Horizon 2020

INFRADEV-1-2014 - Design studies

RICHFIELDS Working Package 6

Deliverable D6.3

List of quality criteria

Development of a Quality Evaluation Framework for Consumer-Generated Domestic Food Preparation Data

Date delivered:
M32

Authors:

Naomi Klepacz, Marcus Maringer, Susanne Ekman, Anne Norman, Muriel Verain, Anouk Geelen, Monique Raats

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:	List of quality criteria - Development of a Quality Evaluation Framework for Consumer-Generated Domestic Food Preparation Data.
Deliverable No.:	D 6.3
Authors:	Naomi Klepacz, Marcus Maringer, Susanne Ekman, Anne Norman, Muriel Verain, Anouk Geelen, Monique Raats
Reviewer:	Karin Zimmermann – Project Coordinator Pieter van't Veer – Scientific Coordinator Bent Egberg Mikkelsen – RICHFIELDS Phase 2 Leader
Start date:	01.03.2016
Delivery date:	30.05.2018
Due date of deliverable:	15.07.2016
Dissemination level:	PU
Status:	Final

Change history:

Version	Notes	Date
001	Draft version	15.07.2016

Karin Zimmermann

Project Coordinator

Prof. dr. ir. 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.

SUMMARY

This deliverable formulates a set of quality criteria for the evaluation of this consumer-generated food preparation data in terms of its scientific relevance and technical and legal governance. These three area were selected as indicators of quality as they allow for the assessment of data in relation to key questions relating to domestic food preparation behaviour (i.e., What/Who/Why/How and Where). This is, in addition to assessing the legal limitations, organizational restrictions, confidentiality and privacy concerns related to collection, integration and dissemination of consumer-generated data and the technical protocols and standards for data access and data processing. Information about these topics is crucial for developing the blueprint of a data platform, such as RICHFIELDS, as well as for its data governance structure.

In addition to providing a framework for the evaluation of data quality, the result of this deliverable also provides structure and guidance for the data collection process of deliverable 6.1, which is an inventory of consumer-generated food preparation data tools. More specifically, this quality framework provides an operationalised definition for each quality criteria in the form of a set of relevant questions that should be answered for each tool included in the RICHFIELDS Inventory Management System (RIMS). RIMS is an online management system for the storage and assessment of tools that produce consumer-generated data on the purchase, preparation, consumption of food and/or beverages and their associated lifestyle data that could potentially be of use to social science researchers. RIMS comprises two component parts; [1] a typology of the tools stored within the inventory, and [2] a list of quality criteria against which each tool can be evaluated.

The typology is a scheduled framework categorizing the food preparation tools according to defined groupings. The current typology for food preparation is a four-level model. The first level is the overall domain - in this instance, domestic food preparation. The second level reflects the goal of underlying motivation of the behaviour captured by the tool. The third level reflects the specific behaviours captured by the tool and the final level is indicative of the recorded behaviour.

The identified quality criteria are based on aspects of health and lifestyle specific to food consumption. Preparation behaviours are in some respect quite distinct and different from food intake, as they frequently require a degree of pre-behaviour decision making such as looking up a recipe. In this regard the current quality criteria don't sufficiently capture 'intended' behaviours, only enacted behaviours. The next step for these criteria is to test them with the tools currently in RIMS. However, for these tools it will be challenging to validate them according to current criteria at the level required for the inventory presented in deliverable 6.1. As for many tools, it is not possible to respond to these the criteria, particularly with the feasibility parameters worked to in this exercise. That is to say, it is not possible to easily identify certain aspects of a tool's quality without either expert knowledge of the fields of ICT and Law, and without the downloading and the downloading and testing of a tool, the examination of a tool's data structure and/or the examination of a hosting data infrastructure. This is therefore a potentially time consuming and costly process to validate the quality of consumer-generated data produced via a tool.

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.

1. INTRODUCTION

1.1 BACKGROUND

In recent years there has been a move towards open data with governments, public authorities and other organizations around the world have launching initiatives to make data available for broader, public use (Heimstädt et al., 2014). Open data is defined as data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike (Open Knowledge Foundation, 2018). In light of digitalization, knowledge generation is changing all spheres of human activity including social interactions, business and health care resulting in what is often referred to as "Big Data". Such data has been characterized by the four dimensions of volume (the amount of data being generated), velocity (the speed at which data needs to be processed), variety (issues relating to data heterogeneity), and veracity (the degree of uncertainty in the content of user generated data) (Schroeck et al., 2012), see Table 1.

Table 1. *Characteristics of big data (Adapted from Schroeck et al., 2012).*

Volume	Variety	Velocity	Veracity
• refers the amount of data • characteristic most associated with big data, volume refers to the mass quantities of data that organizations are trying to harness to improve decision-making across the enterprise • what constitutes truly "high" volume varies by industry and geography	• refers to different types of data and data sources • is about managing the complexity of multiple data types, including structured, semi-structured and unstructured data • organizations need to integrate and analyse data from a complex array of both traditional and non-traditional information sources, from within and outside the enterprise • the explosion of sensors, smart devices and social collaboration technologies has resulted in data is being generated in countless forms, including: text, web data, tweets, sensor data, audio, video, click streams, log files and more	• refers to data in motion, i.e. the speed at which data is created, processed and analysed continues to accelerate • contributing to higher velocity is the real-time nature of data creation, as well as the need to incorporate streaming data into business processes and decision making impacts latency, i.e. the lag time between when data is created or captured, and when it is accessible • data is continually being generated at a pace that is impossible for traditional systems to capture, store and analyze	• refers to data uncertainty, i.e. the level of reliability associated with certain types of data. • striving for high data quality is an important big data requirement and challenge, but even the best data cleansing methods cannot remove the inherent unpredictability of some data • the need to acknowledge and plan for uncertainty is a dimension of big data that has been introduced as data users seek to better understand the uncertain world

