) allowed the user to perform 'Recipe management', ten (18%), to perform 'Meal/menu planning' and 25 (46%) to carry out 'Documenting/recording of food'. A further 18 (33%) apps fell into the category of 'Meal preparation and cooking', within this classification, nine (17%) apps were classified as 'Interacting with Sensors', and nine (17%) apps 'classified as using apps as cooking aids.'

A total of 8052 individuals completed the survey. The most commonly generated data is search data, the least common is "location" data generated using smart devices that record when, where or how

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

food is purchased, prepared or eaten. Slovenia was the country with the highest percentage of respondents generating data for 14 of the 24 types of data, followed by Italy, with 9 out of the 24 types of data. Germany generated the least data for 8 of the 24 categories, followed by the UK, with 6 out of the 24 types of data.

Discussion: The primary motivation for consumers for using domestic food preparation apps is to develop personal food knowledge, skills and/or abilities. This opens up the potential to answer questions relating to Individual Psychological determinants, such as food beliefs, habits and self-regulation in relation to food. However, the limited availability of contextual data, such as that at the 'Individual/Situation', and 'Interpersonal/Social' levels, means that much of this data is detached from the user. Researchers intending to use this data will have to carefully consider the degree to which additional contextual information is required to draw conclusions.

The interconnectedness of the apps presents new opportunities to further enrich data collected from external sources, i.e. there is the potential to create 'links' between the multiple apps used by a single user. For example, it may be useful to gain domestic food preparation specific information from dedicated apps, and enrich this with demographic, situational and social context data collected through apps such as Facebook, Twitter and Instagram. However, the degree to which users would find this interlinkage acceptable still needs to be investigated. It should be noted that to date this type of data has been used to study food consumption patterns.

A criterion for inclusion into the inventory was that the app provided sufficient details, so as to facilitate the completion of the majority of these quality criteria. The decision was made that this task should only use information that was publically available. That is, not to contact the company and/or app developer for additional information beyond what could be found from either using the app, the retail store (e.g. Google Play), or an accompanying website. This was in order to test the feasibility of applying the quality criteria with only publically available information. It was found that a large proportion of domestic food preparation apps were developed by small independent app developers and therefore didn't have the same level of accompanying information as those produced by larger companies. Many apps did not have an accompanying websites or application programming interfaces (APIs, the sets of requirements that govern how one application can "speak" to another) and thus access to information, such as terms and condition, was limited.

Many domestic food preparation apps do not collect consumer-generated data, they provide the consumer with information, and indeed some apps only provide the consumer with information. The need for information can be said to be a major motivating factor for the use of an app in the preparation of food. Searching for cooking times, oven temperatures or weight conversions are all food preparation tasks for which apps are commonly available and yet no consumer data is collected by the app. These apps were therefore excluded from inclusion into the inventory.

A further point to consider with user-generated domestic food preparation data, is the degree to which it can act as a 'proxy' for intake. The data collected via app usage reflects the motivation to gain knowledge and to develop skills. The degree to which this is translated into intake cannot be directly drawn from the data in its current form. At best, it describes an 'intention' to intake certain foods and/or meals.

The survey data shows that across the EU, a wide range of food-related data is being generated by consumers, and that there are differences between countries with respect to the kind and range of data produced. Whilst this descriptive information in itself is likely to change over short term due to the rapid advances in technology capable of capturing food-related information across the EU. Search data is the most likely data to be in abundant supply and yet, it holds relatively limited value for the study of the actual behaviour. On the other hand, the data on food intake and the relevant context

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

data (e.g. location) is relatively harder to come by. Nevertheless, it is important to understand that the linking of these different types of data may achieve the granularity necessary for breakthrough science of consumer food-related behaviour

PAPER 2: Determinants of sharing personal food-related data for research: trust, moral motives and perceived risk

Aim: The goal of the study described in this paper is to gain insight into the extent of the public's willingness to share data generated by any electronic systems they may use while choosing, purchasing and preparing food with researchers and for what purposes. In this light, the objectives include exploration of factors and identification of variables which explain reasons for choosing to share data for research purposes.

Methods: Participants were 8450 adults from eight European countries. Specifically, 1000 participants were recruited from each of the following countries: France, Italy, the Netherlands, Slovenia, Spain, Sweden, the UK, and Germany. Countries were selected to vary with regard to the extent to which there was evidence of high health privacy concern, time in the EU, cuisine, health care system type and role in the European "Food, Nutrition and Health Research Infrastructure" currently being developed.

Participants filled in an online questionnaire that took about 35 minutes to complete. The instrument contained questions on participants' willingness to share with commercial, academic and government organisations the data that food-related apps generate on their daily habits. Questions also covered participants' reasons for (not) sharing such information, their health, as well as their relevant attitudes and values. Participants were asked what types of food related activities they performed on their computers and smartphones. They could select any number of 24 options, including "search for places to eat and drink", "share views on recipes", and "keep records of what you eat and drink". A number of questions were asked three times, separately referring to willingness to share with scientists in universities, governments, and companies.

Results and discussion: The examination of people's reported willingness to share data showed an interesting pattern: we recorded an above average willingness to share data (above 3, on a 5-point scale) with universities - for the purpose of science and public research - across all countries. People were simultaneously slightly less willing to share their data with government and industry though this pattern only showed weak statistical significance. Whilst the result is encouraging as it demonstrates the publics' continued belief that science has an intrinsic value as a societal endeavour that deserves publics' support, it also highlights the need to clearly articulate the purpose to which consumer generated data is put and the way in which it links with data from other sources.

Exploring in greater depth this premise that science has an intrinsic value, we captured three important variables: trust, moral motives and perceived risk. Our model that examined the relative weighting of these factors in predicting willingness to share showed that almost half of the variance of willingness to share data is explained by these three variables: trust had a medium-to-large positive effect on the willingness to share data; moral motives had a small-to-medium positive effect; and perceived risk had a small negative effect.

The three constructs are important as they underline the ethical dimension of data sharing decisions and the need for RICHFIELDS data platform to be explicit about its commitment to these values. Trust, perception of risk and moral motives are closely linked with the issue of data governance and the respect for privacy, confidentiality and consent. Data linkages that RICHFIELDS is proposing, would typically enable identification of a consumer, even if we strive to ensure anonymity and de-personalisation. Against this context it is important to address how more value from data can be

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

extracted without compromising the citizen's right to privacy (recognised by European Conventions of Human Rights), confidentiality and the role of consent within the matrix of big data and privacy whilst keeping in sight the protection imparted to the individual (data subject) by EU's General Data Protection Regulation (Regulation 2016/679) that came into force in May 2018.

These concepts also are relevant within the broader debate about how to manage competing interests of science and data donors/subjects (citizens). These are currently resolved through a combination of standard operation procedures and good scientific practices guidelines, creation of ethics advisory board, and regulation of financial gains from the data/IPR ownership.

However, big lifestyle-related data research infrastructures are not only research resources but also provide valuable opportunities in terms of 'new economy' (i.e. employment, entrepreneurship, knowledge creation). The question of who owns lifestyle-related big data and has access to the linked data, therefore, is simultaneously an issue of economic development and international standing, as well as research. Fairness, legitimacy and due process are important considerations integral to any decisions about ownership and commercialization. Coupled with this is the ethical issue of broader societal value of who has the right to commodify the information based on linked lifestyle-related data and if it should be rightfully 'owned' by anyone, shifting the discussion away from economic sphere, towards the human rights domain (and the associated legal frameworks).

In the light of these findings, the possible considerations for the future of data-driven science need to address the following issues for the purpose of developing RICHFIELDS data platform:

- The research infrastructure that wishes to make use of consumer-generated data will to identify appropriate means of maintaining trust, minimising risk to individual and society and enhancing the perceived moral authority of science. In a nutshell, these endeavours could be achieved through appropriate governance and technical frameworks, but perhaps more importantly, through the engagement with public and constant communication that would develop a strong moral identity for the research infrastructure in this domain.
- The research infrastructure should closely observe the recommendations of how to achieve ethical design for the future. Given RICHFIELDS' purpose is to use the data sets in its repository for research, pseudonymisation is suggested as a means to process the data, provided appropriate safeguards are in place. In order to raise the integrity profile of RICHFIELDS externally the setting up of an independent ethics committee is also suggested. This would also help in bolstering the confidence of data in the utility of research infrastructures such as RICHFIELDS as a research tool for promoting well-being and over time might usher in an era where data subjects in the spirit of altruism give their data for the sake of research and innovation.
- A more nuanced understanding of the way in which the public's perceive the possible solutions and models for RICHFIELDS needs to be obtained through the use case studies and validation of our business, governance and technical models.
- The research infrastructure must be mindful of possible cross-country differences in sensitivities about the issue of data sharing. This necessitates constant monitoring of public attitudes to privacy, science and their perceptions of the food system. The latter is particularly apposite in the context of the publics' increasing awareness of the unsustainability of the current food system, and the growing calls for science to engage ethically with the food-related issues that are fundamental not only to the health of individuals, but also to the health and livelihood of our planet.

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

CONTENTS

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

PREFACE

This deliverable contains two draft manuscripts (PAPERS 1 and 2) for submission to peer-reviewed journals:

- PAPER 1: Understanding the availability of personal food preparation data
- PAPER 2: Determinants of sharing personal food-related data for research: trust, moral motives and perceived risk

The manuscripts will be submitted to peer-reviewed journals and once accepted be made available as open access.

PAPER 1 describes a study that aims to identify food-preparation related data consumers generate through everyday food- related activities, either actively or passively, through the use of tools such as apps and sensors, "outside the research environment". A further goal of the study was to gain insight into the extent of public is currently generating food preparation data generated by any electronic systems and including data related to the purchasing and consumption of food.

The goal of the study described in **PAPER 2** is to gain insight into the extent of public's willingness to share data generated by any electronic systems they may use while choosing, purchasing and preparing food with researchers and for what purposes. In this light, the objectives include exploration of factors and identification of variables which explain reasons for choosing to share data for research purposes. This is necessary because little empirical research exists on the extent to which members of the public are consciously willing to share personal data with scientists outside the context of medicine, where data generated through interactions with healthcare professionals, for example medical records, are sometimes used in an aggregated form for research into specific conditions for the purposes of improving treatment or provision of care. In particular there is a dearth of large scale quantitative research comparing data from a range of countries with diverse cultures and disparate socioeconomic backgrounds who may have different attitudes towards healthy eating and use technology in different ways.

ACKOWLEDGEMENTS

The authors would like to thank colleagues from the RICHFIELDS partner institutions for their support with translation of the questionnaire:

- German Institute of Food Technologies (DIL), DE: German questionnaire
- Institute of Industrial Technologies and Automation - National Research Council of Italy (ITIA-CNR), IT: Italian questionnaire
- Javier de la Cueva, ES: Spanish questionnaire
- Jožef Stefan Institute, SI: Slovenian questionnaire
- RISE Research Institutes, SE: Swedish questionnaire
- University of Surrey, UK: Dutch, Italian, Spanish and French questionnaires
- Wageningen University, NL: Dutch questionnaire

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

PAPER 1: 1. INTRODUCTION

Food skills can be defined in terms of two behavioural components. These are, Planning and Organisation behaviours, and Food Knowledge and Understanding behaviours. Both can be considered as necessary antecedents to the mechanical preparation and/or cooking of foods [Fordyce-Voorham, 2009]. Planning and Organisation are skills are that are reflective of the decision-making process involved in domestic food preparation [McGowan et al., in press]. They may include behaviours such as; (1) Documenting and recording food (e.g. making of shopping lists, or recording of expiration dates for perishable food items); (3) Meal and/or menu planning (e.g. the planning of an individual meal or a series of meals both in terms of menu choice and the timing of individual meals over varying time periods (e.g. days, weeks, months etc.); (3) Recipe management, e.g. the collecting and categorising of recipes for future use.

Knowledge and understanding are skills that reflect a person's need for information relevant to intended preparation behaviour or the reflection on a previously carried out behaviour (Stead et al., 2004). Such skill may include: (1) Sharing knowledge and experience. For example, bookmarking or favouring information within an app for the intention of future use, and/or the reading or writing of reviews and sharing of knowledge and experiences via social media; and (2) Searching for information. For example, searching for knowledge that will assist with future food preparation behaviours, such as searching recipe databases and/or understanding terminology associated with food preparation.

This study aims to identify food-related data consumers generate through everyday food- related activities, either actively or passively, through the use of tools such as apps and sensors, "outside the research environment". The large-scale generation of such data has the potential to be able to provide data for use in research thus providing insights regarding food choice or determinants thereof. Food choice operates at physical, biological, psychological, and sociocultural levels (Sobal, 1991), all which operate simultaneously and interact (Sobal et al., 2014) and include the acquisition, preparation, serving, eating, storage, giving away of and cleaning up of food (Sobal and Bisogni, 2009). Whilst food choice behaviour is seemingly simple, it is in fact very complicated behaviour that is influenced by many interacting factors that each belong to the traditional domains of one of a large diversity of scientific disciplines and as a result each of these disciplines claims to have at least a partial answer to the central question in food choice research: ''Why does who eat what, when, and where?" (Köster, 2009). A further goal of the study was to gain insight into the extent of public is currently generating food preparation data generated by any electronic systems and including data related to the purchasing and consumption of food.

PAPER 1: 2. METHODS

PAPER 1: 2.1 INCLUSION AND EXCLUSION CRITERIA

Applications – or apps – can be defined as software programmes developed specifically for use on small, wireless computing devices – such as a smartphone or tablet. In this instance, the search was restricted to those applications available on the IOS and/or Android platforms. It is possible to apply this search protocol to the identification of prototypical examples of websites and other similar ICT technology. However, due to the volume of apps available in the marketplace, it was decided that in the first instance the search should be restricted to apps only.

In order for an app to be included it had to be:

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

- able to collect consumer-generated data on domestic food preparation at a person, household and/or population level according to the definition of domestic food preparation as set out above;
- not be a clinical or medical;
- available for use by the wider population (e.g. research only apps were excluded);
- currently available to users, or due to be released shortly (Beta testing);
- have sufficient information publicly available to enable the completion of the majority of inventory fields;
- not be specifically aimed at use by children, such as games;
- not be aimed at the preparation of food for children and infants (e.g. breastfeeding and weaning);
- in English language; and
- In the case of Google Play and Itunes Stores, available for download through the UK storefront.

PAPER 1: 2.2 SEARCH PROTOCOL

Based on the definition of domestic food preparation set down in the above section, a protocol was devised for the identification of relevant apps. An internet-based search of the following sources was conducted of the online mobile application stores iTunes Store and Google Play. An additional search was made using the search engines Google and fnd.io. This search was restricted to UK store fronts. Additional apps were identified from reference lists of searched articles, company newsletters, and/or were identified by the research team, so as to identify apps that are not yet on the marketplace. All searches were undertaken by a single researcher and were conducted during the period April 2016 and September 2016.

An initial search of these sources was made using the search terms; food, nutrition, lifestyle and behaviour, as this would allow for the capture of the most comprehensive range of relevant apps, and would avoid the search being restricted by pre-entered search terms or keywords. However, in order to refine the list of apps generated by these searches, a further search was made using the following terms, and their synonyms, specific to domestic food preparation: "food", "food preparation", "cooking", "food production", "diet", "shopping list", "food diary", "food and drink", "food glossary", "meal planner", "cooking skills", "kitchen", "smart kitchen device", "dinner kit" and "smart food scale".

The purpose of this exercise was to identify the scope of domestic food preparation apps available. The aim was to identify a range of apps that captures the breadth of apps available. Apps were categorized on an ongoing basis to develop a typology that captured all types of available apps. Apps with similar functions were grouped together. As data collection continued and more apps were logged, new app functions and types data were identified and the typology was adapted, i.e. a dynamic search process was used. The nature of the consumer-generated data collected by each app generated by the search was noted and a single app was identified that typified that particular consumer-generated data was included into the inventory. The search continued until no new types of apps were identified. In addition, inclusion into the inventory was based on the inclusion and exclusion criteria set down in the following section.

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

PAPER 1: 2.3 APP CHARACTERISATION

Due to the lack of knowledge with respect to the quality of current ICT tools and the data they procedure, great efforts have been made with regards to the development of frameworks and guidelines for the evaluation of such applications (e.g. Brown et al., 2013; Kumar et al., 2013; Meulendijk et al., 2014; Stoyanov et al., 2015). Private, as well as public, companies and institutions now offer guidelines, services and infrastructures for reviewing, evaluating and certifying health applications.

A condition for inclusion on to the final list of quality criteria was the availability of information. That is to say, the questions regarding a tool's description, scientific, technical and legal/ethical characteristics had to be answerable from the information available through sources such as an 'app store' (e.g. ITunes/Google Play) or a tool's homepage. Any criteria that necessitated, for example, the downloading and testing of a tool, the examination of a tool's data structure and/or the examination of a hosting data infrastructure were discarded. The apps were characterised according to four types of quality criteria [1] descriptive criteria, [2] scientific criteria, [3] technical criteria and [4] legal/ethical criteria. An overview of the quality criteria used to characterise the tools is provided in Table 1.

