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

Horizon 2020
INFRADEV-1-2014 - Design studies

RICHFIELDS Working Package 4
Deliverable D4.2

Report on the Synthesis of the findings for WP5-7 Phase 1

Date delivered:
M35

Authors:
Charo Hodgkins, Lada Timotijevic,
Monique Raats, Anne Normann, Marcus Maringer,
Naomi Klepacz

Deliverable lead beneficiaries:
USURREY

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 the Synthesis of the findings for WP5-7
Deliverable No.:	D4.2
Authors:	Charo Hodgkins, Lada Timotijevic, Monique Raats, Anne Normann, Marcus Maringer, Naomi Klepacz
Reviewer:	Karin Zimmermann – Project Coordinator Pieter van 't Veer – Scientific Coordinator
Start date:	1.09.2017
Delivery date:	30.08.2018
Due date of deliverable:	31.05.2018
Dissemination level:	PU
Status:	Final

Change history:		
Version	Notes	Date
001	First draft submitted for review	30.11.2017
002	Amended draft submitted	29.05.2018

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.

Summary

In order to support the design of the Research Infrastructure (RI) Phase 1 was tasked with exploring the range of consumer-generated data currently collected, mainly via smart phone applications and tools (APPS) in terms of the type and quality of consumer-generated data collected and consumers' perspectives on willingness to share their data with researchers.

The purpose of this deliverable, is to provide input to the final design of the proposed RICHFIELDS RI by Phase 3 and insight for the development of the wider roadmap proposal for the FNH-RI by synthesising the findings across the three domains within Phase 1 (purchase, preparation and consumption); highlighting opportunities and potential limitations for the scientific use of consumer-generated data; identifying potential opportunities/issues that are relevant for the final design of the RICHFIELDS RI/data platform (Phase 3), but which may not be covered specifically in the Phase 1 deliverables.

The main recommendations arising from this synthesis report are summarised below:

Scientific usefulness of consumer generated data

- The scientific limitations for consumer-generated purchase and preparation data identified by Phase 1 are potentially possible to overcome **by linking to data from consumer-generated consumption APPS** allowing a more extensive mapping of food choice and eating behaviour from preparation through to consumption for an individual.
- Unstandardized or undocumented food intake assessment procedures, data exchange protocols and formats, terms of use and privacy regulations, limit possibilities to integrate, process and share user-documented food consumption data in a scientifically robust way. Therefore **best practice guidelines, quality standards and protocols are needed for the effective integration** of consumer-generated food purchase, preparation and particularly composition data in a scientifically meaningful way.
- A vital source for better understanding the possible drivers and barriers for people's food purchase, preparation and consumption behaviour is likely to come from **associations between these data and other relevant social, health and lifestyle data**. For example, to gain domestic food purchase, preparation and consumption data from dedicated APPS and link this with health and lifestyle APPS for an individual. This combined **data could be further enriched with demographic, situational and social context data** collected through APPS such as Facebook, Twitter and Instagram.

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

Technical Considerations

- In the first instance it is suggested that the RIMS database and the **typologies developed by Phase 1** for the data collection in the purchase an preparation domains (see D5.1 & D6.1) should inform the development of the RICHFIELDS ontology, linkage and harmonisation methods in WP12.
- It is key to connect to developers/owners of AGGREGATORS already in the marketplace for the further development of the RICHFIELDS technical data infrastructure and to facilitate access to a wider breadth of consumer data.
- As part of the RICHFIELDS design, consideration might also be given to the development of a **RICHFIELDS specific APP** that could not only act as an AGGREGATOR to link with other APPS used by an individual, but also as a means of collecting additional standardised data from a cohort of individuals that are of interest for research purposes.
- It is recommended that the RI/data platform develops a means of **capturing and describing new APPS and AGGREGATORS** that are being used by consumers in the food and wellbeing domain so that researchers can easily see what data is being created by consumers that might best inform future research designs.

Business model, governance and ethical considerations

- Providing **different levels of consent options to consumers** (potentially via the RICHFIELDS APP) would allow consumers to specify exactly those stakeholder categories they are willing for their data to be shared with and those that they are not. In this way direct consent could be obtained from the consumer for the use of their data either for research, policy development or commercial activities and that consent held as meta-data within the RICHFIELDS data platform which from a governance perspective is the most desirable scenario.
- In parallel, by **developing relationships with the AGGREGATORS** already in the marketplace the appropriate levels of overarching consent could be built into their systems and then the data shared between the AGGREGATOR and RICHFIELDS with the consumer's consent.
- The initial results from the user survey (UK only) indicate that due to the increased trust and reduced risk associated with universities when compared to government or commercial organisations the RICHFIELDS RI **Governance Model may benefit from clearly identifying a university/universities as the lead** rather than a more diverse model where all stakeholders have equal leadership.