The Horizon2020 project RICHFIELDS recognizes that the open data movement in research and innovative ways of data collection including user-generated (big) data provide

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

unprecedented possibilities to study diet, lifestyle, health and their determinants. Data can be collected by using new media, e.g. in the social space (the Web, GIS) and real-time (apps, wearables, GIS, sensors) at the individual and group (e.g. household) level. These data could provide valuable information on the association between determinants and dietary intake which is of high societal and scientific relevance. Particularly in the area of infectious disease monitoring, the use of user-generated data has been heralded as an opportunity to improve public health surveillance (Velasco et al., 2014). Health agencies have been reluctant to incorporate these data sources into their systems because many technical issues have not yet been addressed (Velasco et al., 2014). Considerations of data protection and privacy, such as legal and ethical implications related to using Internet and social media data are also needed (Velasco et al., 2014). In the area of diet and health, 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 to predict whether users will or will not meet their daily caloric goals.

Food choice operates at physical, biological, psychological, and sociocultural levels (Sobal, 1991), all which operate simultaneously and interact (Sobal et al., 2014). In scientific research, data is collected in controlled conditions to provide insights. The types and sequences of food-related behaviours are depicted in Figure 1 and include the acquisition, preparation, serving, eating, storage, giving away of and cleaning up of food.

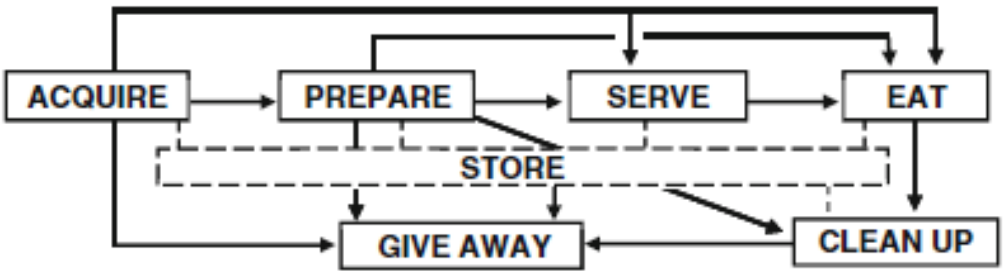

Figure 1. *Summary of types and sequences of food behaviours (from Sobal and Bisogni, 2009)*

Phase 1 of the RICHFIELDS project seeks to identify food-related data that is being actively or passively generated by consumers through the use of tools such as apps and sensors. An example of a food preparation related sensors are Smart Chopsticks from Baidu which measures PH levels, temperature, calories and freshness of cooking oil (Yie, 2014). “Outside the research environment”, people are generating data through everyday food- related activities. These might include banking transactions from which food-related purchase can be estimated, food related (e.g. recipes, restaurant reviews) search behaviour on the internet and the use of apps to record food intake or disclose food-related images or text. The large-scale generation of such data could have the potential to provide data for the purpose of

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

research which can provide insights regarding food choices and can relate to the purchase, preparation and consumption of food.

1.2 AIMS

The aim of this deliverable (D6.3) is to formulate a set of quality criteria for the evaluation of this consumer-generated data in terms of its scientific relevance and technical and legal governance. These three area were selected as indicators of quality as they allow for the assessment of data in relation to key questions relating to domestic food preparation behaviour (i.e., What/Who/Why/How and Where). This is, in addition to assessing the legal limitations, organizational restrictions, confidentiality and privacy concerns related to collection, integration and dissemination of consumer-generated data and the technical protocols and standards for data access and data processing. Information about these topics is crucial for developing the blueprint of a data platform, such as RICHFIELDS, as well as for its data governance structure.

In addition to providing a framework for the evaluation of data quality, the result of this deliverable provides structure and guidance for the data collection process of deliverable 6.1, which is an inventory of consumer-generated food preparation data tools (see section 1.3). More specifically, this quality framework will provide an operationalised definition for each quality criteria in the form of a set of relevant questions that should be answered for each tool included in the RICHFIELDS Inventory Management System (RIMS).

The aim of this deliverable is therefore not to create an exhaustive list of criteria for the validation of 'Big data' sets in terms of their potential use in social science research. Although, it is acknowledged that such validation is of course crucial for the use of 'Big Data' in social science research and warrants closer examination in relation to specific research questions.

Instead, the list of quality criteria as set out in this document aim to access whether the data collected through individual tools (i.e., smartphone apps, websites, and sensors) has the potential for use in social scientific research. That is to say, is the consumer-generated data 'fit for purpose'?

Furthermore, we believe that such a 'tool' or source level validation of quality should form an integral part of the Research Infrastructure and should therefore be considered and incorporated into the overall design of the RICHFIELDS data platform.

1.3 RICHFIELDS INVENTORY MANAGEMENT SYSTEM (RIMS)

The RICHFIELDS Inventory Management System (RIMS) was created in response to Task 6.1 which required the creation on an inventory of types of preparation and lifestyle data, and data collection methodologies. In brief, RIMS is an online management system for the storage and assessment of tools that produce consumer-generated data on the purchase, preparation, consumption of food and/or beverages and their associated lifestyle data that could potentially be of use to social science researchers. RIMS comprises two component

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

parts; [1] a typology of the tools stored within the inventory (see section 2), and [2] a list of quality criteria against which each tool can be evaluated (see section 3).

2. THE TYPOLOGY

2.1 OVERVIEW OF TYPOLOGY

The typology is a scheduled framework categorizing the food preparation tools according to defined groupings. The current typology for food preparation is a four-level model. The first level is the overall domain - in this instance, domestic food preparation. The second level reflects the goal of underlying motivation of the behaviour captured by the tool. The third level reflects the specific behaviours captured by the tool and the final level is indicative of the recorded behaviour. See Figure 2 for diagram of typology.