PAPER 1: Table 1. *Overview of the quality criteria used to characterise the tools.*

Descriptive Criteria *What is it?*	Scientific Criteria *Is it useful?*	Technical Criteria *Can we access it?*	Legal/Ethical Criteria *Can we use it?*
• Data Types • Home page • Contact Information • Supported platforms • Paid Services • Medical Device • Preparation Categories • Price of IOS app • Languages • Itunes user rating • Itune Genre • Current IOS apps • Minimum Android version	• Lifestyle Data • Situational Characteristics • Types of Situational Characteristics • Product Characteristics • External Device • Data integration with partner tools • What was purchased/ prepared/consumed? • What was purchased/ prepared/consumed? • What was prepared? • Act or Intention? • Units of purchase/ preparation/ consumption?	• Is data accessible? • Types of Access • Data Formats • Authentication • Price • Amount	• Terms of use • Privacy Policy • Data ownership • Data usage vendor • Personal information • Types of personal information • Public profile • Privacy settings • Device data • Types of device data • Cookies • Web beacons • Data storage

Quality criteria necessary for the assessment of the technical governance of consumer-generated food preparation data were identified. These criteria reflect the now widely accepted and recommend FAIR data principle (see Wilkinson et al., 2016). However, for the benefit of this exercise, focus remained on those FAIR data principles that did not require us to examine the data structure of the tool, or data access documentation in detail. The Technical Criteria therefore focus around accessibility of data.

Accessibility of data refers to how easy it is to retrieve data and metadata (e.g. Dufty et al., 2014) including the technical infrastructure (e.g. application programming interfaces (APIs, the sets of requirements that govern how one application can "speak" to another)) for data access (e.g. Dedeke,

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

2000). Also, whether data is retrievable using an open, free and university implementable communications protocol (e.g. REST) and is represented in a formal, accessible, shared and broadly applicable language (e.g. Wilkinson et al., 2016). In addition to standardised data access, the protocol should also allow for an authentication and authorisation procedure (e.g. Wilkinson et al., 2016).

This criteria 'is data accessible' seeks to answer the question 'is the data collected by the tool accessible directly via the tool's infrastructure (and not via integrated aggregators, i.e. other tools)? Further accessibility criteria aim to identify whether the tool has any accompanying access documentation, and whether there is a URL to this documentation. The criteria also aims to identify whether the tool has documentation concerning the terms under which the data can be accessed and whether there is a URL to this documentation that users can access. Furthermore, it is an important indicator of data quality that the data can actually be accessed and the form that this access to take (e.g. Email export, web feed, web API).

Consent is a key issue for consumer trust, indeed perceived lack of consent due to data acquisition and usage may undermine public trust. Furthermore, there is a requirement that all tools cover data ownership and data privacy in their licensing agreement, which the consumer accepts at the time of initial use (e.g. Cummings et al., 2013; Adhikari, Richards and Scott, 2014; Blenner et al., 2016). To this end, quality criteria identifying the 'terms of use' of the tool, and the source of this information, together with information regarding the tool's privacy policy were included in the list of legal quality criteria.

Another factor of relevance to data quality is that of data privacy. Data privacy can be defined as the disclosure of all data that a tool - or, other within-tool advertisers – collects or accesses via consumer devices and the applied methods and technologies (e.g. Boulos et al., 2014). This includes, the collection, storage, and network transmission of user generated data, including personal identifiable data and whether the data is securely encrypted during and after those workflows (e.g. Njie, 2013), and the duration and termination of data storage (e.g. Cummings et al., 2013).

Furthermore, data privacy may also refer to the (secondary) usage of the user generated data, such as making data accessible to the general public or sharing data with other affiliated or unaffiliated third-parties, such as analytics and advertising services, or data brokers (e.g. Cummings et al., 2013). Issues surrounding data privacy were deemed of particular relevance by the consortium. Therefore, legal quality criteria relating to data privacy were included on the list, covering issues surrounding the collection of both 'personal identifiable information' about the consumer and also data about the device the consumer is using to access the tool. Other criteria, focus on the storage and sharing of this information, such as with an affiliated or third party. Criteria also cover the consumer's use of homepages/websites and usage trackers such as cookies (data sent from the homepage/website to monitor usage) and web beacons (information embedded in, for example, emails that monitor whether a consumer has accessed particular content).

Data ownership concerns both the possession of and responsibility for information. Ownership implies power as well as control. The control of information includes not just the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to others (Loshin, 2002). Loshin (2002) identifies a list of parties laying a potential claim to data, such as the party that creates or generates the data (e.g. the app user), the enterprise in which the data is created (e.g. the app vendor) or the individual or organisation that buys or licenses data (e.g. third parties and business partners). Both data privacy and ownership may have a significant influence on the intended use of the data given legal limitations, organisational restrictions, and confidentiality and privacy concerns. Legal quality criteria have been included that aim to establish the owner of the consumer-generated data and whether the vendor has the right to access and exploit

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

this data by publishing, distributing, and otherwise publically displaying this data is either its original or another form.

A further criterion of relevance is that of data security. Data security refers to the extent to which access to information is restricted appropriately to maintain its security (e.g. by authentication; e.g. Knight and Cowan, 2005; Schulze and Kromker, 2010; Martinez-Perez et al., 2013). Data security may be assessed on several levels, such as the data level, application level, network level and host level (e.g. Ho et al., 2013). In addition, data security can refer to the storage of data, for example local storage as opposed to cloud-based storage or a 'backup' data system (e.g. Ho et al., 2013). To this end, the quality criteria aim to establish whether the consumer-generated data is securely stored on either a storage device or a web server storage.

PAPER 1: 2.4 GENERATION OF FOOD PREPARATION-RELATED DATA

Participants were 8450 adults from eight European countries. Specifically, we aimed to recruit 1000 participants from each of the following countries: France, Italy, the Netherlands, Slovenia, Spain, Sweden, the UK, and Germany. Stratified sampling was employed in each country, so that an equal number of participants was recruited from the following age groups: 18-29 years; 30-39 years; 40-49 years; 60-59 years; and 60+ years. In each age group, we aimed to recruit an equal number of men and women. All participants were recruited through Lightspeed Research (www.lightspeedresearch.com).

Participants filled in an online questionnaire that took about 35 minutes to complete. The instrument contained questions on participants' willingness to share with commercial, academic and government organisations the data that food-related apps generate on their daily habits. Questions also covered participants' reasons for (not) sharing such information, their health, as well as their relevant attitudes and values. Participants were asked what types of food related activities they performed on their computers and smartphones. They could select any number of 24 options, including "search for places to eat and drink", "share views on recipes", and "keep records of what you eat and drink".

Participants' use of health apps was assessed with a single item (Ernsting et al., 2017). They were asked to list the purposes, if any, for which they had used smartphone apps over the previous 12 months. Participants could tick any number of seven options (e.g. to quit smoking; to lose weight). Since a large number of participants (47 to 70%, depending on the country) used no health apps, the item was dichotomised (0 – has not used health apps; 1 – has used health apps).

Basic **demographic** data were also collected. These included gender, age, height, weight, employment status, income, highest educational level (from 1 – no formal education to 9 – university degree), and household composition. Based on height and weight, we also computed each participant's body mass index. Because 88% of participants did not report their income, we did not use this variable. Since 99.6% of the participants identified as either male or female (as opposed to identifying with another gender, 0.2%, or refusing to answer, 0.2%), we only used the data from men and women when exploring gender issues.

Food practices were assessed with a series of brief questions. First, participants were asked how many minutes they spent cooking on a typical weekday and on a typical weekend day. From these data, we estimated each participant's average cooking time per week (5*typical weekday +2* typical weekend day). We then explored whether participants had the responsibility for shopping (Raats et al., 2015; Hieke et al., 2016) and for cooking (Lavelle et al., 2016; McGowan et al., 2016; Lavelle et al., 2017)) respectively, within their households. Responsibility for shopping was assessed on a three-point Likert

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

scale ranging from "no" through "shared" to "yes". Finally, we asked participants how often in a week they had a takeaway, a ready meal, and pub/restaurant meal, respectively. These questions were answered on a Likert scale ranging from every day (1) to never (6). This was later reverse-coded so that larger scores indicate eating more of these meals.

The study was conducted according to the guidelines laid down in the Declaration of Helsinki and in accordance with the University of Surrey's ethical procedures. Participants provided informed consent, then answered the questions in the order provided above. The survey was administered via Qualtrics™. The survey was developed in English and then translated, checked by native speakers and put into Qualtrics™. Data collection for each country was run separately. Should other researchers wish to conduct a comparable study, advice can be sought from the authors with regard to translation procedures, question selection, dataset preparation and analytic strategy.

Summary scores for the measures above were computed in each of the eight country-level datasets, and incomplete cases were deleted. All variables of interest (including demographics) were aggregated into a master dataset. The master dataset was saved both in a wide and a long format. Descriptive statistics and country comparisons were then computed. This stage of data processing was performed in IBM SPSS 24 and 25.

PAPER 1: 3. RESULTS

PAPER 1: 3.1 A TYPOLOGY OF DOMESTIC FOOD PREPARATION APPS

The focus of the research was on domestic food preparation, i.e. food prepared for one's own consumption, or that of close others (e.g. family members), in the home or another non-commercial environment. A typology of domestic food preparation apps was constructed based the definition of domestic food preparation (Figure 1). Level 1 of this typology reflects the specific domain of interest, that is, domestic food preparation. The second level classifies the domain into three constructs; planning and organisation (food skills), knowledge and understanding (food skills), and meal preparation/cooking (cooking skills). These constructs are said to be the 'antecedents' preceding the act of using an app. The results suggest that the primary motivation for using domestic food preparation apps was to develop one's food knowledge, skills and abilities.

The second level reflects the motivation underlying the behaviour captured by the app (e.g. to gain knowledge and/or understanding). The third level reflects the specific behaviours captured by the app (e.g. searching for information). The second and third levels of this typology reflect the definition of domestic food preparation. The final level of the typology is indicative of the consumer-generated data collect by the selected prototypical apps (e.g. the specific search term used).

The 54 prototypical apps identified were classified according to this typology, these classifications can be seen in Table 2. For 48 (89%) of the apps, the motivation for use was classified as 'Knowledge and understanding' with 33 (61%) allowing the user to 'search for information', and 15 (28%) for the user to 'share knowledge and experience'. A further 53 (98%) were classified as having the 'Planning and organisation' as their primary motivation for use, of these 18 (33%) allowed the user to perform 'Recipe management', ten (18%), to perform 'Meal/menu planning' and 25 (46%) to carry out 'Documenting/recording of food'. A further 18 (33%) apps fell into the category of 'Meal preparation and cooking', within this classification, nine (17%) apps were classified as 'Interacting with Sensors', and nine (17%) apps 'classified as using apps as cooking aids.'

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

Level 1: What is the activity domain?	Domestic food preparation						
Level 2: What is the user aiming to do?	Planning and organisation (food skills)			Knowledge and understanding (food skills)		Meal preparation/cooking (cooking skills)	
Level 3: What is the user doing?	Documenting / recording food	Meal/ menu planning	Recipe management	Sharing knowledge and experience	Searching for information	Using apps as cooking aids	Interacting with sensors
Level 4: What is the recordable user activity?	e.g. shopping lists, pantry lists, fridge contents lists, expiration dates	e.g. meal plans (including daily, weekly, monthly plans); meal choices	e.g. recipe collections; user inputted recipes	e.g. 'favouriting'; bookmarking; reviews; ratings; sharing via social media	e.g. free search of recipe database, ingredient database; glossary terms; filtered searches (inc. meal types, special diet)	e.g. setting timers, measures and conversions	e.g. 'smart' kitchen equipment and appliances

PAPER 1: Figure 1. *Typology of domestic food preparation.*

Apps in in the category 'Meal preparation and cooking' represented by far the smallest proportion of apps currently available in the market place. Although, it is worth bearing in mind that the aim of the task was to identify the range – or variance – of apps currently available within this domain, rather than the depth of apps available in any one category. However, in contrast to the two food skills categories ('Knowledge and understanding' and 'Meal preparation and cooking'), the motivation for using 'Meal preparation and cooking' apps was to assist directly with the cooking – or physical food preparation – process. As such, this category includes apps, such as timers and also 'Smart' technologies found in the connection home. There are several possible explanations for this underrepresentation. One is that the partner technology – such as connected fridges or scales – is in its infancy, and so many people do not own or have access to these technologies – thus they do not require an app. A further explanation is that many apps identified in this category do not collect user-generated data. That is, users are making use of these apps to assist with their cooking skills, such as an egg timer or temperature conversion app, but the apps themselves are not directly collecting any user-generated information and thus were excluded. In short, the user is using these apps in a similar way to a traditional stopwatch or book. Finally, users may simply just be motivated primarily to use apps in the pre-preparation process, rather than for the actual preparation of food stuffs.

The user-generated data is represented in the typology at Level 3. Analysis of the content of the apps has allowed for the identification of seven behavioural constructs. For example, a person's motivation may be to develop their planning and organisational skills, the app allows them to achieve this by providing a function for documenting and/or recording foods. This may be in the form of writing a grocery list of foods to purchase, or recoding expiration dates of foods already purchased. The resulting user-generated activity or 'data', is therefore a list of food items. This is conceptualised in Level 4 of the typology.

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

PAPER 1: Table 2. *The classification of domestic food preparation apps by motivation and behaviour.*

Tool name	Knowledge and understanding (food skills)		Meal preparation and cooking (food skills)		Planning and organisation (cooking skills)		
	Searching for information	Sharing knowledge and experience	Interacting with sensors	Using apps as cooking aids	Recipe management	Meal/ menu planning	Documenting / recording food
8500 Drink and Cocktail Recipes	X				X		
Allrecipes Dinner Spinner		X					X
AnyList		X			X		X
Avocado Meal Planner		X			X	X	
BBC Good Food	X			X	X	X	X
BigOven 350,000+ Recipes and Grocery List	X				X	X	X
Change4Life Smart Recipes	X	X			X		X
Cocktail Making	x	x					
Cook With MandS	X	X			X		
Chronometer							X
Culinary Fundamentals – Cooking School	X						
Culinary Herbs and Spices	X	X			X		
Drinks and Cocktails	X	X					
Drop Recipes	X		X				
Epicurious	X			X			X
Escali SmartConnect			X				
Fat Flush Diet Plan and Meal Tracker	X						X
Fit Men Cook – Healthy Recipes	X					X	X
Food Science 101	X						
Food Intolerances	X						
Forage – free food from the wild	X		X				X
FridgePal			X				X
Glossary of Food Science Terms	X						
Grocery List						x	
HelloFresh	X	X				X	X
Jamie's Recipes	X				X		X

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

Tool name	Knowledge and understanding (food skills)		Meal preparation and cooking (food skills)		Planning and organisation (cooking skills)		
	Searching for information	Sharing knowledge and experience	Interacting with sensors	Using apps as cooking aids	Recipe management	Meal/ menu planning	Documenting / recording food
Kitchen Calculator PRO	X			X			
Kitchen Units: Unit conversion calculator				X			
KitchenPad Timer				X			
Let's Cook – Meal Preparation Timer				X			
LG Smart Range	X		X				
Lose It!	X	X				X	X
MealBoard Meal and Grocery Planner					X	X	X
Meal Planner Pal						X	
My Recipe Book	X	X			X		X
Oh She Glows	X				X		X
Paleo Food List	X						
Pantelligent			X				
Paprika Recipe Manager	X				X	X	X
Prep Pad for iPhone	X	X	X	X	X		
Recipe, Menu and Cooking Planner		X			X	X	
SITU Scale							X
Smart Diet Scale							X
Substitutions	X	X					
Tesco Groceries	X						
The Monash University Low FODMAP Diet	X						X
The Perfect Boiled Egg		X					
The perfect egg timer			X	X			
Time to Roast				X			
Top Chef University	X						
Vitamins Glossary	X	X			X		X
What's In My Fridge							X
Whole Foods Market	X				X		X
Yummly Recipes	X				X		X

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

PAPER 1: 3.2 USER-GENERATED FOOD PREPARATION DATA AS DETERMINANTS OF FOOD BEHAVIOUR

In their paper, Sobal and Bisogni (2009) propose a staged model of the processes involved in food decision-making. They put forward the stages as the 'acquisition', 'preparation', 'serving', and 'eating' of food stuffs. They further suggest that additional decisions need to be made surrounding the storage, giving away and throwing away of food. However, food decisions are not simply related to 'food stuffs' and in this respect they reflect the work of Bisogni et al., (2007) who advocate that food decision are dependent on a range of situational factors, such as location, social interactions, time of day and other actions. The decision is therefore not just 'what' am I going to eat, but with 'whom', 'where', 'how' and 'why'.

PAPER 1: Table 2. *User-generated domestic food preparation data types categorised by the DONE Framework of determinants of nutrition and eating.*

Broad categories of derminants	Sub-categories of derminants	User-generated Data Types
Individual	Biological	Exercise [2]; IBS symptoms [1]; Body measurement [1]; weight goals [1]; body weight [1]; BMI [1]; Body composition [1]; Biometrics [1].
	Demographic	Email address [12]; Home address [8]; Name [7]; Phone number [7]; Financial information [6]; username and/or password [6]; photo/self-select image [2]; Date of birth [2]; Gender [2]; Postcode [2]; Delivery Address [2]; Location [2]; personal video [2]; social network handle [1]; online interactions [1]; IP address [9]; Device location [1].
	Psychological	Notifications [4]; Reminders [1]; cooking advice and instructions [7]; Database search [25]; shopping list [19]; favourite recipes [14]; Favourite food item [3]; filtered search terms [14]; eating patterns [3]; cooking technique/skills [1]; recipe directions [21]; food preferences [7]; Diet plans [2]; personal notes [5]; Meal Planning [8]; Recipe Management [8]; import recipes [5]; list of fridge items [3]; create pantry list [1]; saved Searches [1]; list of expire dates [1].
	Situational	
Interpersonal	Social	Social media network [10]; social media shares/emails [12]; Food Photo [2]; Posts [4]; Comments [4]; Recipe reviews [2]; Share experience via social media [2].
	Cultural	Cuisine [5]; Dish [2]; Occasion [44].
Interpersonal	Product	Ingredients [19]; product weight [5]; product volume [4]; visual properties [3]; brand name [3]; energy content [2]; food [2]; special diet [2]; allergy information [2]; availability [1]; storage conditions [2]; price [1]; food group [1]; vitamins [1]; Food description [4]; cooking temperature [4]; unit of measurement [2]; macro nutrient [2]; micro nutrient [2]; Enter food/ingredient characteristics [6].
	Micro	Geo Coordinates [1]; physical environment (other)[1]; Domestic Kitchen [1]; smart scales [5]; stored in fridge [1]; smart oven [1], Smart refrigerator [1]; GPS data [3]; Select Oven type [3]; Set timer [2].
	Meso/macro	Physical environment [1]; venue name [1]; Altitude [1].
Policy	Industry	
	Government	

Numbers in [ ] represent the number of apps capturing that user-generated data type.