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

The vision of the RICHFIELDS project is to design a **Research Infrastructure (RI)** to collect, align and share consumer, business and research data in order to provide the scientific research community access to innovative data sets and the ability to generate new knowledge and breakthroughs in the consumer food and health domain. This will enable policymakers and other stakeholders to develop, evaluate and implement effective food and health strategies at the level of both individuals and populations.

It is proposed that the RI will provide an unprecedented opportunity to address the determinants of consumer behaviour relevant to food and health across three distinct instances of behaviour: purchase, preparation and consumption. By building on determinants and intake ('DI' components) of the proposed DISH-RI (www.eurodish.eu), the design proposal arising from the RICHFIELDS project will be an important building block for subsequently constructing an ESFRI roadmap proposal for a pan- European Food Nutrition and Health Research Infrastructure (FNH-RI).

In order to support the design of the Research Infrastructure (RI) Phase 1 was tasked with exploring the range of consumer-generated data currently collected, mainly via **smart phone applications and tools (APPS)** in terms of

- the type and quality of consumer-generated data collected
- consumers' perspectives on willingness to share their data with researchers.

1.1 Aim

The purpose of this deliverable, being prepared as part of the WP4 workplan, is to provide input to the final design of the proposed RICHFIELDS RI by Phase 3 and insight for the development of the wider roadmap proposal for the FNH-RI by

1) synthesising the findings across the three domains within Phase 1 (purchase, preparation and consumption)
2) highlighting opportunities and potential limitations for the scientific use of consumer-generated data
3) identifying potential opportunities/issues that are relevant for the final design of the RICHFIELDS RI/data platform (Phase 3), but which may not be covered specifically in the Phase 1 deliverables.

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

1.2 Summary of Phase 1 activities

1.2.1 Type and quality of consumer-generated data

In terms of exploring the type and quality of consumer-generated data collected, Phase 1 developed a common tool for their data collection methods across the three domains of purchase (WP5), preparation (WP6) and consumption (WP7). However, the specific search criteria utilised to identify the APPS of interest, the labels for the data collected on them and the quality criteria utilised were tailored for each of three domains. These are described in full in the respective deliverables; purchase (D5.1, D5.3), preparation (D6.1, D6.3) and consumption (D7.1, D7.3).

These Phase 1 data collection activities have resulted in the creation of the **RICHFIELDS Inventory Management System (RIMS)**. This has provided a methodology and repository for the capture of descriptive data on health and lifestyle APPS used by and accessible to the general public, the methodologies they implement and the parameters/data types they collect and integrate. The aim of developing this inventory was to provide the basis for identification of the scientific, technical and legal/ethical opportunities/issues regarding the potential use and integration of consumer-generated food behaviour data within a Research Infrastructure/shared data platform such as is being proposed by RICHFIELDS.

1.2.2 Consumers' perspectives on willingness to share their data

In order to gain a deeper understanding of consumers' perspectives, Phase 1 have also developed a study to collect data via an online questionnaire in 8 EU countries (France, Germany, Italy, the Netherlands, Slovenia, Spain, Sweden, United Kingdom). The results from this study (n= 1000 per country) will enable RICHFIELDS to

1) understand the extent to which consumers are willing to share their food and health related data, across three different potential user groups/stakeholders; (1) publicly funded researchers, (2) governments and (3) industry.
2) describe differences in willingness to share by country, age, gender, education or socio-economic status.
3) establish the relevant predictors to willingness to share with a view to gaining a deeper understanding of how we might encourage future sharing of data by consumers with the proposed RICHFIELDS data platform/RI.

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

The survey contained the following items:

- The range of data types being shared across the three domains (purchase, preparation and consumption)
- Effect of data sharing context (publicly funded researchers, governments and industry)
- Predictors of willingness to share data
    - Trust and confidence in organization handling data
    - Privacy concerns
    - Reasons for sharing
    - Values (how they see the world)
        - Self-conservation (tradition, acceptance)
        - Self-transcendence (success, adventure)
- Attitudes
    - to science
    - to food and health
- Cooking/shopping practices
- Perceived health

Data has been collected and the analysis of the full dataset is ongoing. It will be reported to Phase 3 as soon as it is available to inform the final design of the governance structure for the proposed RICHFIELDS RI/data platform. Initial analysis of data from the UK sample is discussed in section 3.2.