Level 1: What is the activity domain?	Domestic food preparation						
Level 2: What is the user aiming to do?	Planning & organisation (food skills)			Knowledge & 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 & 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

Figure 2. *Typology of Domestic Food Preparation. A four-level model: Level 1: overall domain - in this instance, domestic food preparation; Level 2: goal of underlying motivation of the behaviour captured by the tool; Level 3: specific behaviours captured by the tool; and Level 4: indicative of the recorded behaviour.*

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

2.2 DEFINITION OF DOMESTIC FOOD PREPARATION

Research by Stead and colleagues (2004) showed how 'cooking' encompasses a wide range of skills needed to feed people and includes specific elements of meal preparation (e.g. chopping, mixing, heating ingredients, understanding the language and terminology of recipes, following recipes, understanding measurements and cooking techniques) as well as knowledge of how to plan and budget for food and organise and plan meals that are acceptable to other household members. Food preparation is a multifaceted set of interconnecting behaviours that centre around the preparation of food either for one's own consumption, or the consumption of others such as in a domestic (e.g., family) and/or commercial (e.g., restaurant) setting. For the purposes of this project, food preparation will refer to domestic food preparation only as food service related data is digitalised via business processes and thus in this project conceptualized as food purchase behaviour. That is, food prepared for one's own consumption, or that of close others (e.g., family members), in the home or other non-commercial environment.

In this instance food preparation can be said to focus on two core skills sets; Food Skills (FS) and Cooking Skills (CS).

Food Skills: Food skills can be defined in terms of two behavioural components. That is, (a) Planning and Organisation, and (b.) Food Knowledge and Understanding. Both can be considered as necessary antecedents to preparation of foods (Fordyce-Voorham, 2009).

Planning and Organisation: These skills are reflective of the decision-making process involved in food preparation (McGowan et al., 2015). They may include behaviours such as

[i] *Documenting and recording food. F*or example, the making shopping lists, or recording expiration dates of food items.

[ii] *Meal/menu planning*. For example, 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).

[iii] *Recipe management*. For example, collecting and categorising recipes for future use.

Knowledge and Understanding: These skills reflect a persons need for information relevant to intended preparation behaviour or the reflection on a previously carried out behaviour (Stead et al., 2004), such skills include;

[i] *Sharing knowledge and experience*. For example, bookmarking or favouriting information within an app for the intention of future use, and/or the reading a writing of reviews and sharing of knowledge and experience via social media.

[ii] *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.

Cooking Skills: These can be defined as a set of mechanical and/or physical skills used in the preparation of foods, such as chopping, mixing and heating (Short, 2003). Cooking Skills also

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

encompass perceptual and conceptual skills, such as understanding how a food will react when heated (Short, 2003).

Meal Preparation and/or Cooking: The underlying goal or motivation of this behaviour is the actual preparation of foods for consumption. Behaviours relevant to meal preparation and/or cooking may include;

[i] *Using apps as cooking aids*. For example, the use of an egg timer, or digital measurement or conversion chart.

[ii] *Interacting with sensors.* For example, the use of the 'internet of things', intelligent kitchen equipment and appliances, or sensors in the home.

3. QUALITY CRITERIA

In order to create the quality framework a literature search was conducted. Private as well as public companies and institutions offer guidelines, services and infrastructures for reviewing, evaluating and certifying health applications, and the literature search was conducted on these existing quality frameworks. Quality criteria from that overview were selected based on the significance for the quality dimensions related to scientific relevance and legal and technical governance. In order to evaluate the relevance of our selection of quality criteria which reside outside of our own field of expertise (legal and technical governance), we contacted experts in the relevant fields of Law[1] and ICT[2] (one distinguished expert for each field of expertise). Based on the experts' opinions the selection of quality criteria was adjusted. The quality criteria used were further developed during the inventory, i.e. adding variables/inputs to the criteria as result of increased knowledge about different tool types and what data they potentially generate. As existing quality frameworks are rather general in nature with respect to scientific relevance and do not focus on specific scientific fields such as those relevant to RICHFIELDS. It was thus necessary for the assessment of quality within RICHFIELDS to create a unique set of criteria.

3.1 SELECTING QUALITY CRITERIA

The quality criteria currently in RIMS were identified and selected based on the consultation process outlined below. This three-stage process involved the identification of existing quality frameworks, the consultation of experts and an evaluation of feasibility.

3.1.1 EXISTING QUALITY FRAMEWORKS

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;

[1] Prof Indira Carr, Faculty of Arts and Social Sciences, University of Surrey, Guilford, Surrey, United Kingdom.

[2] Dr Barbara Korousic Seljak, Computer Systems Department, Jožef Stefan Institute, Ljubljana, Slovenia.

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

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. However, existing qualify frameworks are rather general in nature with respect to scientific relevance and do not focus on specific scientific fields, such as those relevant to RICHFIELDS. Thus, it is necessary for the assessment of data quality within RICHFIELDS to create a unique set of quality criteria.

3.1.2 EXPERT OPINIONS

In order to establish appropriate legal and technical quality criteria for inclusion onto the list, consortium members representing the fields of ICT and Law were consulted. These experts provided feedback as to the quality criteria required both for the final RICHFIELDS Research Infrastructure, and also identified the quality needs of Phase 2 consortium members.

3.1.3 FEASIBILITY

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 quality 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. An overview of the quality criteria used to characterise the tools is provided in Table 2.

Table 2. *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

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

Full details in Table 3	Full details in Tables 4 and 5	Full details in Table 6	Full details in Table 7

3.2 QUALITY CRITERIA: DESCRIPTIVE

In order to effectively evaluate the quality of a tool, it is essential that certain key characteristics be identified, described and recorded within RIMS. These descriptive characteristics focus on the identification of the source of the information and classification of each tool according to the typology set out in section 2. A full list of the Descriptive Quality Criteria and their descriptions can be found in Table 3.