These drivers are unlikely to be static, rather they are driving choice and behaviour only at the current moment. Recently, an adept at the creation of a dynamic and interactive framework of determinants of nutrition and eating has been made. The DONE (Determinants Of Nutrition and Eating) framework has arisen out of work carried out by the DEterminants of Diet and Physical Activity (DEDIPAC)

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

knowledge hub (Stok et al., 2017). The DONE framework identify determinants as falling into four broad categories, Individual, Interpersonal, Environmental, and Policy (Stok et al., 2017). Each of these categories have multiple sub-levels. Table 2 details the overlay of user-generated domestic food preparation data types onto the DONE framework, so as to identify the potential and limitation of these data types for answering questions relating to determinants of nutrition and eating.

The typology of domestic food preparation apps suggests that the primary motivator to engage with an app is to develop food knowledge, skills and/or abilities. The DONE framework (Stok et al., 2017) places these determinants at the level of 'Individual' and 'Psychological', thus the majority of user-generated domestic food preparation data types collect data at this level. Some examples of data types collected at this level include; 'meal-planning', 'recipe management', 'shopping lists' and 'databases searches'. Further examples of user-generated data collected at the 'Individual' are those categorised as 'Demographic' data types. Such data types include, 'email addresses', 'home addresses', 'data of birth, gender'. Individual Biological level data is also generated through app use, this increase details about 'body weight', 'BMI', 'Body composition', but also general health conditions, such as IBS. It should be noted that the recording of personal biological characteristics moves away from the primary motivation for using these apps – to develop food knowledge, skills and/or abilities. However, this 'gap' in user-generated individual and biological level data may be filled by information derived from 'consumption apps' of which the primary motivation for using an app is to record food intake. There is the potential for researchers to user-generated data from multiple app sources to create a picture of consumer food choice and eating behaviour.

Of the 54 prototypical app examples for domestic food preparation, not one generated user data relating to an individual's situation. As defined in the DONE framework (Stok et al., 2017), the determinants that would fall under the category 'Individual/Situational' relate to factors that impose constraints on an individual's consumption (e.g. access of a car, workload) and also wider health behaviours relating to eating. It is a key limitation of user-generated data, that it potentially tells you little about the individual's situation. It may be possible to derive inferences as to an individual's situation through the analysis of other information. For example, their meal plans may give you some indication as to their ability to access food, or the time they have available for food preparation. However, these are merely guesses on the part of the researcher. Again, as this classification relates largely to consumption, it may be that user-documented food consumption data (see Maringer et al., 2018) would give a better indication of an individual's personal situation.

There is a similar issue for user generated 'Interpersonal' data. Specifically, data at the 'Interpersonal/Social' level. The prototypical apps collect data related to social media use and an individual's interaction in an online environment (e.g. social media network, social media shares/emails, posts, comments, reviews). However, no information is collected as to an individual's family structure, or the socio-economic status of a household. Again, analysis of certain aspects of the user-generated data – such as meal plans – may reveal information relating to the make-up of a household or its socio-economic status. The apps do however collect some user-generated data potentially relating to cultural food customs. For example, data is collected on the type of cuisine, the dish and occasion – that is, whether the food is being prepared for an event such as Christmas, Easter, a birthday, or a drinks party.

The Environmental category at the product level is well represented. For example, user-generated data is collected about ingredients, product weight, product volume, nutritional value. However, again data relating directly to the micro environment – home environment – is limited. The DONE framework (Stok et al., 2017) would suggest that determinants in this area relate to the availability and accessibility of food in the home, the meal environment and portion size. It is possible that some of this information may be derived from the analysis of other data. None of the apps analysed collect

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

policy data, whether in relation to industry (e.g. reference to industry regulations such as nutrition composition regulations or portion size regulations) or government (e.g. governmental regulations, food and eating related campaigns). This may limit researchers in drawing conclusions as to the influence of regulation on food choice and eating behaviour.

PAPER 1: 3.3 AVAILABILITY AND ACCESSIBILITY OF INFORMATION

The data used in this analysis were derived from publically available information about the app. In most instances this was the companion website to the app, although information was also sought from the app store's metadata, terms and conditions documents and privacy statements. It was considered an important constituent of the exercise to discover the extent to which information about the apps could be derived from publically available sources and thus in instances where the quality criteria could not effectively be answered it was decided not to seek additional information directly from the company or app developer. In short, the public availability of the information is in and of itself an important quality criterion. However, only 32 (64%) of apps were found to have a companion homepage/website. Of those that did have a homepage/website, contact information was provided for 23 (71%) of the apps. For those apps that did have a website, 18 (56%) provided 'terms of use' documentation and 23 (71%) provided a 'privacy policy'. Of those apps that provided details regarding terms of use, nine (50%) indicated that the data was owned by the vendor and three (16%) by the users.

There are further gaps in the availability and accessibility of information. For example, for on four (8%) of apps was data collected by the app directly accessible via the apps existing infrastructure. For 11 (22%) the data was not accessibility, but for the majority of apps, 35 (70%), no information was publically available regarding the accessibility of the data.

PAPER 1: 3.4 GENERATION OF FOOD PREPARATION-RELATED DATA

A total of 8052 individuals completed the survey. Demographic statistics and descriptors that characterize this cohort are provided in Table 3. Details of the nature and degree of generating food-related data in Table 4 and Figures 2-6. As shown, the targeted sampling strategy was effective at recruiting a sample of individuals that met the stratification targets. Most respondents were responsible for food shopping. Variability across countries is apparent in terms of the types of data generated. The most commonly generated data is search data, the least common is "location" data generated using smart devices that record when, where or how food is purchased, prepared or eaten. Slovenia was the country with the highest percentage of respondents generating data for 14 of the 24 types of data, followed by Italy, with 9 out of the 24 types of data. Germany generated the least data for 8 of the 24 categories, followed by the UK, with 6 out of the 24 types of data.

PAPER 1: Table 1. *Characteristics of the sample.*

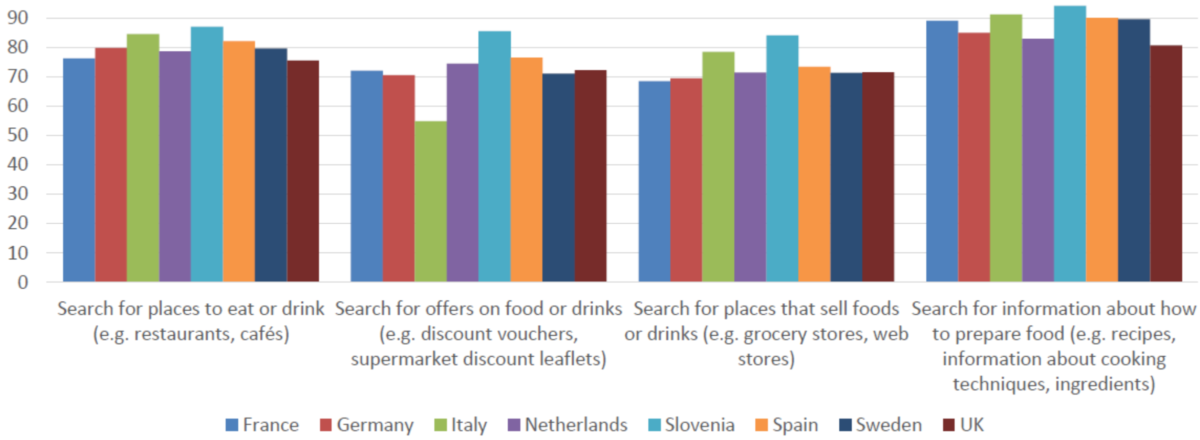
Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Gender								
Male	50.7 (531)	50.2 (523)	50.0 (460)	49.6 (503)	43.3 (446)	49.6 (463)	50.5 (528)	50.0 (510)
Female	49.0 (513)	49.3 (514)	49.5 (455)	49.5 (502)	56.7 (585)	50.0 (467)	49.1 (513)	49.7 (506)
Missing	0.3 (3)	0.5 (5)	0.5 (5)	0.9 (9)	0.0 (0)	0.4 (4)	0.4 (4)	0.3 (3)

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

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Age								
18 - 29 years	20.8 (218)	20.2 (211)	9.6 (88)	19.8 (201)	20.0 (206)	10.3 (96)	20.8 (217)	19.9 (203)
30 - 39 years	20.1 (210)	20.6 (215)	22.5 (207)	19.8 (201)	22.9 (236)	22.7 (212)	20.2 (211)	19.9 (203)
40 - 49 years	20.1 (210)	19.8 (206)	22.5 (207)	19.9 (202)	21.8 (225)	22.7 (212)	19.8 (207)	19.9 (203)
50 -59 years	19.5 (204)	19.7 (205)	22.2 (204)	20.2 (205)	21.4 (221)	22.4 (209)	19.5 (204)	19.9 (203)
60 + years	19.4 (203)	19.3 (201)	23.2 (213)	19.8 (201)	13.7 (141)	21.7 (203)	19.5 (204)	20.1 (205)
Missing	0.2 (2)	0.4 (4)	0.1 (1)	0.4 (4)	0.2 (2)	0.2 (2)	0.2 (2)	0.2 (2)
Internet use								
A few times a week	4.0 (42)	2.7 (28)	1.6 (15)	3.6 (37)	3.2 (33)	1.1 (10)	2.7 (28)	3.7 (38)
Almost every day	6.9 (72)	8.1 (84)	8.9 (82)	16.5 (167)	13.5 (139)	11.5 (107)	6.6 (69)	8.8 (90)
Every day	88.9 (931)	89.2 (929)	89.5 (823)	79.6 (807)	83.3 (859)	87.4 (816)	90.0 (941)	87.1 (888)
I don't know	0.2 (2)	0.1 (1)	0.0 (0)	0.3 (3)	0.0 (0)	0.1 (1)	0.7 (7)	0.3 (3)
Devices used to generate food related data								
Computer	89.4 (936)	78.3 (816)	91.1 (838)	77.4 (785)	89.8 (926)	88.0 (822)	78.7 (822)	74.7 (761)
Phone	62.2 (651)	66.3 (691)	79.0 (727)	68.1 (691)	79.1 (816)	81.0 (757)	74.2 (775)	62.1 (633)
Tablet	35.7 (374)	34.4 (358)	39.6 (364)	41.4 (420)	31.2 (322)	44.6 (417)	34.4 (359)	38.7 (394)
Health app use								
No	70.3 (736)	67.9 (708)	59.6 (548)	65.9 (668)	47.2 (487)	50.0 (467)	64.9 (678)	65.4 (666)
Yes	29.7 (311)	32.1 (334)	40.4 (372)	34.1 (346)	52.8 (544)	50.0 (467)	35.1 (367)	34.6 (353)
Take-away meals								
Never	39.3 (411)	21.5 (224)	32.7 (301)	27.8 (282)	44.6 (460)	30.1 (281)	21.5 (225)	21.2 (216)
Less than once a week	36.4(381)	40.9 (426)	37.0 (340)	46.6 (473)	35.7 (368)	36.6 (342)	54.4 (569)	48.3 (492)
Once a week	12.2 (128)	19.7 (205)	14.8 (136)	13.7 (139)	9.5 (98)	14.6 (136)	15.5 (162)	17.3 (176)
2-3 times a week	5.6 (59)	11.4 (119)	8.0 (74)	5.4 (55)	5.6 (58)	9.7 (91)	4.6 (48)	7.4 (75)
4-6 times a week	3.7 (39)	3.7 (39)	4.9 (45)	3.1 (31)	3.1 (32)	3.6 (34)	1.5 (16)	3.0 (31)
Every day	2.8 (29)	2.8 (29)	2.6 (24)	3.4 (34)	1.5 (15)	5.4 (50)	2.4 (25)	2.8 (29)
Ready meals								
Never	34.4 (360)	26.8 (279)	36.7 (338)	40.0 (406)	45.6 (470)	36.5 (341)	34.1 (356)	23.1 (235)
Less than once a week	31.5 (330)	36.6 (381)	30.2 (278)	34.3 (348)	35.8 (369)	30.4 (284)	38.9 (407)	36.8 (375)
Once a week	14.0 (147)	21.4 (223)	16.1 (148)	12.2 (124)	10.3 (106)	15.1 (141)	15.3 (160)	18.6 (190)
2-3 times a week	12.0 (126)	10.5 (109)	10.2 (94)	7.0 (71)	4.8 (50)	9.2 (86)	7.0 (73)	12.8 (130)
4-6 times a week	5.7 (60)	3.6 (37)	4.6 (42)	4.5 (46)	2.2 (23)	4.6 (43)	3.3 (35)	6.0 (61)
Every day	2.3 (24)	1.2 (13)	2.2 (20)	1.9 (19)	1.3 (13)	4.2 (39)	1.3 (14)	2.7 (28)
Eating out								
Never	29.5 (309)	19.7 205)	22.7 (209)	25.6 (260)	30.4 (313)	15.8 (148)	24.2 (253)	15.7 (160)

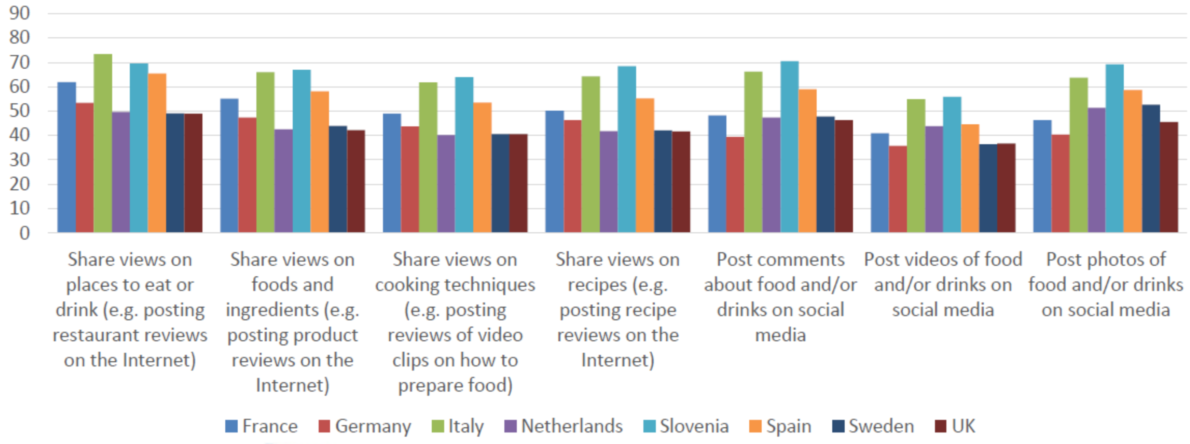
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654280.