It was considered an important indicator of quality that each tool be 'traceable', both in terms of the search strategy used to locate the tool and in terms of where additional information can be found about the tool (e.g., a website), or – in the case of apps – downloaded. Descriptions of apps and services were obtained from publicly available information made available by app vendors, i.e. without the need to install and use the app, e.g. the technical details, app descriptions and screenshots provided in the respective app stores (iTunes and Google Play Store) and, where available, feature and service descriptions, documentation, and frequently asked questions on associated homepages. Quality criteria requiring the installation and usage of the apps were not included, e.g., criteria related to the functionality of the tools or the resulting user experience, such as feasibility, intuitiveness, learnability, efficiency, engagement, etc.

Therefore, in addition to basic information, such as the name of the tool as it appears in ITunes, for each tool the 'search type' used to find the tool needs to be recorded, in additional to information about the 'search engine', 'reference tool' and 'search term'. Descriptions of these criteria and variables, currently identified within RIMS, can been seen in the top row of Table 3. Further identifying factors about the tool are also collected. This is to assist in maintaining reliability, as it enables users to ensure they are using the same tool obtained via the same source. These additional identifying factors include; a copy of the 'tool logo', the 'tool description' as written by the tool developer, 'languages' supported by the tool as well as 'supported platforms'. In addition, information on how to access the tool is also recorded, including where the tool can be accessed or downloaded, the website and homepage for the tool and the name of the company, or app developer, who currently owns the tool. Descriptions of these characteristics can be found in the bottom row of Table 3.

Information was also collected on the tool type. That is, whether the tool is an app, website or software. In the case of apps obtainable via the ITunes sore, further information concerning the characteristics of these apps can be obtained directly via the apps's unique ITunes ID number. Thus, for each app – where possible – the ITunes ID number was recorded.

An important indicator for the quality of a tool is whether it meets the overarching aim of the RICHFIELDS project. That is to say, the tool must collect consumer-generated data on, in this instance, domestic food preparation. Thus the tool – and the data it collects - must be classifiable according to the typology set out in Section 2. The descriptive quality criteria therefore list criteria for 'data types,' that is whether or not the app collects consumer-generated 'food preparation data'. Further descriptive quality criteria allow for the

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

identification of the consumers' 'goals', according to the typology (i.e., 'category preparation'), and also their 'behaviour', according to the typology (i.e., 'subcategory preparation').

Table 3. *Descriptive quality criteria collected for each tool.*

Criteria (sub-criteria)	Descriptive Question	Criteria Description	Variables
Search Type	How did you find the tool?	To identify the type of search strategy that was used to identify the tool.	Search Engine Reference
Search Engine	What search engine has been used to find the tool?	Variable contingent upon the select of the variable 'search engine'. Name of the specific search engine that has been used to identify the tool.	Appcrawl Vinoza Google Play Google Search ITunes PubMed Fnd.io
Search Term	What is the search you used to find the tool?	Variable contingent upon the selection of the variable 'search engine'. The search term or string of terms used to identify the tool.	Text Entry
Tool Reference	Where did you find the tool	Variable contingent upon the selection of the variable 'reference'. The scientific reference used to identify the tool (e.g., journal article).	Text entry
Tool Type	What category does the tool belong to?	To identify the type of tool and the category to which it belongs.	App Website Software
Query ITunes Store	If selected, after you save the item application information will be collected from ITunes search API and automatically inserted.	The ITunes ID number for the tool.	Text entry
Data Types	What type of data does the tool collect?	To identify the type of consumer-generated data that the tool collects.	Lifestyle Purchase Preparation Consumption
Category lifestyle	What lifestyle category does the tool belong to?	To identify the category relating to the type of consumer-generated lifestyle data that the tool collects.	Mind Health Environment Activity

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

Criteria (sub-criteria)	Descriptive Question	Criteria Description	Variables
Subcategory purchase	What lifestyle subcategory does the tool belong to?	To identify the subcategory relating to the type of consumer-generated lifestyle data that the tool collects.	Stress level (health) Social environment (Environment) Physical environment (Environment) Body composition (Health) Blood Sugar (Health) BMI (Health) Mood (Mind) Thoughts (Mind) Heart Rate (Health) Weight (Health) Blood Pressure (Health) Sleep (Activity) Exercise (Activity) Steps (Activity)
Category purchase	What purchase category does the tool belong to?	To identify the category relating to the type of consumer-generated purchase data that the tool collects.	Rating/Review Banking Eating Out Retail/groceries Vouchers/Coupons Shopping List Search Engine
Subcategory purchase	What purchase subcategory does the tool belong to?	To identify the subcategory relating to the type of consumer-generated purchase data that the tool collects.	Survey(Rating/Reviews) Comments (Rating/Review) Scores (Rating/Reviews) Budgeting (Banking) Transaction (Banking) Credit Card (Banking) Order/Take-away (Eating out) Foodbags/Meals (Retail/Groceries) Online shopping (Retail/Groceries) Specific (Discount/Coupons) General (Discount/Coupons) Pre-generated (shopping list) Self-generated (shopping List) Product info (Search Engine) Locate/Book restaurant (Search Engine) Price Comparison (Search Engine) Locating Store (Search Engine)

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

Criteria (sub-criteria)	Descriptive Question	Criteria Description	Variables
Category preparation	What preparation category does the tool belong to?	To identify the category relating to the type of consumer-generated preparation data that the tool collects.	Meal preparation / cooking Knowledge and understanding Planning and organisation
Subcategory preparation	What preparation subcategory does the tool belong to?	To identify the subcategory relating to the type of consumer-generated preparation data that the tool collects.	Interacting with sensors (meal prep/cooking) Using apps (meal prep/cooking) Searching for info (Knowledge and understanding) Recipe management (Planning and Organisation) Meal Planning (Planning and Organisation) Documenting/Recording (Planning and Organisation)
Category consumption	What consumption category does the tool belong to?	To identify the category relating to the type of consumer-generated consumption data that the tool collects.	Productivity Medical Support Behaviour Change
Subcategory consumption	What consumption subcategory does the tool belong to?	To identify the subcategory relating to the type of consumer-generated consumption data that the tool collects.	Intake recording (Productivity) Memory/sharing (Productivity) Alcohol/Coffee (Behaviour change) Special diet (Behaviour change) Food intolerance (clinical diagnosis) Healthy diet (behaviour change) Hydration (behaviour change) Diabetes (clinical diagnosis) Weight management (behaviour change)
Tool Logo	Add a URL to a logo image	The Uniform Resource Locator (URL) web 'address' for the logo image associated with the tool	Text entry
Tool description	Tool description	The description of the tool provided by the app developer and visible to the app user in ITunes.	Text entry
Languages	Which languages are supported?	A list of languages supported by the tool.	Text entry
Download / Access URL	Where can the tool be downloaded or accessed?	The URL web address from which the tool can be accessed or downloaded.	Text entry

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

Criteria (sub-criteria)	Descriptive Question	Criteria Description	Variables
Company Name	What is the name of the company who owns the tool?	The name of the company who owns the tool.	Text entry
Website URL	What is the web address of the tool?	The URL web address for the tool. This website may be a supporting website, or a website for the company or web developer who owns the tool.	Text entry
Has home page	Does the tool provide a link to a working home page?	To identify whether or not the tool has a working home page	Yes No
Supported platforms	What are the supported platforms?	To identify which platforms are supported by the tool.	Kindle Watch OS Android wear Blackberry HTML5 Windows Phone OSX Windows Android IOS

3.3 QUALITY CRITERIA: SCIENTIFIC – GENERAL

3.3.1 OVERVIEW OF SCIENTIFIC CRITERIA

The aim of the scientific criteria is to establish whether the consumer-generated data collected via a tool can be used to answer established research questions regarding food intake and food choice. In addition, the overarching aim of RICHFIELDS is to identify determining factors in order to answer wider research questions about what we choose to eat, and how and why we make these choices. Such factors may include, biological determinants (e.g., physical ability to prepare a food), economic determinants (e.g., cost and income), physical determinants (e.g., availability of food, access to services and information), social determinants (e.g., class, culture, social and situational contexts), psychological determinants (e.g., mood, goals, attitudes, beliefs and knowledge) and medical determinants (e.g., health status).

Food choice behaviour in general is a seemingly simple, but in fact very complicated behaviour that is influenced by many interacting factors. Moreover, these factors 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). The quantity eaten is of course also of particular interest. The complexity also goes for food purchase behaviour.

To this end, quality criteria were established around a traditional research question framework of ‘what, who, why, how and where’. Both general scientific quality criteria and

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

domain specific (purchase, preparation, consumption) quality criteria were created. The following sections will outline the general scientific quality criteria (Table 4) and those criteria specific to domestic food and/or beverage preparation (Table 6).

3.3.2 GENERAL SCIENTIFIC QUALITY CRITERIA

General scientific quality criteria were established that cross all three domains (purchase, preparation, consumption). These criteria include questions of associated lifestyle data. That is, additional consumer-generated lifestyle data collected through associated purchase, preparation and consumption tools. There has also been a growing trend towards the self-monitoring of lifestyle activities and patterns, such as monitoring steps taken, through the use of wearable devices and wireless sensors (e.g., Evenson et al., 2015). To this end, quality criteria regarding the use of external devices were included in the list of criteria. There is also a similar consumer use of social media and a desire to link, evaluate and/or share various aspects of food purchase, preparation and consumption activities. Information about these associated lifestyle activities adds important context to a data set. So much so, that the additional criteria of situation characteristics was added. This criteria relates to the characteristics surrounding both the situation under which the data was generated, and the characteristics of the product that was purchased, prepared and/or consumed were deemed as important indicators of a tool's quality.

Table 4. *General scientific quality criteria collected for each tool.*

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
Lifestyle Data	What type of lifestyle data does the tool collect?	To identify the type of consumer-generated lifestyle data collected by an app relating to purchase, preparation and/or consumption behaviour.	Text Entry
.Situational Characteristics	Does the tool collect information about the situation of a consumer?	To identify whether or not the app collects data regarding the context and/or situation in which the consumer-generated data collected by the app was generated.	Yes No
Type of situational characteristics	*What type of situational characteristics does the tool collect?*	*The type of context and/or situational data collected by the app.*	*Text Entry*
Product Characteristics	Does the tool collect information about the characteristics of the product what has been consumed, prepared or purchased?	To identify whether or not the app collects consumer-generated data regarding the type of product purchase, prepared or consumed by the user.	Yes No
Type of product characteristics	*What type of product characteristics does the tool collect?*	*The type of product characteristic(s) collected by the apps.*	*Text Entry*

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
External devices of same vendor	Does the tool support external devices owned by the vendor of the tool?	To identify whether or not the app supports external devices manufactured by the same company as the current tool with the express intent of being used in partnership.	Yes No
Device Type	What type of external devices does the tool support?	The type of external device(s) supported by the app.	Text Entry
Data integration with partner tools	Does the tool integrate data from other tools?	To identify whether or not the app integrates consumer or non-consumer-generated data (such as demographic data) from other tools included in RIMS.	Yes No
Partner tools	Of which other tools does the tool integrate data?	The name of the tool included in RIMS from which the current tool takes its information.	Text Entry

3.3.3 SCIENTIFIC QUALITY CRITERIA FOR PREPARATION

Possible criteria for assessing the scientific quality of tools collecting consumer-generated food preparation data were discussed amongst Phase 1 consortium members. The decision was taken to follow a traditional research question framework which would explore questions surrounding the 'what, who, why, how and where' of food preparation. This decision was also in line with the overarching aim of RICHFIELDS which is to use consumer-generated data to identify determinates of food intake and food choice.