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Less than once a week	40.6 (425)	57.5 (599)	39.8 (366)	55.6 (564)	46.8 (482)	52.9 (494)	56.4 (589)	59.7 (608)
Once a week	13.2 (138)	13.5 (141)	22.1 (203)	8.8 (89)	11.8 (122)	18.2 (170)	10.9 (114)	13.9 (142)
2-3 times a week	8.8 (92)	6.1 (64)	8.4 (77)	5.0 (51)	7.0 (72)	7.1 (66)	5.5 (57)	5.6 (57)
4-6 times a week	5.4 (57)	2.1 (22)	4.8 (44)	3.1 (31)	2.8 (29)	2.9 (27)	1.9 (20)	2.7 (28)
Every day	2.5 (26)	1.1 (11)	2.3 (21)	1.9 (19)	1.3 (13)	3.1 (29)	1.1 (12)	2.4 (24)
Responsible for shopping								
No	1.7 (18)	2.9 (30)	1.7 (16)	5.3 (54)	4.1 (42)	2.0 (19)	3.5 (37)	3.1 (31)
Shared	21.1 (220)	29.8 (311)	29.8 (274)	28.1 (284)	42.2 (434)	28.3 (264)	36.1 (377)	29.9 (303)
Yes	77.2 (807)	67.3 (701)	68.5 (630)	66.6 (673)	53.7 (552)	69.7 (650)	60.3 (630)	67.1 (680)
Responsible for cooking								
Never	4.9 (51)	6.3 (66)	5.3 (49)	7.0 (71)	6.8 (70)	5.2 (49)	5.0 (52)	6.8 (69)
1 or 2 times per week	20.5 (215)	20.5 (214)	21.7 (200)	21.3 (216)	24.2 (250)	16.8 (157)	16.3 (170)	17.4 (177)
3 or 4 times per week	17.5 (183)	24.7 (257)	17.5 (161)	20.1 (204)	24.0 (247)	24.9 (233)	24.9 (260)	20.0 (204)
5 or 6 times per week	14.3 (150)	16.6 (173)	13.4 (123)	20.3 (206)	13.6 (140)	1.4 (13)	16.0 (167)	18.4 (187)
Every day	42.8 (448)	31.9 (332)	42.1 (387)	31.3 (317)	31.4 (324)	51.6 (482)	37.9 (396)	37.5 (382)



PAPER 1: Figure 2. *Type of data generated: % respondents producing search data.*

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

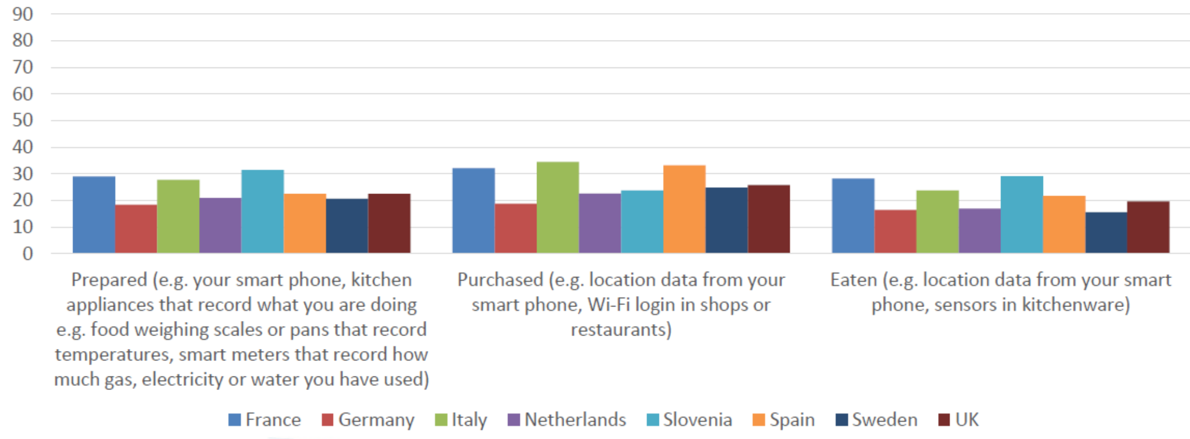


PAPER 1: Figure 3. *Type of data generated: % respondents producing "opinion sharing" data.*

PAPER 1: Figure 4. *Type of data generated: % respondents producing "planning and buying" data.*

PAPER 1: Figure 5. *Type of data generated: % respondents producing "record keeping" data.*

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



PAPER 1: Figure 6. *Type of data generated: % respondents producing "location" data generated using smart devices that record when, where or how food is....*

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

PAPER 1: Table 2. *Type of data generated: % respondents (number of respondents).*

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
SEARCH DATA								
Searching for information on how to prepare food (e.g. recipes, information about cooking techniques, ingredients)	89.0 (932)	84.9 (885)	91.2 (839)	82.94 (841)	94.18 (971)	90.04 (841)	89.57 (936)	80.67 (822)
Searching for places to eat or drink (e.g. restaurant)	76.2 (798)	79.9 (832	84.6 (778)	78.60 (797)	87.00 (897)	82.12 (767)	79.62 (832)	75.56 (770)
Searching for offers on food or drinks (e.g. discount vouchers)	72.0 (754)	70.5 (735	54.9 (505)	74.46 (755)	85.55 (882)	76.55 (715)	71.10 (743)	72.23 (736)
Searching for places that sell foods or drink	68.5 (717)	69.4 (723)	78.5 (722)	71.40 (724)	84.09 (867)	73.34 (685)	71.29 (745)	71.54 (729)
"OPINION SHARING" DATA								
Sharing views on places to eat or drink (e.g. posting restaurant reviews on the internet)	61.9 (648)	53.4 (556)	73.4 (675)	49.6 (503)	69.5 (717)	65.4 (611)	49.1 (513)	49.0 (499)
Sharing views on foods and ingredients (e.g. post reviews on the internet)	55.1 (577)	47.3 (493)	65.9 (606)	42.6 (432)	66.9 (690)	58.1 (543)	43.9 (459)	42.2 (430)
Sharing views on cooking techniques (e.g. posting recipes or clips on how to prepare food)	49.0 (513)	43.7 (455)	61.7 (568)	40.2 (408)	63.9 (659)	53.4 (499)	40.6 (424)	40.5 (413)
Sharing views on recipes (e.g. posting reviews on the internet)	50.1 (525)	46.4 (483)	64.2 (591)	41.7 (423)	68.5 706)	55.3 (516)	42.1 (440)	41.6 (424)
Posting comments about food and/or drink on social media	48.1 (504)	39.4 (411)	66.2 (609)	47.4 (481)	70.5 727)	59.0 (551)	47.9 (500)	46.3 (472)
Posting videos of food and/or drink on social media (e.g. YouTube, Facebook, Twitter)	41.0 (429)	35.8 (373)	54.9 (505)	43.9 (445)	55.9 (576)	44.7 (417)	36.4 (380)	36.7 (374)
Posting photos of food and/or drink on social media (e.g. Facebook, Twitter)	46.3 (485)	40.4 (421)	63.7 (586)	51.3 (520)	69.2 (713)	58.6 (547)	52.6 (550)	45.5 (464)
"PLANNING AND BUYING" DATA								
Booking places to eat (e.g. restaurants)	62.3 (652)	59.9 (624)	77.3 (711)	76.1 (772)	66.5 (686)	71.4 (667)	61.2 (639)	65.1 (663)
Comparing food/drink products and prices	65.8 (689)	65.9 (687)	75.8 (697)	67.4 (683)	74.6 (769)	70.3 (657)	67.6 (706)	67.4 (687)
Buying food or drinks (e.g. online grocery shopping/takeaways)	55.2 (578)	50.3 (524)	64.2 (591)	55.0 (558)	55.4 (571)	61.1 (571)	53.3 (557)	63.1 (643)
Creating shopping lists	49.5 (518)	46.6 (485)	59.5 (547)	49.4 (501)	55.2 (569)	58.6 (547)	51.6 (539)	47.7 (486)
Planning menus/meals	41.2 (431)	42.3 (441)	50.4 (464)	40.9 (415)	46.1 (475)	55.7 (520)	48.5 (507)	39.9 (407)
"RECORD KEEPING" DATA								
Creating or managing recipe collections	50.5 (529)	47.1 (491)	54.8 (504)	51.9 (526)	63.2 (652)	61.6 (575)	51.7 (540)	42.4 (432)
Keeping records of money spend on food/drinks (i.e. budgeting)	40.9 (428)	37.3 (389)	59.0 (543)	46.7 (473)	49.1 (506)	38.0 (355)	42.5 (444)	40.8 (416)
Keep records of food/drink stores in the household (e.g. what's in the freezer)	33.5 (351)	28.8 (300)	42.2 (388)	30.5 (309)	36.0 (371)	38.0 (355)	27.6 (288)	31.8 (324)

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

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Keep records of what you eat and drink	33.1 (347)	31.0 (323)	43.3 (398)	34.3 (348)	31.0 (320)	33.9 (317)	30.1 (314)	33.9 (345)
Keep records of what people in your household eat or drink	31.4 (329)	26.2 (273)	39.7 (365)	27.6 (280)	30.4 (313)	33.0 (308)	22.6 (236)	29.6 (302)
“LOCATION” DATA GENERATED USING SMART DEVICES THAT RECORD WHEN, WHERE OR HOW FOOD IS....								
Purchased (e.g. location data from your mobile phone, Wi-Fi login in shops or restaurants)	32.1 (336)	18.7 (195)	34.5 (317)	22.6 (229)	23.8 (245)	33.2 (310)	24.8 (259)	25.8 (263)
Prepared (e.g. your mobile phone, kitchen appliances that record what you are doing e.g. food weighing scales or pans that record temperatures, smart meters that record how much gas, electricity or water you have used)	29.04 (304)	18.4 (192)	27.7 (255)	20.9 (212)	31.4 (324)	22.5 (210)	20.7 (216)	22.5 (229)
Eaten (e.g. location data from your mobile phone, sensors in kitchenware)	28.3 (296)	16.4 (171)	23.7 (218)	16.9 (171)	29.1 (300)	21.7 (203)	15.6 (163)	19.6 (200)

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

PAPER 1: 4. DISCUSSION

The current growth in the use of mobile technology and specifically app use, makes this an important and interesting time in terms of the potential for this data. If researchers are able to again access to user-generated data sets, it emanates many of the problems traditionally associated with research. However, with it comes a new set of problems that have to be considered.

The primary motivation for using domestic food preparation apps is to develop personal food knowledge, skills and/or abilities. This opens up the potential to answer questions relating to Individual Psychological determinants, such as food beliefs, habits and self-regulation in relation to food. However, the limited availability of contextual data, such as that at the 'Individual/Situation', and 'Interpersonal/Social' levels, means that much of this data is detached from the user. Researchers intending to use this data will have to carefully consider the degree to which additional contextual information is required to draw conclusions.

The interconnectedness of the apps presents new opportunities to further enrich data collected from external sources, i.e. there is the potential to create 'links' between the multiple apps used by a single user. For example, it may be useful to gain domestic food preparation specific information from dedicated apps, and enrich this with demographic, situational and social context data collected through apps such as Facebook, Twitter and Instagram. However, the degree to which users would find this interlinkage acceptable still needs to be investigated. It should be noted that to date this type of data has been used to study food consumption patterns, e.g. Twitter (Abbar et al.2014; Fried et al., 2014) and Instagram (Mejova et al., 2015; Sharma and De Choudhury, 2015).

A criterion for inclusion into the inventory was that the app provided sufficient details, so as to facilitate the completion of the majority of these quality criteria. The decision was made that this task should only use information that was publically available. That is, not to contact the company and/or app developer for additional information beyond what could be found from either using the app, the retail store (e.g. Google Play), or an accompanying website. This was in order to test the feasibility of applying the quality criteria with only publically available information. It was found that a large proportion of domestic food preparation apps were developed by small independent app developers and therefore didn't have the same level of accompanying information as those produced by larger companies. Many apps did not have an accompanying websites or API and thus access to information, such as terms and condition, was limited.

Many domestic food preparation apps do not collect consumer-generated data, they provide the consumer with information, and indeed some apps only provide the consumer with information. The need for information can be said to be a major motivating factor for the use of an app in the preparation of food. Searching for cooking times, oven temperatures or weight conversions are all food preparation tasks for which apps are commonly available and yet no consumer data is collected by the app. These apps were therefore excluded from inclusion into the inventory.

As the focus of our search was on food-specific apps, the role of social media apps in food preparation was also not thoroughly explored in this study. Common social media application, such as Facebook, Twitter and Instagram, can all be used by the consumer to share and gain information about food preparation – and indeed, food purchase and food consumption. However, the complex and disjointed nature of the consumer-generated data collected via these platforms makes it difficult to unpick the relevant data points. Thus it was decided to include only those apps designed specifically for domestic food preparation according to the definition set down in the document. It is therefore recommended that further investigation is made of the consumer-generated domestic food preparation data gathered through social media platforms.

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

A further point to consider with user-generated domestic food preparation data, is the degree to which it can act as a 'proxy' for intake. The data collected via app usage reflects the motivation to gain knowledge and to develop skills. The degree to which this is translated into intake cannot be directly drawn from the data in its current form. At best, it describes an 'intention' to intake certain foods and/or meals.

The survey data shows that across the EU, a wide range of food-related data is being generated by consumers, and that there are differences between countries with respect to the kind and range of data produced. Whilst this descriptive information in itself is likely to change over short term due to the rapid advances in technology capable of capturing food-related information across the EU. Search data is the most likely data to be in abundant supply and yet, it holds relatively limited value for the study of the actual behaviour. On the other hand, the data on food intake and the relevant context data (e.g. location) is relatively harder to come by. Nevertheless, it is important to understand that the linking of these different types of data may achieve the granularity necessary for breakthrough science of consumer food-related behaviour

.

PAPER 1: REFERENCES

Abbar, S., Mejova, Y., Weber, I. (2014) You tweet what you eat: studying food consumption through Twitter, *CoRR abs*/1412.4361

Adhikari, R., Richards, D., Scott, K. (2014) Security and privacy issues related to the use of mobile health apps. Paper presented at the *25th Australasian Conference on Information Systems.*

Blenner, S. R., Kollmer, M., Rouse, A. J., Daneshvar, N., Williams, C., Andrews, L. B. (2016) Privacy policies of android diabetes apps and sharing of health information. *JAMA*, 315(10), 1051-1052.

Boulos, M. N., Brewer, A. C., Karimkhani, C., Buller, D. B., Dellavalle, R. P. (2014) Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics*, 5(3), 229.

Brown, W., 3rd, Yen, P. Y., Rojas, M., Schnall, R. (2013).Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *Journal of Biomedical Informatics*, 46(6), 1080-1087.

Cummings, E., Borycki, E. M., Roehrer, E. (2013) Issues and considerations for healthcare consumers using mobile applications. Studies in Health Technology and Informatics, 183, 227-231.

Dedeke, A. (2000) A Conceptual Framework for Developing Quality Measures for Information Systems. Paper presented at the *Conference on Information Quality* Retrieved from http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202000/Papers/AConceptualFramework4DevelopQualityMeasure.pdf

Dufty, D., Bérard, H., Lefranc, S., Signore, M. (2014) *A Suggested Framework for the Quality of Big Data.*

Ernsting, C., Dombrowski, S. U., Oedekoven, M., O Sullivan, J. L., Kanzler, M., Kuhlmey, A., Gellert, P. (2017) Using smartphones and health apps to change and manage health behaviors: A population-based survey. *Journal of Medical Internet Research*, 19(4), e101.

Fordyce-Voorham, S. (2009) Essential food skills in skill-based healthy eating program. *Journal of the Home Economics Institute of Australia,* 16(2), 16-20.

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

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., Bell, D. (2014) Analyzing the language of food on social media, in IEEE International Conference on Big Data.

Ho, B., Lee, M., Armstrong, A. W. (2013) Evaluation criteria for mobile teledermatology applications and comparison of major mobile teledermatology applications. *Telemedicine and e-Health*, 19(9), 678-682.

Knight, S., Cowan, E. (2005) Developing a framework for assessing information quality on the World Wide Web. *Informing Science Journal*, 8.

Köster, E.P. (2009) Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference* 20, 70–82

Kumar, S., Nilsen, W.J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., Riley, W. T., Shar, A., Spring, B., Spruijt-Metz, D., Hedeker, D., Honavar, V., Kravitz, R., Lefebvre, R.C., Mohr, D.C., Murphy, S.A., Quinn, C., Shusterman, V., Swendeman, D. (2013) Mobile health technology evaluation: the mHealth evidence workshop. *American Journal of Preventive Medicine,* 45(2), 228-236.

Lavelle, F., McGowan, L., Hollywood, L., Surgenor, D., McCloat, A., Mooney, E,. Caraher, M., Raats, M., Dean, M. (2017) The development and validation of measures to assess cooking skills and food skills, *International Journal of Behavioral Nutrition and Physical Activity*, 14, 1, 118.

Lavelle, F., Spence, M., Hollywood, L., McGowan, L., Surgenor, D., McCloat, A., Mooney, E., Caraher, M., Raats, M. & Dean, M. (2016) Learning cooking skills at different ages: a cross-sectional study, *International Journal of Behavioral Nutrition and Physical Activity*, 13, 11, 119.

Loshin, D. (2002) Knowledge integrity: data ownership. Data Warehouse magazine (Online) June 8, 2004.

Martinez-Perez, B., de la Torre-Diez, I., Candelas-Plasencia, S., Lopez-Coronado, M. (2013) Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications*. Journal of Medical Systems*, 37(5), 9976.

McGowan, L., Pot, G. K., Stephen, A. M., Lavelle, F., Spence, M., Raats, M., Hollywood, L., McDowell, D., McCloat, A., Mooney, E., Caraher, M., Dean, M. (2016) The influence of socio-demographic, psychological and knowledge-related variables alongside perceived cooking and food skills abilities in the prediction of diet quality in adults: a nationally representative cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*. 13, 13 p., 111.

Mejova, Y., Haddadi, H., Noulas, A., Weber I. (2015) #foodporn: Obesity patterns in culinary interactions, in Proceeding DH '15 *Proceedings of the 5th International Conference on Digital Health,* pp 51-58.

Meulendijk, M. C., Meulendijks, E. A., Jansen, P. A. F., & Numans, M. E. M. (2014) What concerns users of medical apps? Exploring non-functional requirements of medical mobile applications. Paper presented at *the Twenty Second European Conference on Information Systems*, Tel Aviv.

Njie, C. M. L. (2013). *Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications.*

Raats M. M., Hieke, S., Jola, C., Hodgkins, C., Kennedy, J., Wills, J. (2015) Reference amounts utilised in front of pack nutrition labelling; impact on product healthfulness evaluations. *European Journal of Clinical Nutrition*, 69(5): 619-625 16.

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

Schulze, K., Kromker, H. (2010) A framework to measure user experience of interactive online products. Paper presented at the *7th International Conference on Methods and Techniques in Behavioral Research*.

Sharma S., De Choudhury, M. (2015) Detecting and characterizing nutritional information of food and ingestion content in Instagram, in *Proceeding WWW '15 Companion Proceedings of the 24th International Conference on World Wide Web*, pp 115-116

Sobal, J., Bisogni, C. A. (2009) Constructing Food Choice Decisions. *Annals of Behavioral Medicine*, 38, 1, s37–s46.

Sobal, J. (1991) Obesity and socioeconomic status: A framework for examining relationships between physical and social variables. *Medical Anthropology*, 13, 231–247.

Sobal, J., Bisogni, C. A., Jastran, M. (2014). Food choice is multifaceted, contextual, dynamic, multilevel, integrated, and diverse. *Mind, Brain, and Education*, 8: 6–12.

Stead, M., Caraher, M., Wreiden, W., Longbottom, P., Valentine, K., Anderson, A. (2004) Confident, fearful and hopeless cooks. *British Food Journal*, 106(4) 247-287.

Stok, F. M., Hoffmann, S., Volkert, D., Boeing, H., Ensenauer, R., Stelmach-Mardas, M., Eva Kiesswetter, E., Alisa Weber, A., Harald Rohm, H., Nanna Lien, N., Brug, J., Holdsworth, 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

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., Mani, M. (2015) Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth*, 3(1), e27.