The aim with the 'What' quality criteria is to identify specific aspects of food preparation activity and/or behaviour. The quality criteria 'What was prepared?' aims to capture these specifics of food preparation. Additional criteria also aim to enhance this measure of quality by capturing information regarding the method/technique by which the tool captures data relating to food preparation (e.g., photograph, written review, menu plan, shopping list), the unit form of this data (e.g., product units) and the population level to which this data refers (e.g., Individual, household, group).

The 'How much' quality criteria aim to identify whether the tool captures information about the quantity of food being prepared. Thus the quality criteria 'How much was prepared' was created, along with the accompanying criteria of 'method', which aims to identify whether or not the tool captures data on the method/technique by which the tool captures this food preparation related data (e.g., sensor, consumer self-report).

The quality criteria 'when it was prepared?' aims to identify the time at which the preparation behaviour took place, and the accompanying question 'time unit' seeks to identify the time unit that was recorded by the tool (e.g., hours, minutes). A similar criteria is 'where was it prepared?' which aims to established whether the tool collects information about the location in which the behaviour took place, and the unit of measurement that the tool captures. 'How was it prepared?' aims to identify how the food was prepared (e.g., cooked,

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

frozen) and the method used by the tool to capture this information. The final quality question seeks to establish where the preparation behaviour captured by the tool relates to a specific occasion, such as Christmas.

In addition, criteria are available that aim to identify whether the behaviour captured through the tool was an 'actual' behaviour carried out by the consumer, or just a behavioural intention. The theory of planned behaviour (TPB) maintains that behaviour is directly influenced by one's decision to act (i.e., intention) and the control one perceives one's self to have over the behaviour. Thus, intention to act, in turn, is dependent on attitudes toward the act, subjective normative pressure to act, and perceived behavioural control (Bagozzi and Dholakia, 2006). Many of the food preparation behaviours captured by tools reflect behavioural intentions, such a planning and meal or searching for knowledge by looking at a recipe, and thus are challenging to classify according to current scientific criteria.

Table 5. *Scientific quality criteria relating to food preparation collected for each tool.*

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
What was prepared?	Does the tool collect data on what was prepared?	To identify whether or not the app collects data on 'what' domestic food(s) and/or beverage(s) have been prepared. That is, the app collects observable data relating to the nature and characteristics of prepared domestic food(s) and/or beverages.	Yes No No information
Methods: What was prepared?	Which method(s) have been used to collect data on what was prepared?	To identify the method(s) of data collection used by the app to capture data relating to nature and characteristics of food(s) and/or beverage(s) prepared.	Text entry
What was prepared: act or Intention?	Does the data about what was prepared refer to intentions to prepare or actual acts of preparation?	To identify whether the data on what food was prepared captured by the app relates to an 'actual' act(s) that have taken place – or, whether the data relates to an 'intended' act(s). That is, data captured relates to an act(s) that have yet to take place.	Act Intention Both
What unit: preparations	What is the unit of measurement?	The unit in which the nature and characteristics of the food(s) and/or beverage(s) that have been prepared have been measured.	Text entry
What is the consumer unit: preparations?	What is the consumer unit for which preparations have been measured	The population for which data the nature and characteristics of the food(s) and/or beverage(s) that have been prepared have been captured.	No Information Group Household Individual

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
How much was prepared?	Does the tool collect data on how much was prepared?	To identify whether or not an app collects observable data relating to the quantity of domestic food(s) and/or beverage(s) that have been prepared.	Yes No No Information
Methods: How much was prepared?	Which method(s) have been used to collect data on how much was prepared?	To identify the method(s) of data collection used by the app to capture data relating to the quantity of food(s) and/or beverage(s) prepared.	Text entry
How much was prepared: act or intention?	Does the data about how much was prepared refer to intentions to prepare or actual acts of preparations?	To identify whether the data on the quantity of food prepared captured by the app relates to an 'actual' act(s) that have taken place – or, whether the data relates to an 'intended' act(s). That is, data is the captured relates to an act(s) that have yet to take place.	Act Intention Both
When was it prepared?	Does the tool collect data on when the preparation took place?	To identify whether or not an app collects observable data relating to the time at which food preparation took place.	Yes No No Information
Time unit: Preparations	In which unit(s) of time has preparation been measured?	The unit in which the time that preparation took place has been measured.	Date Weeks No Information Snack Meal Exact time Hours Day Periods Days Months Years
When was it prepared: Act or Intention?	Does the data about when food was prepared refer to intentions to prepare or actual acts of preparations?	To identify whether the data on when food preparation took place captured by the app relates to an 'actual' act(s) that have taken place – or, whether the data relates to an 'intended' act(s). That is, data captured relates to an act(s) that have yet to take place.	Act Intention Both
Where was it prepared?	Does the tool collect data on where the preparations took place?	To identify whether or not an app collects observable data relating the physical location in which the preparation food(s) and/or beverage(s) took place.	Yes No No Information

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
Location unit: preparations	In what unit has the location of preparations been measured?	The unit in which the physical location in which food preparation took place was measured/recorded.	No Information Venue Name Physical Environment Social Environment Geo - Coordinates In or out of the home
Where was it prepared: Act or intention?	Does the data about where was prepared refer to intentions to prepare or actual acts of preparations?	To identify whether the data regarding the physical location in which food preparation took place relates to an 'actual' act(s) that have taken place – or, whether the data relates to an 'intended' act(s). That is, data captured relates to an act(s) that have yet to take place.	Act Intention Both
How was it prepared?	Does the tool collect information about how the preparations took place?	To identify whether or not an app collects observable data relating to how food(s) and/or beverage(s) were prepared. That is, the apps captures data relating to the mechanism of food preparation.	Yes No Both
Method: How was it prepared?	What data does the tool collect about how the food was prepared?	To identify the method(s) of data collection used by the app to capture data relating to the mechanism of food(s) and/or beverage(s) preparation.	Free Text
How was it prepared: Act or intention?	Does the data about how was prepared refer to intentions to prepare or actual acts of preparations?	To identify whether the data regarding the mechanism of food preparation relates to an 'actual' act(s) that have taken place – or, whether the data relates to an 'intended' act(s). That is, data captured relates to an act(s) that have yet to take place.	Act Intention Both
Occasion	Does the tool collect information about the occasion of the preparations?	To identify whether or not an app collects observable data relating to the occasion on which food was prepared. Occasion can be operationalised as to the circumstances under which food preparation took place. This may include, meals times (i.e., breakfast) or a celebratory occasion (i.e., birthday).	Yes No No Information

3.4 QUALITY CRITERIA: TECHNICAL

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

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

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 Quality 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., API) for data access (e.g., Dedeke, 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).

The Technical Quality Criteria in RIMS operationalises 'data accessibility' can be seen in Table 6. Firstly, it is important to ascertain whether the data is accessible. This criteria 'is data accessible' seeks to answer the question 'is the data collected by the tool accessible directly via the tools infrastructure (not via integrated aggregators)? 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). Also, whether this data can be accessed using a commonly used access protocol.

Table 6. *Technical quality criteria collected for each tool.*

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
Is data accessible	Is the data collected by the tool accessible directly via the tools infrastructure (not via integrated aggregators)?	To identify whether or not the consumer-generated data collected by the tool is accessible either directly via the tool itself, or via its associated infrastructure (e.g., an API).	Yes No No Information
Access documentation	*Does the tool provide access documentation?*	*To identify whether or not the owner of the data has provided written documentation instructing users on how to access the consumer-generated data associated with the tool.*	*Yes* *No* *No Information*
Terms of access	*Does the tool provide a term of access document?*	*To identify whether or not the owner of the data has provided written terms by which a user may or may not be permitted to access the data.*	*Yes* *No* *No Information*
URL Terms of access	*Add URL to terms of use of the data access*	*Provide the URL, if available, to direct users to the terms of access associated with the tool.*	*Text Entry*
Implements access protocols?	Can the data be accessed using a commonly used access protocol?		Yes No No Information

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
Types of access protocols	*What commonly used protocol must be implemented to access the data?*		*Text Entry*
Data formats	*In what format is the data accessible?*	*To identify the format in which the user generated consumer data is accessible to the user (e.g., Excel, PDF, CSV).*	*Text Entry*
Authentication	Does access require authentication?	To identify whether or not access to the data set require the user to be authenticated. That is, some form of validation process is required to authenticate the identification of the user. This may be in the form of a user account held with the data owner.	Yes No No Information
Price	Does data access require payment?	To identify, whether or not the owners of the data require a fee/subscription or some other form of payment to access the data.	Yes No No Information

3.5 QUALITY CRITERIA: LEGAL

Following discussion with consortium members, issues surrounding data privacy, consumer consent, data ownership and data security were highlighted as important indicators of data quality.

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. These criteria can be seen in full in Table 7.

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 in-app advertiser – 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

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

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 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.

Table 7. *Legal quality criteria collected for each tool.*

Criteria (sub-criteria)	**Scientific Question**	**Criteria Description**	**Variables**
Terms of use	Does the tool provide a terms of use document?	To establish whether or not the tool provides a statement or legal documentation that discloses the terms under which the consumer may use the tool.	Yes No
URL Terms of use	Add URL to the terms of use document	To provide a URL to a website (or similar) on which the documentation relating to the terms under which the consumer may use the tool are found.	Text Entry
Privacy Policy	Does the tool have a privacy policy document?	To establish whether or not the tool provides a statement or legal document that discloses some or all of the ways in which the company responsible for the tool gathers, uses, discloses and/or manages the tool users data.	Yes No

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
URL Privacy Policy	Add URL to the privacy statements document	To provide a URL to website (or similar) on which a statement or legal documentation that discloses some or all of the ways in which the company responsible for the tool gather, uses, discloses and/or manages the tool users data can be found.	Text Entry
Data ownership	Who holds the ownership of the user generated data (user content)?	To identify the individual and/or company that has legal rights and control over a single piece or set of data generated through consumer use of the tool.	Text Entry
Data usage vendor	Does the tool vendor retain the right to access and exploit the user generated data (publish, distribute, publically display)?	To establish whether or not the vendor of the tool (e.g., app store), retains the right to access and exploit the consumer-generated data produced by the tool.	Yes No
Personal information	Does the tool collect personal identifiable information (e.g., during registration)?	To identify what, if any, personal identifiable information about the user of the tool is collected during the registration process.	Text Entry
Informed consent	Is the user asked permission about collecting personal identifiable information?	To establish whether or not the tool service user has granted permission for the data they generate to be used by another party in the full knowledge of the possible consequences.	Yes No No information
Types of Personal information	What types of personal identifiable information does the tool collect?	To identify the types of personal identifiable information collected by the tool about the tool user.	Text Entry
Public profile	Does the tool create a public profile of the users' personal data?	To establish whether or not the tool creates a public profile of either the consumers' personal identifiable information, or consumer-generated data collected via the tool.	Yes No No Information
Device Data	Does the app collect device data after installation/visit?	To establish whether or not the tool is collecting data as to the device via which the consumer uses the tool.	Yes No No Information
Type of device data	What type of device data does the tool collect?	To identify the type of information that is being collected about the device via which the consumer uses the tool (e.g., IP address).	Text Entry
Cookies	Does the homepage/website of the tool store cookies on a user's computer?	To establish whether or not the homepage/website of the tool stores cookies (data sent from the website to the user's browser).	Yes No No Information
Web Beacons	Does the homepage/website of the tool store web beacons to track the online moments of users?	To establish whether or not the homepage/website of the tool uses web beacons (embedded objects that invisibly check whether the user has accessed content).	Yes No No Information