Wilkinson, M. D.; Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L. B., Bourne, P. E.; Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L.; Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A. Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

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

PAPER 2: 1. INTRODUCTION

PAPER 2: 1.1 BACKGROUND AND RATIONALE

The ongoing digital revolution is creating the potential for huge datasets generated by computers, apps and mobile devices to be put to work in solving the grand challenges facing society today. Much of this "big data" is generated as a by-product of app use and only wants the appropriate infrastructure and the consent of the user to be established before it can be put to fruitful use in learning more about the diet, lifestyle and health of millions of individuals. Typically, though, the technological capacity to create and capture this data is developing far faster than the knowledge required to use it ethically and wisely (Nilsen et al., 2012, Bennett and Raab, 2018).

Use of this data has so far been sporadic. On the one hand researchers in the area of infectious disease monitoring have already seized upon the potential of user-generated data to improve public health surveillance; on the other health agencies have been reluctant to fully embrace new data sources due to the many technical issues, not to mention the huge ethical questions raised, that have yet to be fully resolved (Velasco et al., 2014). In the area of diet and health specifically, researchers have recently used data collected through Twitter (Abbar et al., 2014; Fried et al., 2014) and Instagram (Mejova et al., 2015; Sharma and De Choudhury, 2015) to study food consumption patterns. Weber and Achananuparp (2016) used data from public food diaries collected using the application MyFitnessPal to construct models predicting whether users will or will not meet their daily calorie goals. This ably demonstrates the potential of repurposing data generated by social media and mobile apps for use in health research; what is still wanting is an endeavour with the vision and scope to bring all these different kinds of data together to create a broader understanding of the food choices consumers make and why they make them.

Food choice is influenced by multifarious physical, biological, psychological, and sociocultural factors (Sobal, 1991) all acting (Sobal et al., 2014). The relationship between the consumer and their food is therefore a complex and emotionally layered one with resonances in areas as sensitive, diverse and culturally loaded as gender and mental health. Crucially for food behaviour research, work on collecting eHealth data has revealed that while progress is rapidly being made with the technical and logistical challenges involved we still understand little of the human factors affecting willingness to contribute data (Dinev et al., 2016). What research has been done with users of eHealth devices thus far suggests that most have scant awareness of the nature and severity of its potential risks and frequently underestimate them as a result (Bellekens et al., 2016). What is already clear is that a world driven by data runs not on technology but on trust (Wilton, 2015). Surrendering a degree of control over their own privacy requires trust that those who use their data will do so in ways commensurate with their own values and wants.

What is needed, therefore, is a broad survey of motivations for and against sharing food-related data, one incorporating a comprehensive range of measures. Given the variability in food supply, consumer lifestyles and culture across Europe it is essential this be applied across a large, representative population in a number of countries. It must incorporate insight into not just who will share which data and why, but for what purposes they would be willing to see it used. The utility of comparing the same individual's attitudes towards sharing data to be used for different research purposes has been demonstrated by Kim and colleagues (2015), who also draw attention to a lack of large scale research into consumer's feelings about issues of privacy, consent and data security, especially work which considers different kinds of usage. This is a key issue for big data which, though often collected for a specific type of use, is often detailed and flexible enough to be employed for a range of purposes by

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

diverse organisations with disparate goals – goals which may or may not be commensurate with the values of those who must consent to its use.

As outlined in Richfield's Deliverable 13.2 (Carr, 2018), unlike medical research which has some accepted standards such as the World Health Organisation's 'Standards and Operational Guidance for Ethics Review of Health-related Research with Human Participants' (WHO, 2011) and those from the UK Medical Research Council (2012). The primary issues of ethical concern with regard to personal data for research purposes are privacy, informed consent and ownership of data (Carr, 2018). Article 8 of the Charter of Fundamental of the European Union provides every individual with the right to protection of personal data about him or her, it reads:

Protection of personal data

1. *Everyone has the right to the protection of personal data concerning him or her.*
2. *Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.*
3. *Compliance with these rules shall be subject to control by an independent authority.*

Consent is key to legitimising data processing consent is pivotal. The recently introduced General Data Protection Regulation emphasises consent and imposes significant responsibilities on entities that are responsible for processing data. For consent to be valid it must be freely given; a proper explanation of what the individual is consenting to must have been provided before the consent is obtained; separate consents must be given for separate purposes; consent can be refused; and (most important of all) consent can be withdrawn at any time.

Thus the use of personal data in research is limited by the extent to which people are willing give consent to share their data with researchers. Whilst there is has been significant research conducted to understand people's willingness to share health and medical data with researchers, to date little research has been conducted to understand this issue particularly with regard to personal health data collected through mobile devices (Seifert et al., 2018; Bietz et al., 2016; Chen et al., 2016; Pevnik et al., 2016; Prashad, 2014). A recent systematic review and thematic synthesis of qualitative studies (Quiroz-Aitken et al., 2016) examining public attitudes towards the sharing or linkage of health data for research widespread general, albeit conditional, support for data linkage and sharing for research purposes. Whilst a variety of concerns were raised (e.g.) in cases Where participants perceived there to be actual or potential public benefits from research and had trust in the individuals or organisations conducting and/ or overseeing data linkage/sharing, they were generally supportive of sharing data. Concerns raised related to confidentiality, control over one's own data, as well as the potential uses and abuses of data and possible problems that might arising as a result. The authors (Quiroz-Aitken et al., 2016) concluded that there was a need public engagement and deliberation to ensure the legitimacy of future health informatics research. Stockdale and colleagues (2018) reviewed the literature on UK and Irish public opinions of medical data use in research found that whilst there was a widespread willingness to share electronic health records for research for the common good, this was contingent on their evaluation of the trustworthiness of research organisations as assessed by their competence in data-handling and motivation for accessing the data.

Studies by Bietz et al. (2016) and Chen et al. (2016) found large proportions (77-78%) of younger and middle-aged people to be open to sharing their data with researchers. In a recent study with older respondents Seifert and colleagues (2018) found that 57.2% of the participants who tracked their health data were willing to share it with researchers. Skatova and colleagues (2014) found the majority

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

of respondents in their studies were willing to share personal data for the purpose of researcher if the research were to lead to public good; where they were less willing, they wanted assurances of direct personal benefits. In an explorative anthropological study on the sharing of mental health data Sleigh (2018) found that regardless privacy and surveillance concerns, participants were driven by altruistic motivations to engage with valued health researchers, in the hope it would contribute to future avenues of research.

PAPER 2: 1.2 RESEARCH AIMS

The goal of the study is to gain insight into the extent people are willing to share food-related data generated by any electronic systems they may use while choosing, purchasing and preparing food with researchers and for what purposes. In this light, the objectives include exploration of factors and identification of variables which explain reasons for choosing to share data for research purposes. This is necessary because little empirical research exists on the extent to which members of the public are consciously willing to share personal data with scientists outside the context of medicine. Where data generated through interactions with healthcare professionals, for example medical records, are sometimes used in an aggregated form for research into specific conditions for the purposes of improving treatment or provision of care. In particular there is a dearth of large scale quantitative research comparing data from a range of countries with diverse cultures and disparate socioeconomic backgrounds who may have different attitudes towards healthy eating and use technology in different ways.

PAPER 2: 2. METHODS

PAPER 2: 2.1 SAMPLING STRATEGY

Participants were 8450 adults from eight European countries. Specifically, 1000 participants were recruited from each of the following countries: France, Italy, the Netherlands, Slovenia, Spain, Sweden, the UK, and Germany. Countries were selected to vary with regard to the extent to which there was evidence of high health privacy concern, time in the EU, cuisine, health care system type and role in the European “Food, Nutrition and Health Research Infrastructure” currently being developed. Stratified sampling employed in each country, so that an equal number of participants was recruited from the following age groups: 18-29 years; 30-39 years; 40-49 years; 60-59 years; and 60+ years. In each age group, we aimed to recruit an equal number of men and women. All participants were recruited through Lightspeed Research (www.lightspeedresearch.com).

PAPER 2: 2.2 INSTRUMENT

Participants filled in an online questionnaire that took about 35 minutes to complete. The instrument contained questions on participants’ willingness to share with commercial, academic and government organisations the data that food-related apps generate on their daily habits. Questions also covered participants’ reasons for (not) sharing such information, their health, as well as their relevant attitudes and values. Participants were asked what types of food related activities they performed on their computers and smartphones. They could select any number of 24 options, including “search for places to eat and drink”, “share views on recipes”, and “keep records of what you eat and drink” (see Table

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

2 for details of the types of consumer-generated food data). A number of questions were asked three times, separately referring to scientists in universities, governments, and companies.

Participants were asked separate questions (adapted from Bietz et al., 2016) on whether they were **willing to share data** with scientists, governments, and companies. Participants were only asked these questions for the types of data (such as restaurant bookings or recipe collections) they generated, according to their previous responses. Thus, each participant stated how likely they were to share one or more of 24 data types. For each type of data, they stated how likely they are to share such data with the respective stakeholder on a Likert scale from 1 (I definitely would not share…) to 5 (I definitely would share…). Cronbach's alpha was .987 for scientists, .988 for governments and .989 for companies.

Trust in scientists, governments, and companies, as well as **perceived risk** in sharing data with each of these stakeholders was assessed with a set of bespoke questions we developed based on the extant literature (Bhattacherjee, 2002; Jarvenpaa, et al., 1999). Eight questions measured trust (e.g. "Companies that produce or sell foods and drinks keep the public's best interest in mind when handling their data."), while four measured risk perceptions (e.g. "In general, it would be risky to give my data to governments."). Scores ranged from 1 to 5; higher scores reflected a higher level of trust and a heightened perception of risk, respectively. For trust, Cronbach's alpha was .922 for sharing with scientists, .951 for governments, and .913 for companies. For risk, Cronbach's alpha was .743 for scientists, .628 for governments and .689 for companies.

Questions related to the following topics were asked only once (i.e. not for every stakeholder group) from each participants.

Attitudes to science were assessed with three questions (from the MRC Ipsos MORI, 2007). On a 10-point Likert scale, higher scores reflected more positive attitudes (e.g. "Science and technology are making our lives healthier, easier, and more comfortable"). Cronbach's alpha was .850.

Subjective health status was assessed with a single Likert-type item ranging from 1 (very bad) to 5 (very good) (from the European Social Survey, http://www.europeansocialsurvey.org/).

Health interest was measured using an 8-item scale (Roininen, et al., 1999) that was designed to measure general health interest in a food context. On a 5-point Likert scale, higher scores reflected more interest in the healthiness of food (e.g. "It is important for me that my daily diet contains a lot of fruits and vegetables") Half of the items were reverse-coded (e.g. "I eat what I like and do not worry about the healthiness of food."). Cronbach's alpha was .822.

Participants' "basic human values" were explored with the 10-item Schwartz Values Survey, as used in the European Social Survey. Each item consisted of a brief vignette describing a value configuration (e.g. "Being very successful is important to this person; to have people recognize one's achievements"). Participants had to express their own level of identification with the values expressed, on a scale ranging from 1 (very much like me) to 6 (not at all like me). An exploratory factor analysis (EFA) was performed on the scores in each country, resulting in two factors that reflected Schwartz's constructs of **self-conservation** (i.e. personal values related to respect for tradition, conformity, and security) and **self-transcendence** (i.e. ability to focus attention on doing something for the sake of others).

Reasons for sharing food-related data were assessed with a set of 20 questions answered on a 5-point Likert scale (Skatova, et al., 2014). While the items were designed to explore six types of reasons, including altruism and reputation, we found that the reasons were very strongly correlated in the present sample. We therefore decided to create one moral motives measure by averaging the all items. Cronbach's alpha was .972.

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

Privacy concern was measured with three items from Patil et al. (2016). The items asking participants if they were concerned about their personal health information being accessed by non-medical personnel, being access by private companies, or being misused for harassment. The Likert scale ranged from "not concerned" (1) to "very concerned" (5). Cronbach's alpha was .842.

Participants' **use of health apps** was assessed with a single item (Ernsting et al., 2017). They were asked to list the purposes, if any, for which they had used smartphone apps over the previous 12 months. Participants could tick any number of seven options (e.g. to quit smoking; to lose weight). Since a large number of participants (47 to 70%, depending on the country) used no health apps, the item was dichotomised (0 – has not used health apps; 1 – has used health apps).

Basic **demographic data** were also collected. These included gender, age, height, weight, employment status, income, highest educational level (from 1 – no formal education to 9 – university degree), and household composition. Based on height and weight, we also computed each participant's BMI. Because 88% of participants did not report their income, we did not use this variable. Since 99.6% of the participants identified as either male or female (as opposed to identifying with another gender, 0.2%, or refusing to answer, 0.2%), we only used the data from men and women when exploring gender issues.

Food practices were assessed with a series of brief questions. First, participants were asked how many minutes they spent cooking on a typical weekday and on a typical weekend day. From these data, we estimated each participant's average **cooking time** per week (5*typical weekday +2* typical weekend day). We then explored whether participants had the **responsibility for shopping** (Raats et al., 2015; Hieke et al., 2016) and **for cooking** (Lavelle et al., 2016; McGowan et al., 2016; Lavelle et al., 2017)) respectively, within their households. Responsibility for shopping was assessed on a three-point Likert scale ranging from "no" through "shared" to "yes". Finally, we asked participants how often in a week they had a **takeaway, a ready meal, and pub/restaurant meal**, respectively. These questions were answered on a Likert scale ranging from every day (1) to never (6). This was later reverse-coded so that larger scores indicate eating more of these meals.

PAPER 2: 2.3 PROCEDURE

The study was conducted according to the guidelines laid down in the Declaration of Helsinki and in accordance with the University of Surrey's ethical procedures. Participants provided informed consent, then answered the questions in the order provided above. The survey was administered via Qualtrics™.

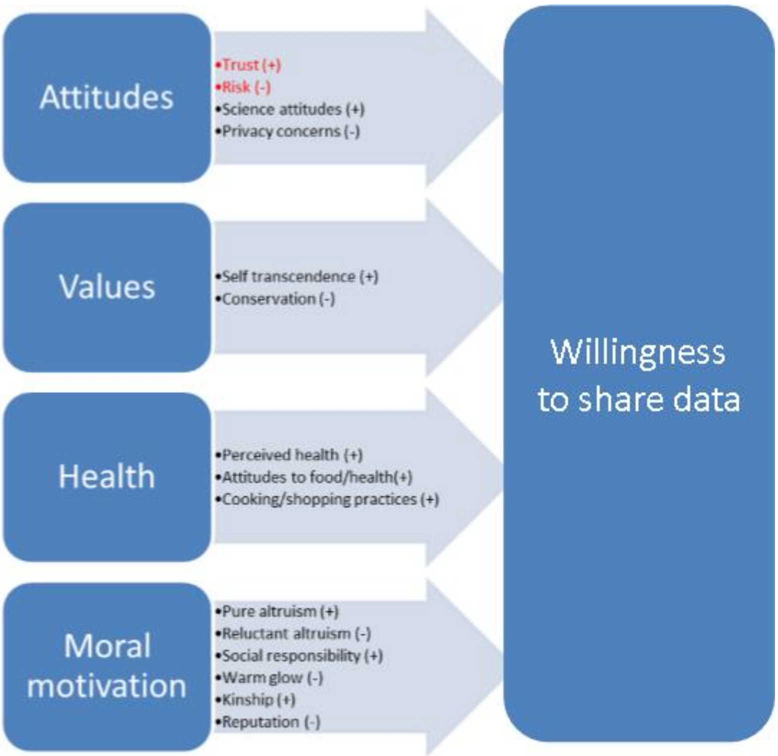
The questionnaire was developed in English and then translated, checked by native speakers and put into Qualtrics™. Data collection for each country was run separately. Should other researchers wish to conduct a comparable study, advice can be sought from the authors with regard to translation procedures, question selection, dataset preparation and analytic strategy.

PAPER 2: 2.4 ANALYTIC STRATEGY

Summary scores for the measures above were computed in each of the eight country-level datasets, and incomplete cases were deleted. All variables of interest (including demographics) were aggregated into a master dataset. The master dataset was saved both in a wide and a long format. Descriptive statistics and country comparisons were then computed. This stage of data processing was performed in IBM SPSS 24 and 25.

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

Next, a data-driven model was developed to predict participants' willingness to share data with universities, governments, and companies. To explore the data and build models that predict willingness to share data with the three stakeholders, a data-driven approach was adopted. See Figure one for an overview of the variables considered in the model. First, a random sample of 20% was extracted from the dataset for developing the model. The final model will be tested on the remaining data. We started with a saturated model and reached a more parsimonious version through backward elimination. This analysis was performed in R 3.4.3 with the lme4 package.



PAPER 2: Figure 1. *Variables considered in building the model and their predicted influence (+ indicates a positive influence; - indicates a negative influence). Variables that are in RED are measured separately for the three types of stakeholders).*

PAPER 2: 3. RESULTS

PAPER 2: 3.1 SAMPLE CHARACTERIZATION

A total of 8052 individuals completed the survey with full data needed for the analysis (see above comments regarding gender). Demographic statistics and descriptors that characterize this cohort are provided in Table 1. Details of the nature and degree of generating food-related data in Table 2 and Figures 2-6. As shown, the targeted sampling strategy was effective at recruiting a sample of individuals that met the stratification targets. Most respondents were responsible for food shopping. Variability across countries is apparent in terms of the types of data generated. The most commonly generated data is search data, the least common is "location" data generated using smart devices that record when, where or how food is purchased, prepared or eaten. Slovenia was the country with the highest percentage of respondents generating data for 14 of the 24 types of data, followed by Italy, with 9 out of the 24 types of data. Germany generated the least data for 8 of the 24 categories, followed by the UK, with 6 out of the 24 types of data.