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

Criteria (sub-criteria)	Scientific Question	Criteria Description	Variables
Identifiable data sharing (Affiliated parties)	Will collected personal identifiable data be shared with affiliated third parties (with confidentiality agreements)?	To establish whether or not personal identifiable information collected about the consumer with be shared with an affiliated third party, either with or without the consent of the consumer.	Yes With consent No No Information
Identifiable data sharing (Unaffiliated parties)	Will collected personal identifiable data be shared with unaffiliated third parties (without confidentiality agreements)?	To establish whether or not the personal identifiable information collected about the consumer will be shared with unaffiliated third parties, either with or without the consent of the consumer.	Yes With consent No No Information
Data storage	Where does the system store the data it generates?	To identify the storage system on which the owner of the data stores the consumer-generated data collected by the tool.	Web server storage Device storage No information
Data deletion	Is the user able to delete or ask for deletion of his or her personal identifiable information (e.g., after account termination)?	To establish whether or not the consumer is able to delete, or request deletion, of his or her personal identifiable information collected via the tool.	Yes No No Information
Usage Analytics	Does the homepage/website of the tool use third-parties for advertising and usage analytics?	To establish whether or not the homepage/website of the tool uses third-party advertising on their homepage/website, and/or a third-party for the analysis of homepage/website usage.	Yes No No Information
Third party services	Does the tool provide any third party services?	To establish whether or not the tool uses any third-party services. That is, are any services provided by the tool outsources to a company, or individual, other than the tool owner.	Yes No No Information
Social Network Sharing	Can the collected data be shared with social networks?	To establish whether or not the tool has the facility for the consumer to share their data collected within the tool with a social network (either their own social network, or that of another).	Yes No No Information
Data Encryption	Does the tool encrypt the collected data?	To establish whether or not the tool encrypts the consumer-generated data. That is, are the data converted into another form which cannot easily be understood by anyone other than an authorised party?	Yes No No Information

4. DISCUSSION

The work underlying this deliverable aimed not only to produce a list of quality criteria, but also to visualise the potential opportunities available to researchers for consumer-generated food data, and in particular domestic food preparation data. Our work aimed to identify

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

factors influencing the quality of this data and also to highlight potential gaps and needs in the collection, integration and dissemination of such data. The quality criteria set out in this deliverable aim to assess the potential for this consumer-generated data - collected through individual tools, such as smartphone apps, websites and sensors - have for use in social scientific research.

The identified quality criteria are based on aspects of health and lifestyle specific to food consumption. Preparation behaviours are in some respect quite distinct and different from food intake, as they frequently require a degree of pre-behaviour decision making such as looking up a recipe. In this regard the current quality criteria don't sufficiently capture 'intended' behaviours, only enacted behaviours. Furthermore, technology is constantly changing, as is legislation. It is therefore difficult to pin down specific quality criteria in such a way as to future proof RICHFIELDS. Thus the design and implementation of relevant quality criteria necessitate expert and currently knowledge in these fields.

The next step for these criteria is to test them with the tools currently in RIMS. However, for these tools it will be challenging to validate them according to current criteria at the level required for the inventory presented in deliverable 6.1. As for many tools, it is not possible to respond to these the criteria, particularly with the feasibility parameters worked to in this exercise. That is to say, it is not possible to easily identify certain aspects of a tool's quality without either expert knowledge of the fields of ICT and Law, and without the downloading and the downloading and testing of a tool, the examination of a tool's data structure and/or the examination of a hosting data infrastructure. This is therefore a potentially time consuming and costly process to validate the quality of consumer-generated data produced via a tool.

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

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.

Bagozzi, R.P., Dholakia, U. M. (2006) Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of Research in Marketing* 23 , 45–61.

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*. Retrieved from http://www1.unece.org/stat/platform/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf?version=1&modificationDate=1420725063663&api=v2

Evenson, K. R., Goto, M. M., Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 159.

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.

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.

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

https://arizona.pure.elsevier.com/en/publications/analyzing-the-language-of-food-on-social-media

Heimstädt, M., Saunderson, F., Heath, T. (2104) F rom toddler to teen: growth of an open data ecosystem. *eJournal of eDemocracy & Open Government*. 6, 123-135.

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 (2009) 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.

Loshin, D. (2002). Knowledge integrity: data ownership. *Data Warehouse magazine (Online) June 8, 2004*. Retrieved from www.datawarehouse.com/article/?articleid=3052

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., Caraher, M., Raats, M., Lavelle, F., Hollywood, L., McDowell, D., Spence, M., McCloat, A., Mooney, E. & Dean, M. (2015). Domestic Cooking and Food Skills: A Review. *Critical Reviews in Food Science and Nutrition,* 57:11, 2412-2431

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 Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1004&context=ecis2014

Njie, C. M. L. (2013). *Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications*. Retrieved from http://www.privacyrights.org/mobile-medical-apps-privacy-technologist-research-report.pdf

Open Knowledge Foundation. Open data handbook. Available from: http://opendatahandbook.org [Accessed 28th May 2018].

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

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P. (2012*). Analytics: The Real-World Use of Big Data*. London: IBM Global Business Services Business Analytics and Optimisation in collaboration with Säid Business School, University of Oxford. Retrieved from http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-bigdata-at-work.html

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. Retrieved from http://measuringbehavior.org/files/ProceedingsPDF(website)/Schulze_FullPaper5.3.pdf

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

Short, F. (2003). Domestic cooking skills- what are they? *Journal of the Home Economics Institute of Australia*, 10(3) 13-22.

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.

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

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.

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.

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*. http://arxiv.org/abs/1510.04802

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.

Jie, Y. (2014). Is your food safe? New 'smart chopsticks' can tell in: China real time. Wall Street Journal. http://blogs.wsj.com/chinarealtime/2014/09/03/is-your-food-safe-baidus-new-smart-chopsticks-can-tell/

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