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

PAPER 2: 3.2 DESCRIPTIVE STATISTICS. DIFFERENCES AMONG COUNTRIES

Recall that the willingness to share data, trust, and risk were assessed separately for universities, governments, and private companies. An inspection of the means and standard deviations of these variables across the eight countries (see Table 3 and Figures 7-9) occasions two observations:

(1) The willingness to share, as well as the trust and perceived risk were above the midpoint of the scale (2.5 on a 5-point scale) for all three stakeholders in all eight countries;

(2) Participants were more willing to share data with universities than with governments and companies; they also expressed more trust in universities, and perceived less risk than for other stakeholders.

The effect of country and stakeholder on sharing was assessed in a mixed-design ANOVA. The results confirm a general preference for universities and significant but small variation across countries. Willingness to share differed both by country, F (7, 7716) = 61.155, $p < .001$, $\eta^2 = .053$, and by stakeholder, F (2, 7716) = 577.249, $p < .001$, $\eta^2 = .070$; the interaction was significant but very weak, F (14, 7716) = 6.414, $p < .001$, $\eta^2 = .006$. Planned contrasts showed that participants were more willing to share with universities than with governments and companies, F (1, 7716) = 584.073, $p < .001$, $\eta^2 = .127$. Participants were also more willing to share with companies than governments, but this effect was very weak, F (1, 7716) = 47.530, $p < .001$, $\eta^2 = .008$.

The correlates of willingness to share data are presented in Tables 4 and 5.

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

PAPER 2: Table 1. *Characteristics of the sample.*

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Gender								
Male	50.7 (531)	50.2 (523)	50.0 (460)	49.6 (503)	43.3 (446)	49.6 (463)	50.5 (528)	50.0 (510)
Female	49.0 (513)	49.3 (514)	49.5 (455)	49.5 (502)	56.7 (585)	50.0 (467)	49.1 (513)	49.7 (506)
Missing	0.3 (3)	0.5 (5)	0.5 (5)	0.9 (9)	0.0 (0)	0.4 (4)	0.4 (4)	0.3 (3)
Age								
18 - 29 years	20.8 (218)	20.2 (211)	9.6 (88)	19.8 (201)	20.0 (206)	10.3 (96)	20.8 (217)	19.9 (203)
30 - 39 years	20.1 (210)	20.6 (215)	22.5 (207)	19.8 (201)	22.9 (236)	22.7 (212)	20.2 (211)	19.9 (203)
40 - 49 years	20.1 (210)	19.8 (206)	22.5 (207)	19.9 (202)	21.8 (225)	22.7 (212)	19.8 (207)	19.9 (203)
50 -59 years	19.5 (204)	19.7 (205)	22.2 (204)	20.2 (205)	21.4 (221)	22.4 (209)	19.5 (204)	19.9 (203)
60 + years	19.4 (203)	19.3 (201)	23.2 (213)	19.8 (201)	13.7 (141)	21.7 (203)	19.5 (204)	20.1 (205)
Missing	0.2 (2)	0.4 (4)	0.1 (1)	0.4 (4)	0.2 (2)	0.2 (2)	0.2 (2)	0.2 (2)
Internet use								
A few times a week	4.0 (42)	2.7 (28)	1.6 (15)	3.6 (37)	3.2 (33)	1.1 (10)	2.7 (28)	3.7 (38)
Almost every day	6.9 (72)	8.1 (84)	8.9 (82)	16.5 (167)	13.5 (139)	11.5 (107)	6.6 (69)	8.8 (90)
Every day	88.9 (931)	89.2 (929)	89.5 (823)	79.6 (807)	83.3 (859)	87.4 (816)	90.0 (941)	87.1 (888)
I don't know	0.2 (2)	0.1 (1)	0.0 (0)	0.3 (3)	0.0 (0)	0.1 (1)	0.7 (7)	0.3 (3)
Devices used to generate food related data								
Computer	89.4 (936)	78.3 (816)	91.1 (838)	77.4 (785)	89.8 (926)	88.0 (822)	78.7 (822)	74.7 (761)
Phone	62.2 (651)	66.3 (691)	79.0 (727)	68.1 (691)	79.1 (816)	81.0 (757)	74.2 (775)	62.1 (633)
Tablet	35.7 (374)	34.4 (358)	39.6 (364)	41.4 (420)	31.2 (322)	44.6 (417)	34.4 (359)	38.7 (394)
Health app use								
No	70.3 (736)	67.9 (708)	59.6 (548)	65.9 (668)	47.2 (487)	50.0 (467)	64.9 (678)	65.4 (666)
Yes	29.7 (311)	32.1 (334)	40.4 (372)	34.1 (346)	52.8 (544)	50.0 (467)	35.1 (367)	34.6 (353)
Take-away meals								
Never	39.3 (411)	21.5 (224)	32.7 (301)	27.8 (282)	44.6 (460)	30.1 (281)	21.5 (225)	21.2 (216)
Less than once a week	36.4(381)	40.9 (426)	37.0 (340)	46.6 (473)	35.7 (368)	36.6 (342)	54.4 (569)	48.3 (492)
Once a week	12.2 (128)	19.7 (205)	14.8 (136)	13.7 (139)	9.5 (98)	14.6 (136)	15.5 (162)	17.3 (176)
2-3 times a week	5.6 (59)	11.4 (119)	8.0 (74)	5.4 (55)	5.6 (58)	9.7 (91)	4.6 (48)	7.4 (75)
4-6 times a week	3.7 (39)	3.7 (39)	4.9 (45)	3.1 (31)	3.1 (32)	3.6 (34)	1.5 (16)	3.0 (31)
Every day	2.8 (29)	2.8 (29)	2.6 (24)	3.4 (34)	1.5 (15)	5.4 (50)	2.4 (25)	2.8 (29)
Ready meals								
Never	34.4 (360)	26.8 (279)	36.7 (338)	40.0 (406)	45.6 (470)	36.5 (341)	34.1 (356)	23.1 (235)
Less than once a week	31.5 (330)	36.6 (381)	30.2 (278)	34.3 (348)	35.8 (369)	30.4 (284)	38.9 (407)	36.8 (375)
Once a week	14.0 (147)	21.4 (223)	16.1 (148)	12.2 (124)	10.3 (106)	15.1 (141)	15.3 (160)	18.6 (190)
2-3 times a week	12.0 (126)	10.5 (109)	10.2 (94)	7.0 (71)	4.8 (50)	9.2 (86)	7.0 (73)	12.8 (130)
4-6 times a week	5.7 (60)	3.6 (37)	4.6 (42)	4.5 (46)	2.2 (23)	4.6 (43)	3.3 (35)	6.0 (61)

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Every day	2.3 (24)	1.2 (13)	2.2 (20)	1.9 (19)	1.3 (13)	4.2 (39)	1.3 (14)	2.7 (28)
Eating out								
Never	29.5 (309)	19.7 205)	22.7 (209)	25.6 (260)	30.4 (313)	15.8 (148)	24.2 (253)	15.7 (160)
Less than once a week	40.6 (425)	57.5 (599)	39.8 (366)	55.6 (564)	46.8 (482)	52.9 (494)	56.4 (589)	59.7 (608)
Once a week	13.2 (138)	13.5 (141)	22.1 (203)	8.8 (89)	11.8 (122)	18.2 (170)	10.9 (114)	13.9 (142)
2-3 times a week	8.8 (92)	6.1 (64)	8.4 (77)	5.0 (51)	7.0 (72)	7.1 (66)	5.5 (57)	5.6 (57)
4-6 times a week	5.4 (57)	2.1 (22)	4.8 (44)	3.1 (31)	2.8 (29)	2.9 (27)	1.9 (20)	2.7 (28)
Every day	2.5 (26)	1.1 (11)	2.3 (21)	1.9 (19)	1.3 (13)	3.1 (29)	1.1 (12)	2.4 (24)
Responsible for shopping								
No	1.7 (18)	2.9 (30)	1.7 (16)	5.3 (54)	4.1 (42)	2.0 (19)	3.5 (37)	3.1 (31)
Shared	21.1 (220)	29.8 (311)	29.8 (274)	28.1 (284)	42.2 (434)	28.3 (264)	36.1 (377)	29.9 (303)
Yes	77.2 (807)	67.3 (701)	68.5 (630)	66.6 (673)	53.7 (552)	69.7 (650)	60.3 (630)	67.1 (680)
Responsible for cooking								
Never	4.9 (51)	6.3 (66)	5.3 (49)	7.0 (71)	6.8 (70)	5.2 (49)	5.0 (52)	6.8 (69)
1 or 2 times per week	20.5 (215)	20.5 (214)	21.7 (200)	21.3 (216)	24.2 (250)	16.8 (157)	16.3 (170)	17.4 (177)
3 or 4 times per week	17.5 (183)	24.7 (257)	17.5 (161)	20.1 (204)	24.0 (247)	24.9 (233)	24.9 (260)	20.0 (204)
5 or 6 times per week	14.3 (150)	16.6 (173)	13.4 (123)	20.3 (206)	13.6 (140)	1.4 (13)	16.0 (167)	18.4 (187)
Every day	42.8 (448)	31.9 (332)	42.1 (387)	31.3 (317)	31.4 (324)	51.6 (482)	37.9 (396)	37.5 (382)

PAPER 2: Table 2. *Type of data generated: % respondents (number of respondents).*

Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Search data								
Searching for information on how to prepare food (e.g. recipes, information about cooking techniques, ingredients)	89.0 (932)	84.9 (885)	91.2 (839)	82.94 (841)	94.18 (971)	90.04 (841)	89.57 (936)	80.67 (822)
Searching for places to eat or drink (e.g. restaurant)	76.2 (798)	79.9 (832	84.6 (778)	78.60 (797)	87.00 (897)	82.12 (767)	79.62 (832)	75.56 (770)
Searching for offers on food or drinks (e.g. discount vouchers)	72.0 (754)	70.5 (735	54.9 (505)	74.46 (755)	85.55 (882)	76.55 (715)	71.10 (743)	72.23 (736)
Searching for places that sell foods or drink	68.5 (717)	69.4 (723)	78.5 (722)	71.40 (724)	84.09 (867)	73.34 (685)	71.29 (745)	71.54 (729)
“Opinion sharing” data								
Sharing views on places to eat or drink (e.g. posting restaurant reviews on the internet)	61.9 (648)	53.4 (556)	73.4 (675)	49.6 (503)	69.5 (717)	65.4 (611)	49.1 (513)	49.0 (499)
Sharing views on foods and ingredients (e.g. post reviews on the internet)	55.1 (577)	47.3 (493)	65.9 (606)	42.6 (432)	66.9 (690)	58.1 (543)	43.9 (459)	42.2 (430)
Sharing views on cooking techniques (e.g. posting recipes or clips on how to prepare food)	49.0 (513)	43.7 (455)	61.7 (568)	40.2 (408)	63.9 (659)	53.4 (499)	40.6 (424)	40.5 (413)
Sharing views on recipes (e.g. posting reviews on the internet)	50.1 (525)	46.4 (483)	64.2 (591)	41.7 (423)	68.5 706)	55.3 (516)	42.1 (440)	41.6 (424)
Posting comments about food and/or drink on social media	48.1 (504)	39.4 (411)	66.2 (609)	47.4 (481)	70.5 727)	59.0 (551)	47.9 (500)	46.3 (472)
Posting videos of food and/or drink on social media (e.g. YouTube, Facebook, Twitter)	41.0 (429)	35.8 (373)	54.9 (505)	43.9 (445)	55.9 (576)	44.7 (417)	36.4 (380)	36.7 (374)
Posting photos of food and/or drink on social media (e.g. Facebook, Twitter)	46.3 (485)	40.4 (421)	63.7 (586)	51.3 (520)	69.2 (713)	58.6 (547)	52.6 (550)	45.5 (464)
“Planning and buying” data								
Booking places to eat (e.g. restaurants)	62.3 (652)	59.9 (624)	77.3 (711)	76.1 (772)	66.5 (686)	71.4 (667)	61.2 (639)	65.1 (663)
Comparing food/drink products and prices	65.8 (689)	65.9 (687)	75.8 (697)	67.4 (683)	74.6 (769)	70.3 (657)	67.6 (706)	67.4 (687)
Buying food or drinks (e.g. online grocery shopping/takeaways)	55.2 (578)	50.3 (524)	64.2 (591)	55.0 (558)	55.4 (571)	61.1 (571)	53.3 (557)	63.1 (643)
Creating shopping lists	49.5 (518)	46.6 (485)	59.5 (547)	49.4 (501)	55.2 (569)	58.6 (547)	51.6 (539)	47.7 (486)
Planning menus/meals	41.2 (431)	42.3 (441)	50.4 (464)	40.9 (415)	46.1 (475)	55.7 (520)	48.5 (507)	39.9 (407)
“Record keeping” data								
Creating or managing recipe collections	50.5 (529)	47.1 (491)	54.8 (504)	51.9 (526)	63.2 (652)	61.6 (575)	51.7 (540)	42.4 (432)
Keeping records of money spend on food/drinks (i.e. budgeting)	40.9 (428)	37.3 (389)	59.0 (543)	46.7 (473)	49.1 (506)	38.0 (355)	42.5 (444)	40.8 (416)
Keep records of food/drink stores in the household (e.g. what’s in the freezer)	33.5 (351)	28.8 (300)	42.2 (388)	30.5 (309)	36.0 (371)	38.0 (355)	27.6 (288)	31.8 (324)

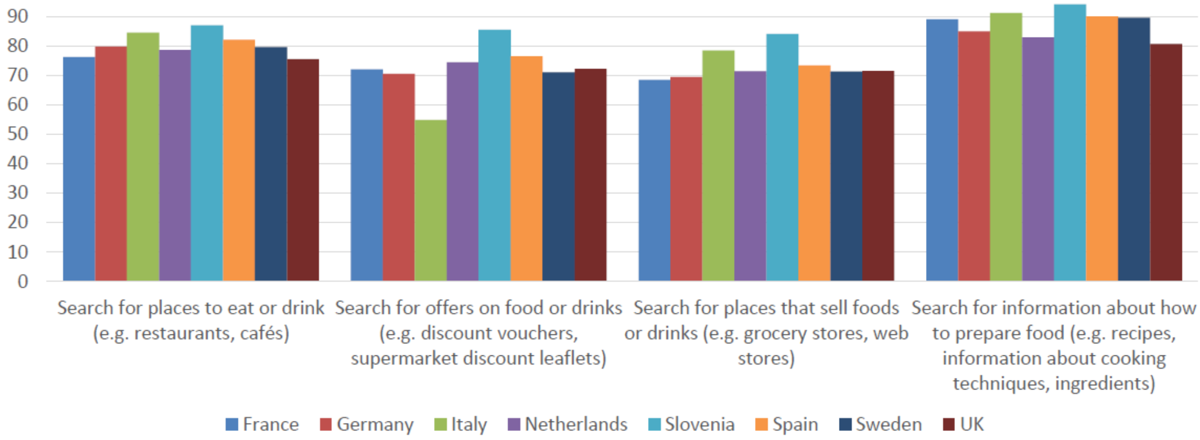
Variable % (N)	France	Germany	Italy	Netherlands	Slovenia	Spain	Sweden	UK
Keep records of what you eat and drink	33.1 (347)	31.0 (323)	43.3 (398)	34.3 (348)	31.0 (320)	33.9 (317)	30.1 (314)	33.9 (345)
Keep records of what people in your household eat or drink	31.4 (329)	26.2 (273)	39.7 (365)	27.6 (280)	30.4 (313)	33.0 (308)	22.6 (236)	29.6 (302)
"Location" data generated using smart devices that record when, where or how food is....								
Purchased (e.g. location data from your mobile phone, Wi-Fi login in shops or restaurants)	32.1 (336)	18.7 (195)	34.5 (317)	22.6 (229)	23.8 (245)	33.2 (310)	24.8 (259)	25.8 (263)
Prepared (e.g. your mobile phone, kitchen appliances that record what you are doing e.g. food weighing scales or pans that record temperatures, smart meters that record how much gas, electricity or water you have used)	29.04 (304)	18.4 (192)	27.7 (255)	20.9 (212)	31.4 (324)	22.5 (210)	20.7 (216)	22.5 (229)
Eaten (e.g. location data from your mobile phone, sensors in kitchenware)	28.3 (296)	16.4 (171)	23.7 (218)	16.9 (171)	29.1 (300)	21.7 (203)	15.6 (163)	19.6 (200)

PAPER 2: Table 3. *Means and standard deviations of these variables across the eight countries*

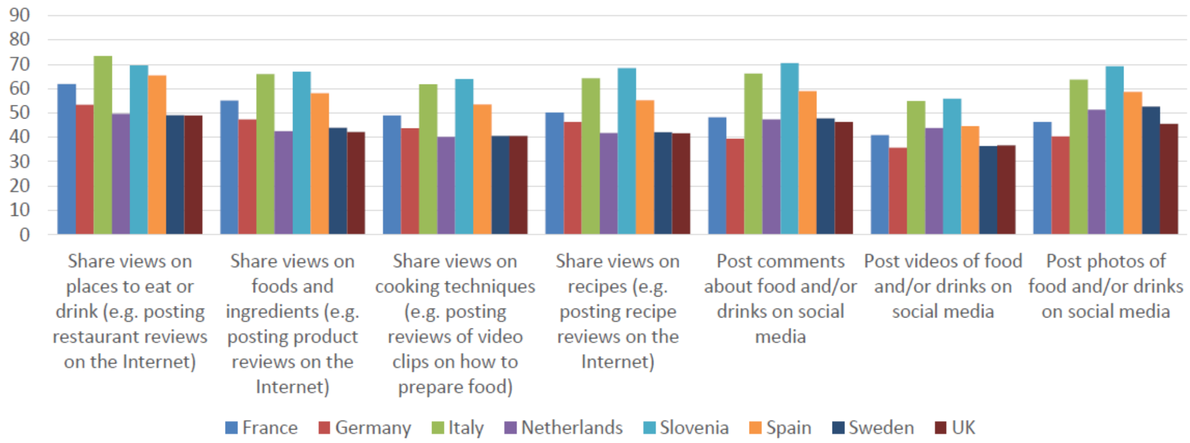
Variable	France		Germany		Italy		Netherlands		Slovenia		Spain		Sweden		UK	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	44.69	15.11	44.64	14.97	39.88	13.48	45.06	15.17	43.01	13.63	39.32	13.08	45.02	15.91	45.40	15.71
BMI	24.51	5.09	25.92	5.40	24.62	4.45	25.51	5.59	26.11	5.50	25.11	4.25	25.75	5.07	26.04	5.20
Willingness to share with universities[1]	3.36	1.09	3.31	1.08	3.65	.88	3.17	1.08	3.85	.81	3.63	.89	3.44	1.03	3.57	1.02
Willingness to share with governments[a]	3.00	1.20	2.87	1.18	3.40	.97	2.91	1.10	3.53	.99	3.29	1.03	3.21	1.12	3.26	1.15
Willingness to share with companies[a]	3.09	1.12	3.08	1.12	3.41	.98	2.95	1.11	3.72	.86	3.40	.95	3.15	1.04	3.31	1.06
Trust in universities[b]	3.38	.80	3.33	.75	3.48	.68	3.28	.65	3.67	.76	3.48	.68	3.43	.77	3.52	.80
Trust in governments[b]	3.10	.99	3.00	.94	3.37	.83	3.09	.82	3.28	.88	3.11	.90	3.26	1.01	3.18	.91
Trust in companies[b]	3.06	.80	3.07	.74	3.25	.68	3.00	.70	3.37	.82	3.15	.69	2.97	.77	3.12	.87
Perceived risk in sharing data with universities[c]	3.22	.75	3.10	.69	3.22	.65	3.09	.64	2.99	.69	3.19	.66	3.05	.69	3.00	.73
Perceived risk in sharing data with governments[c]	3.26	.71	3.26	.74	3.04	.65	3.11	.66	3.16	.70	3.22	.75	3.00	.79	3.20	.71
Perceived risk in sharing data with companies[c]	3.11	.79	3.05	.72	3.17	.65	2.98	.71	3.12	.66	3.10	.65	2.96	.71	3.25	.68
Perceived health[d]	3.47	.86	3.58	.87	3.70	.71	3.59	.79	3.62	.86	4.06	.51	3.53	.96	3.67	.85
Health interest[e]	3.33	.71	3.03	.72	.	.	3.25	.67	3.31	.65	3.49	.65	3.16	.75	3.24	.69
Data privacy concerns	3.36	1.18	3.47	1.06	2.89	.98	3.08	1.07	2.92	1.21	3.42	1.05	2.75	1.14	3.17	1.21
Science mostly good	5.72	1.68	6.66	1.99	7.11	1.74	6.64	1.61	5.65	2.17	7.23	1.73	7.01	1.98	6.96	1.93
Moral motives	3.14	.83	2.93	.85	2.97	.75	2.87	.76	3.39	.74	3.34	.70	3.09	.82	3.10	.76
Amount of data	14.02	9.11	12.77	8.69	16.20	8.74	13.48	8.64	15.69	7.63	15.11	8.66	13.11	8.12	13.19	9.19
Take-away (per week)[f]	2.06	1.23	2.43	1.21	2.23	1.27	2.19	1.19	1.91	1.13	2.36	1.38	2.17	1.02	2.31	1.15
Ready meals (per week) [f]	2.30	1.33	2.31	1.15	2.22	1.28	2.07	1.23	1.86	1.07	2.28	1.38	2.11	1.13	2.50	1.29
Eating out (per week)	2.28	1.26	2.17	.95	2.40	1.19	2.10	1.05	2.09	1.07	2.38	1.11	2.08	.96	2.27	1.04
Responsible for shopping[g]	2.76	.47	2.64	.54	2.67	.51	2.61	.59	2.50	.58	2.68	.51	2.57	.56	2.64	.54
Responsible for cooking[h]	2.30	1.33	2.53	1.30	2.35	1.35	2.52	1.31	2.61	1.33	2.23	1.36	2.34	1.27	2.38	1.32
Cooking time (minutes per week)	215.08	162.44	282.19	184.92	259.21	190.29	255.89	151.73	452.46	282.95	293.43	232.09	271.70	172.53	245.66	171.79

[a] Scale used: 1 (I definitely would not share…) to 5 (I definitely would share…)
[b] Scale used: 1 to 5; higher scores reflect a higher level of trust
[c] Scale used: 1 to 5; higher scores reflect a higher perceived risk
[d] Scale used: 1 (very bad) to 5 (very good)
[e] Scale used: 1 to 5; higher scores reflect more interest in the healthiness of food

[f] Scale used: Never (1); Less than once a week (2); Once a week (3); 2-3 times a week (4); 4-6 times a week (5); Every day (6)
[f] Scale used: No (1); Shared (2); Yes (3)
[g] Scale used: Never (1); 1 or 2 times per week (2); 3 or 4 times per week (3);5 or 6 times per week (4); Every day (5)



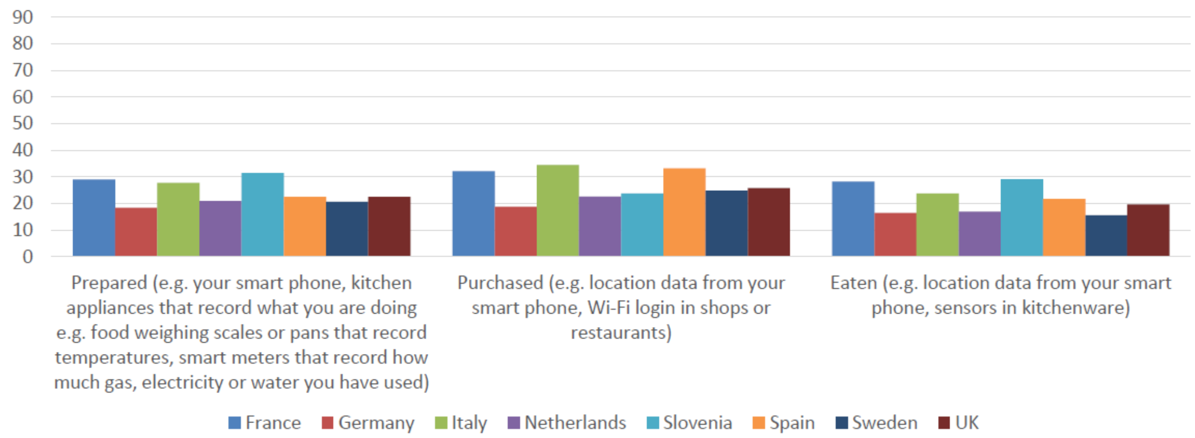
PAPER 2: Figure 2. *Type of data generated: % respondents producing search data.*



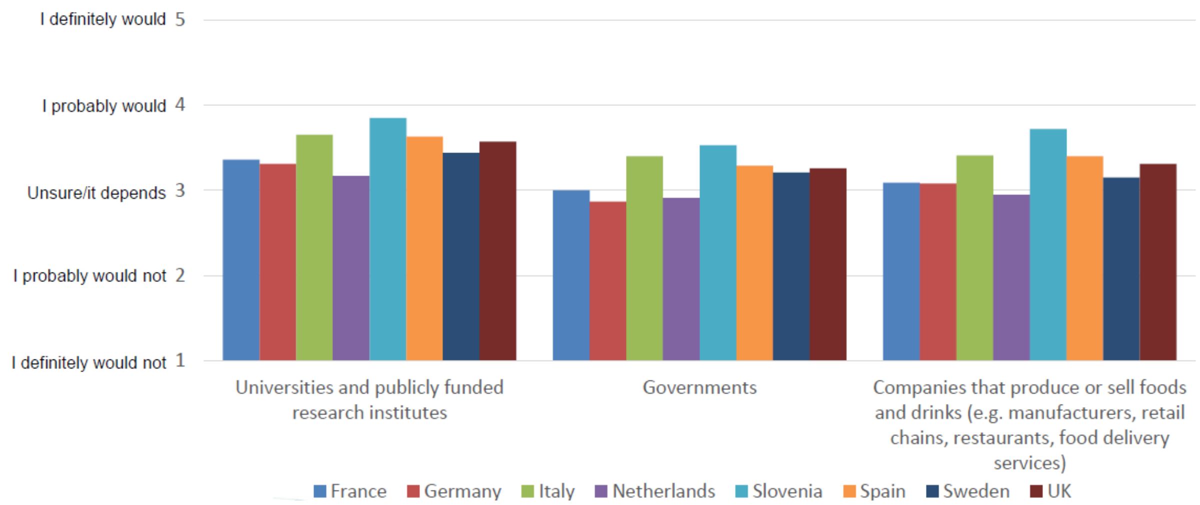
PAPER 2: Figure 3. *Type of data generated: % respondents producing "opinion sharing" data.*

PAPER 2: Figure 4. *Type of data generated: % respondents producing "planning and buying" data.*

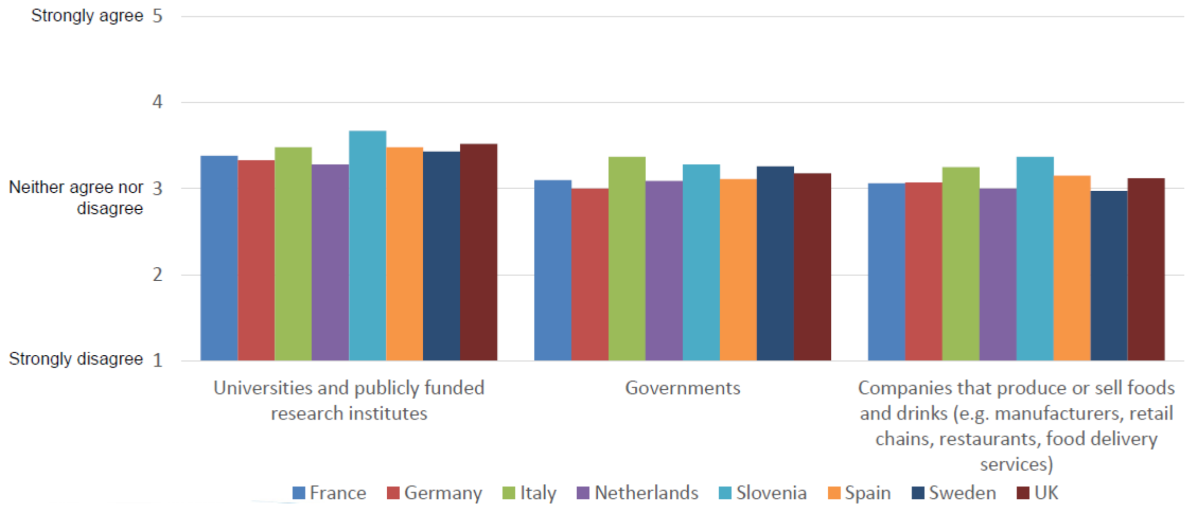
PAPER 2: Figure 5. *Type of data generated: % respondents producing "record keeping" data.*



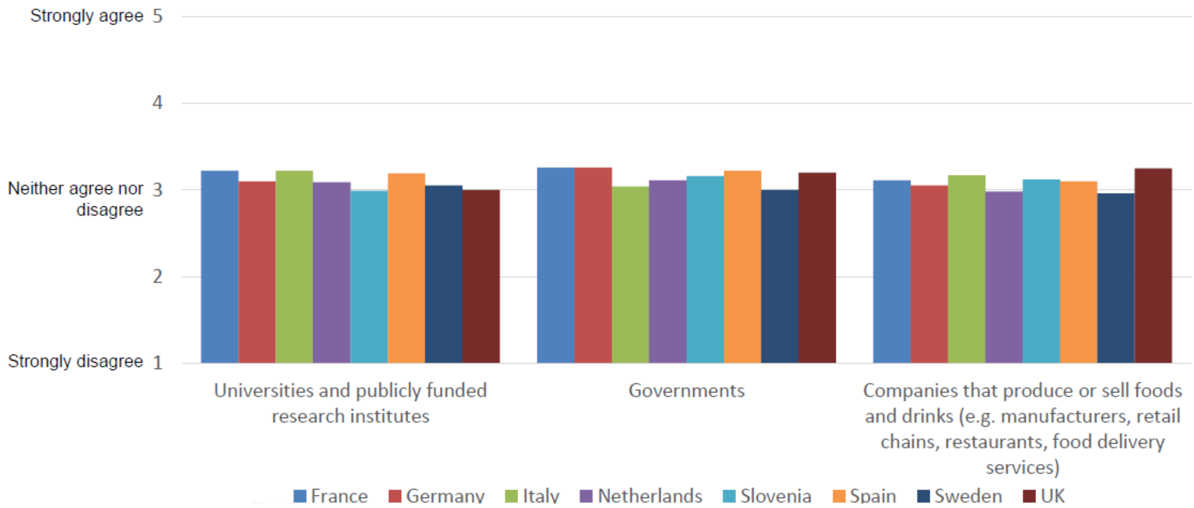
PAPER 2: Figure 6. *Type of data generated: % respondents producing "location" data generated using smart devices that record when, where or how food is....*



PAPER 2: Figure 7. *Respondents' willingness to share data.*



PAPER 2: Figure 8. *Respondents' trust in ability to handle data.*



PAPER 2: Figure 9. *Respondents' perceived risk in sharing data.*

PAPER 2: Table 4. *Correlates of willingness to share data.*

	Share: universities			Share: governments			Share: companies		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Trust in universities	.605	.000	7744	.511	.000	7739	.527	.000	7725
Perceived risk in sharing data: Universities	.046	.000	7744	.081	.000	7739	.093	.000	7725
Trust in governments	.447	.000	7744	.619	.000	7739	.481	.000	7725
Perceived risk in sharing data with governments	-.192	.000	7744	-.306	.000	7739	-.195	.000	7725
Trust in companies	.436	.000	7744	.493	.000	7739	.581	.000	7725
Perceived risk in sharing data with companies	.167	.000	7744	.183	.000	7739	.211	.000	7725
Perceived health	.126	.000	7697	.116	.000	7692	.111	.000	7678
BMI	.026	.037	6609	.013	.287	6604	.030	.014	6593
Age	-.073	.000	7744	-.081	.000	7739	-.104	.000	7725
Health interest	.100	.000	6837	.088	.000	6831	.065	.000	6817
Data privacy concerns	-.224	.000	7739	-.281	.000	7734	-.253	.000	7720
Science mostly good	.207	.000	7744	.165	.000	7739	.168	.000	7725
Income	.046	.166	916	.050	.132	916	.052	.113	916
Log Income	.041	.229	876	.055	.101	876	.016	.632	876
Moral motives	.446	.000	7744	.494	.000	7739	.506	.000	7725
Gender (binary)	.031	.007	7712	.023	.041	7707	.034	.003	7693
Amount of data	.210	.000	7744	.263	.000	7739	.301	.000	7725
Responsible for cooking	-.057	.000	7744	-.034	.003	7739	-.038	.001	7725
Responsible for shopping	.059	.000	7729	.054	.000	7724	.059	.000	7710
Cooking time (per week)	.107	.000	7550	.073	.000	7546	.099	.000	7533
Take-away (per week)	.088	.000	7744	.132	.000	7739	.146	.000	7725
Ready meals (per week)	.086	.000	7744	.136	.000	7739	.145	.000	7725
Eating out (per week)	.118	.000	7744	.172	.000	7739	.168	.000	7725

PAPER 2: Table 5. *Correlates of willingness to share data.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1: Trust: Universities																						
2: Perceived risk in sharing data with universities	.230																					
3: Trust: Governments	.613	.149																				
4: Perceived risk in sharing data with governments	-.176	.306	-.384																			
5: Trust: Companies	.617	.266	.616	-.121																		
6: Perceived risk in sharing data with companies	.264	.498	.235	.207	.394																	
7: Perceived health	.142	.056	.142	-.036	.125	.067																
8: BMI	.008	.007	-.010	.010	.000	.008	-.170															
9: Age	-.106	-.093	-.095	.007	-.149	-.077	-.188	.190														
10:Health interest	.097	-.040	.067	-.025	.016	-.007	.085	-.069	.172													
11: Data privacy concerns	-.257	.016	-.345	.274	-.294	-.042	-.061	.004	.083	.099												
12: Science mostly good	.269	-.022	.223	-.137	.153	.034	.153	.013	.001	.064	-.070											
13: Income	.059	.087	.063	.021	.069	.083	.050	-.040	-.033	-.011	-.057	.050										
14: Log Income	.036	.030	.062	-.012	.025	-.023	.067	-.022	-.002	.063	.028	.077	.393									
15: Moral motives	.527	.222	.497	-.126	.557	.257	.126	.026	-.098	.168	-.217	.168	.080	.075								
16: Gender (binary)	-.001	-.073	.005	-.027	.022	-.040	-.020	-.146	-.012	.138	-.001	-.044	-.032	-.122	-.023							
17: Amount of data	.272	.231	.289	-.025	.369	.220	.104	-.029	-.381	.033	-.121	.044	.051	.061	.387	-.045						
18: Responsible for cooking	-.030	.018	-.013	.020	-.018	-.002	-.004	.052	-.118	-.156	-.023	-.049	-.033	.028	-.021	-.337	.037					
19: Responsible for shopping	.055	.046	.059	-.011	.065	.055	.026	-.028	.080	.070	.006	.040	.021	.022	.041	.238	.059	-.491				
20: Cooking time (per week)	.076	-.059	.019	-.014	.042	-.021	.018	.049	.065	.099	.038	.009	-.027	.042	.053	.102	-.013	-.152	.015			
21: Take-away (per week)	.151	.208	.175	.025	.236	.195	.046	.013	-.326	-.203	-.124	.009	.102	.050	.206	-.150	.380	.153	.002	-.171		
22: Ready meals (per week)	.145	.193	.175	.028	.231	.199	-.002	.000	-.231	-.195	-.101	.015	.087	-.023	.203	-.151	.327	.120	.028	-.206	.648	
23: Eating out (per week)	.185	.208	.216	.017	.252	.218	.107	-.022	-.248	-.050	-.125	.029	.113	.105	.241	-.137	.402	.142	.007	-.123	.583	.527

PAPER 2: 3.3 MODEL BUILDING

Recall that a random sample representing 20% of the total data was selected for the purpose of exploratory analyses. This sample contains 1689 participants.

We aimed to develop a 2-level model, with individual participants on Level 1 and countries on Level 2. Two variables Level 2 predictors have been explored: country rankings based on gross domestic product per capita (GDPpc), and country rankings based on the percent of people who said they trusted their co-nationals in the World Values Survey. Country-level means for individual variables and country rankings are strongly correlated. To avoid multicollinearity issues (i.e. the phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy), and given the small number of countries, only the country ranking based on GDPpc was used in the models below.

Three separate models were built to predict willingness to share data with universities, with governments and with companies, respectively. All individual-level predictors were mean-centred within each country. To reach a parsimonious model, a backward elimination approach was employed. Saturated models (containing all predictors with random slopes and intercepts) failed to converge. Therefore, models with fixed slopes and a random intercept were taken as a starting point *(full model)*:

> *willingness to share data ~ Crank + trust + perceived risk in sharing data + attitudes to science + concern + moral motives + perceived health + health interest + age + education + self-transcendence + self-conservation + (1 | country)*

Backward elimination was performed with a threshold of $p < .001$ for fixed effects and $p < .10$ for random effects. The same model was reached for all three stakeholders *(final model)*:

> *willingness to share data ~ trust + perceived risk in sharing data + moral motives + (1 | country)*

This model was compared to the full model, a null model containing only a random intercept,

> *willingness to share data ~ 1 + (1 | country)*

and a model similar to the final model but also including random slopes for all three predictors *(expanded model)*:

> *willingness to share data ~ trust + perceived risk in sharing data + moral motives +(0+trust|country)+(0+ perceived risk in sharing data|country)+ (0+model|country)+(1 | country).*

See Table 6 for the results. The final model fit the data better than the null model, and as well as the expanded model. The final model did not fit the data as well as the full model, but the difference in explained variance was within 1%.

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

PAPER 2: Table 6. *Results for the Model Building Stage*

		Universities	Governments	Companies
β of the final model	Trust	0.444	0.425	0.389
	Perceived risk in sharing data	-0.151	-0.100	-0.107
	Moral motives	0.218	0.254	0.258
Final model vs null	χ^2 (3)	890.08, $p < .001$	903.71, $p < .001$	817.74 , $p < .001$
Final model vs full	χ^2 (9)	14.96, $p = .092$	27.09, $p = .001$	33.62, $p < .001$
Final model vs expanded	χ^2 (9)	7.64, $p = .054$	1.44, $p = .697$	0.34, $p = .953$
Explained variance: final model (full model)	R^2	.46 (.46)	.46 (.47)	.44(.45)

Therefore, the final model was retained. In this model, trust had a medium-to-large positive effect on the willingness to share data; moral motives had a small-to-medium positive effect; and perceived risk had a small negative effect. The model explained almost half of the variance of willingness to share data.

The final model was tested on the full data set (Table 7). It fit the data well and explained a large portion of the variance.

PAPER 2: Table 7. *Results for the Model Confirmation Stage*

		Universities	Governments	Companies
β of the final model	Trust	.499	.433	.405
	Perceived risk in sharing data	-.118	-.100	-.030
	Moral motives	.210	.255	.279
Final model vs null	χ^2 (3)	3391.4, $p < .001$	3844.9, $p < .001$	3184.0, $p < .001$
Explained variance: final model (full model)	R^2	.42	.46	.41

PAPER 2: 4. DISCUSSION

The current study set off with an important realisation that any promise of linking large consumer-generated data for the purpose of research, innovation or policy development, necessitates better undestanding of data subjects' willingness to share their personal data for these diverse purposes. We also highlighted that there are fundamental, as of yet unexplored differences between publics' willingness to share medical and health data for the purpose of health delivery on one hand, and the the willigness to share the loosely defined "lifestyle" data for the variety of purposes that may not have any tangible or immediate benefits for the individual consumer, on the other. Our core concern in this survey therefore has been to describe the current attitudes, beliefs and intentions to share life-style (more specifically, food-related) data drawn from a representative sample across Europe, and to establish a reliable model of factors that may explain this willingness. These two questions – the type of data that is considered "sharable" and the kind of factors that underpin willingness to share - are a critical aspect of user needs research needed to develop a workable technical, governance and business model for RICHFIELDS.

Overall, our survey shows that across the EU, a wide range of food-related data is being generated by consumers, and that there are differences between countries with respect to the kind and range of data produced. Whilst this descriptive information in itself is likely to change over short term due to

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

the rapid advances in technology capable of capturing food-related information across the EU, it nevertheless gives us an immediate sense of the potential of consumer-generated data to provide value to RICHFIELDS. Search data is the most likely data to be in abundant supply and yet, it holds relatively limited value for the study of the actual behavior. On the other hand, the data on food intake and the relevant context data (e.g. location) is relatively harder to come by, a particular challenge in the context of the need for contextualized intake data for the purpose of RICHFIELDS. Nevertheless, it is important to understand that the linking of these different types of data may achieve the granularity necessary for breakthrough science of consumer food-related behavior - which RICHFIELDS would need to address in the next phases of its development.

The examination of people's reported willingness to share data showed an interesting pattern: we recorded an above average willingness to share data (above 3, on a 5-point scale) with universities - for the purpose of science and public research - across all countries. People were simultaneously slightly less willing to share their data with government and industry though this pattern only showed weak statistical significance. Whilst the result is heartening as it demonstrates the publics' continued belief that science has an intrinsic value as a societal endeavour that deserves publics' support, it also highlights the need to clearly articulate the purpose to which consumer generated data is put and the way in which it links with data from other sources.

Exploring in greater depth this premise that science has an intrinsic value, we captured three important variables: trust, moral motives and perceived risk. Our model that examined the relative weighting of these factors in predicting willingness to share showed that almost half of the variance of willingness to share data is explained by these three variables:

- trust had a medium-to-large positive effect on the willingness to share data;
- moral motives had a small-to-medium positive effect;
- perceived risk had a small negative effect.

The three constructs are important as they underline the ethical dimension of data sharing decisions and the need for RICHFIELDS data platform to be explicit about its commitment to these values. Trust, perception of risk and moral motives are closely linked with the issue of data governance and the respect for privacy, confidentiality and consent. Data linkages that RICHFIELDS is proposing, would typically enable identification of a consumer, even if we strive to ensure anonymity and de-personalisation. Against this context it is important to address how more value from data can be extracted without compromising the citizen's right to privacy (recognised by European Conventions of Human Rights), confidentiality and the role of consent within the matrix of big data and privacy whilst keeping in sight the protection imparted to the individual (data subject) by EU's General Data Protection Regulation (*Regulation 2016/679*) that came into force in May 2018.

These concepts also are relevant within the broader debate about how to manage competing interests of science and data donors/subjects (citizens). These are currently resolved through a combination of standard operation procedures and good scientific practices guidelines, creation of ethics advisory board, and regulation of financial gains from the data/IPR ownership (Royal Society/British Academy, 2017).

However, big lifestyle-related data research infrastructures are not only research resources but also provide valuable opportunities in terms of 'new economy' (i.e. employment, entrepreneurship, knowledge creation). The question of who owns lifestyle-related big data and has access to the linked data, therefore, is simultaneously an issue of economic development and international standing, as well as research. Fairness, legitimacy and due process are important considerations integral to any decisions about ownership and commercialization. Coupled with this is the **ethical issue** of broader

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

societal value of who has the right to commodify the information based on linked lifestyle-related data and if it should be rightfully 'owned' by anyone, shifting the discussion away from economic sphere, towards the human rights domain (and the associated legal frameworks).

PAPER 2: 5. CONCLUSIONS

In the light of these findings, the possible considerations for the future of data-driven science need to address the following issues for the purpose of developing RICHFIELDS data platform:

- The research infrastructure that wishes to make use of consumer-generated data will need to identify appropriate means of maintaining trust, minimising risk to indivdiual and society and enhancing the perceived moral authority of science. In a nutshell, these endeavours could be achieved throuh appropriate governance and technical frameworks, but perhaps more importantly, through the engagement with public and constant communication that would develop a strong moral identity for the research infrastructure in this domain.
- The research infrastructure should closley observe the recommendations of how to achieve ethical design for the future (e.g. please see Carr (2018)). Given RICHFIELDS' purpose is to use the data sets in its repository for research, pseudonymisation is suggested as a means to process the data, provided appropriate safeguards are in place. In order to raise the integrity profile of RICHFIELDS externally the setting up of an independent ethics committee is also suggested. This would also help in bolstering the confidence of data in the utility of research infrastructures such as RICHFIELDS as a research tool for promoting well-being and over time might usher in an era where data subjects in the spirit of altruism give their data for the sake of research and innovation.
- A more nuanced understanding of the way in which the publics perceive the possible solutions and models for RICHFIELDS needs to be obtained through the use case studies and validation of our business, govenrance and technical models.
- The research infrastructure must be mindful of possible cross-country differences in sensitivities about the issue of data sharing. This necessitates constant monitoring of public attitudes to privacy, science and their perceptions of the food system. The latter is particular apposite in the context of the publics' increasing awareness of the unsustainability of the current food system, and the growing calls for science to engage ethically with the food-related issues that are fundamental not only to the health of individuals, but also to the health and livelihood of our planet.

PAPER 2: REFERENCES

Abbar, S., Mejova, Y., Weber, I. (2014) You tweet what you eat: studying food consumption through Twitter, *CoRR abs*/1412.4361

Bellekens, X., Hamilton, A., Seeam, P., Nieradzinska, K., Franssen, Q. Seeam, A. (2016) Pervasive eHealth services a security and privacy risk awareness survey. In *2016 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (CyberSA),* pp. 1-4.

Bennett, C., Raab, C. D., Revisiting (2018) The Governance of Privacy': Contemporary Policy Instruments in Global Perspective. Original paper Prepared for the Privacy Law Scholars Conference, Berkeley CA, June 1-2, 2017. Revised version forthcoming in "*Regulation and Governance*".

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

Bhattacherjee, A. (2002) Individual Trust in Online Firms: Scale Development and Initial Test, *Journal of management Information Systems,* 19, 211-241.

Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., Sheehan, J., Patrick, K. (2016) Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association*, 23(e1), 42–48.

Carr, I. (2018) RICHFIELDS *Deliverable 13.2. Ethical Design at the Interface of Ethics for Big Data and the European Union's General Data Protection Regulation*.

Chen, J., Bauman, A., & Allman-Farinelli, M. (2016) A study to determine the most popular lifestyle smartphone applications and willingness of the public to share their personal data for health research. *Telemedicine and e-Health*, 22, 655–665.

Dinev, T., Albano, V., Xu, H., D'Atri, A. and Hart, P. (2016) Individuals' attitudes towards electronic health records: A privacy calculus perspective. In *Advances in healthcare informatics and analytics* (pp. 19-50). Springer, Cham.

Ernsting, C., Dombrowski, S. U., Oedekoven, M., O Sullivan, J. L., Kanzler, M., Kuhlmey, A., Gellert, P. (2017) Using smartphones and health apps to change and manage health behaviors: A population-based survey. *Journal of Medical Internet Research,* 19(4), e101.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., Bell, D. (2014) Analyzing the language of food on social media, in *IEEE International Conference on Big Data.*

Hieke, S., Palascha A., Jola, C., Wills, J., Raats, M. M. (2016) The pack size effect: Influence on consumer perceptions of portion sizes. *Appetite* 96, 225–238.

Jarvenpaa, S. L., Tractinsky, N., Saarinen, L. (1999) Consumer Trust in an Internet Store: a Cross-Cultural Validation, *Journal of Computer-Mediated Communication*, 5(2).

Kim, K. K., Joseph, J. G., Ohno-Machado, L. (2015) Comparison of consumers' views on electronic data sharing for healthcare and research. *Journal of the American Medical Informatics Association,* 22(4), 821–830.

Lavelle, F., McGowan, L., Hollywood, L., Surgenor, D., McCloat, A., Mooney, E,. Caraher, M., Raats, M., Dean, M. (2017) The development and validation of measures to assess cooking skills and food skills, *International Journal of Behavioral Nutrition and Physical Activity*. 14, 1, 118.

Lavelle, F., Spence, M., Hollywood, L., McGowan, L., Surgenor, D., McCloat, A., Mooney, E., Caraher, M., Raats, M.,Dean, M. (2016) Learning cooking skills at different ages: a cross-sectional study, *International Journal of Behavioral Nutrition and Physical Activity*. 13, 11 p., 119.

McGowan, L., Pot, G. K., Stephen, A. M., Lavelle, F., Spence, M., Raats, M., Hollywood, L., McDowell, D., McCloat, A., Mooney, E., Caraher, M., Dean, M. (2016) The influence of socio-demographic, psychological and knowledge-related variables alongside perceived cooking and food skills abilities in the prediction of diet quality in adults: a nationally representative cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*. 13, 13 p., 111.

Mejova, Y., Haddadi, H., Noulas, A., Weber I. (2015) #foodporn: Obesity patterns in culinary interactions, in *Proceeding DH '15 Proceedings of the 5th International Conference on Digital Health*, pp 51-58.

MRC Ipsos MORI (2007) The Use of Personal Health Information in Medical Research General Public Consultation.

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

Nilsen, W., Kumar, S., Shar, A., Varoquiers, C., Wiley, T., Riley, W. T., Pavel, M.,Atienza, A. A. (2012) Advancing the science of mHealth. *Journal of Health Communication*, 17(sup1), 5-10.

Patil, S., Lu, H., Saunders, C.L., Potoglou, D., Robinson, N. (2016) Public preferences for electronic health data storage, access, and sharing – evidence from a pan-European survey. *Journal of the American Medical Informatics Association* 2016;0:1–11.

Pevnick, J. M., Fuller, G., Duncan, R., Spiegel, B. M. R. (2016) A large-scale initiative inviting patients to share personal fitness tracker data with their providers: Initial results. *PloS one*, 11, e0165908.

Prasad, A., Sorber, J., Stablein, T., Anthony, D., Kotz, D. (2014) Understanding user privacy preferences for mhealth data sharing. In S. Adibi (Ed*.), mHealth: Multidisciplinary verticals* (pp. 545–569). Boca Raton, LA: CRC Pres.

Quiroz-Aitken, M., De St. Jorre, J., Pagliari, C., Jepson, R., Cunningham-Burley, S. (2016) Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics* 17, 1, 73

Raats M. M., Hieke, S., Jola, C., Hodgkins, C., Kennedy, J., Wills, J. (2015) Reference amounts utilised in front of pack nutrition labelling; impact on product healthfulness evaluations. *European Journal of Clinical Nutrition* 69(5): 619-625 16.

Roininen, K., Lähteenmäki, L., Tuorila, H. (1999) Quantification of consumer attitudes to health and hedonistic characteristics of foods. *Appetite*, 33, 71-88.

Royal Society/British Academy (2017) *Data Management and Use: Governance in the 21st Century*; and *Data Governance: Landscape Review*

Seifert, A., Christen, M., Martin, M. (2018) Willingness of older adults to share mobile health data with researchers. GeroPsych, 31, 41-49.

Sharma S., De Choudhury, M. (2015) Detecting and characterizing nutritional information of food and ingestion content in Instagram, in *Proceeding WWW '15 Companion Proceedings of the 24th International Conference on World Wide Web*, pp 115-116

Skatova, A., Ng, E., Goulding, J. (2014) Data Donation: Sharing Personal Data for Public Good. *Digital Economy 2014: Application of Digital Innovation*, 2014, London, UK.

Sleigh, J. (2018) Experiences of Donating Personal Data to Mental Health Research: An Explorative Anthropological Study. *Biomedical Informatics Insights*.

Sobal, J. (1991) Obesity and socioeconomic status: A framework for examining relationships between physical and social variables. *Medical Anthropology*, 13, 231–247.

Sobal, J., Bisogni, C. A., Jastran, M. (2014) Food choice is multifaceted, contextual, dynamic, multilevel, integrated, and diverse. *Mind, Brain, and Education*, 8: 6–12.

Stockdale J, Cassell J and Ford E (2018) "Giving something back": A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland [version 1; referees: 2 approved with reservations]. *Wellcome Open Research*, 3:6

Velasco, E., Agheneza, T., Denecke, K., Kirchner, G., Eckmanns, T. (2014) Social media and internet-based data in global systems for public health surveillance: a systematic review *Milbank Quarterly* 92, 7

Weber, I., Achananuparp, P. (2016) Insights from machine-learned diet success prediction, in *Pacific Symposium on Biocomputing (PSB) 2016*.h

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

WHO (2011) *Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants*.

Wilton, R. (2015) The data-driven world doesn't run on data: it runs on trust. *Internet Society.*

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