In addition, the design of this Phase 1 study has provided input into the development of the wider user needs survey being implemented as part of the FHN-RI preliminary planning phase.

2 Method

To develop this deliverable, a desk-based review of the deliverables arising from the Phase 1 activities was performed by WP4. The documentation reviewed are detailed in Table 1. The results section of this deliverable focusses on synthesising the relevant Phase 1 outcomes in terms of the scientific usefulness of the consumer data across the three domains of purchase, preparation and consumption. The discussion section seeks to summarise the outcomes to inform the technical considerations (WP11), business model considerations (WP12) and finally governance/ethical considerations (WP13) in order to inform the final design of the RI/data platform

For specific information on the methodologies utilised by Phase 1 reference should be made back to the deliverables detailed in Table 1 although a short summary of the RIMS

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

database development is provided in section 2.1 below to facilitate understanding of this report.

Table 1 – Documentation reviewed for development of this synthesis report D 4.2*

Domain	Deliv. No.	Deliverable Title
WP5 - Purchase	D5.1	Report on inventory of types of purchase data and data collection methodologies
	D5.3	List of quality criteria
	D5.4	Paper on quality criteria and overview of criteria applied to available data/methods
	D5.5	Report on gaps and needs
WP6 - Preparation	D6.1	Report on inventory of types of preparation data and data collection methodologies
	D6.3	List of quality criteria
	D6.5	Report on gaps and needs
WP7 - Consumption	D7.1	Report on inventory of types of consumption data and data collection methodologies
	D7.3	List of quality criteria
	D7.4	Paper on quality criteria and overview of criteria applied to available data/methods
	D7.5	Report on gaps and needs

*Information provided by Phase 1 at the mid-term review both in the scientific reporting and in presentation form were also utilised to assist with the development of this synthesis report.

2.1 The RICHFIELDS Inventory Management System (RIMS)

The Richfields Inventory management System (RIMS) was designed to collect data on, and facilitate the evaluation of APPS that produce consumer-generated data. It details information about the APPS, including: metadata; data characteristics (regarding the theoretical meaning and value of the collected data); legal characteristics (regarding the permissions of data usage – terms and conditions) and technical characteristics (regarding the possibilities and protocols for data access and integration). RIMS is comprised of two main parts:

1) A set of selected APPS described according to a typology relevant for each domain (as defined in D5.3, 6.3 and 7.3).
2) The quality criteria for each of the APPS included in the database (as defined in D5.3, 6.3 and 7.3).

Utilising RIMS, Phase 1 were able to create an inventory and subsequently evaluate, a

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

range of APPS and report on these in terms of the types data collected, its scientific relevance and accessibility from both a technical and legal/ethical perspective. These analyses are reported in the specific deliverables for each domain; purchase (D5.4, D5.5), preparation (D6.4, D6.5) and consumption (D7.4, D7.5).

As a result of the purposeful sampling methodology applied by Phase 1, in its current form, the APPS captured in RIMS do not represent a definitive list of all available APPS but more a representation of the potential range of APPS that are being utilised by consumers (see Table 2). In addition, Phase 1 only included APPS available in the United Kingdom (UK) in their inventory since the legal and privacy documents for each APP needed to be interpretable by any researcher in the RICHFIELDS consortium.

Table 2 –Number of APPS by domain included in RIMS

Domain	No. of APPs catalogued
Purchase	62
Preparation	50
Consumption	257
Lifestyle	78
Aggregators (able to integrate with food consumption data)	12

2.2 *Quality criteria and assessment of APPS included in RIMS*

Once the APPS of interest had been identified for inclusion into RIMS, the next stage was to collect and classify the information available about each APP and identify the meta-data available in order to answer the following 4 key questions defined as being relevant for the development of the proposed RICHFIELDS RI/data platform:

1) ***What is it?*** – Descriptive criteria
2) ***Is it useful?*** – Scientific Criteria
3) ***Can we access it?*** – Technical criteria
4) ***Can we use it?*** – Legal/Ethical criteria

An overview/summary of the variables collected on each APP can be seen in Figure 1.

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

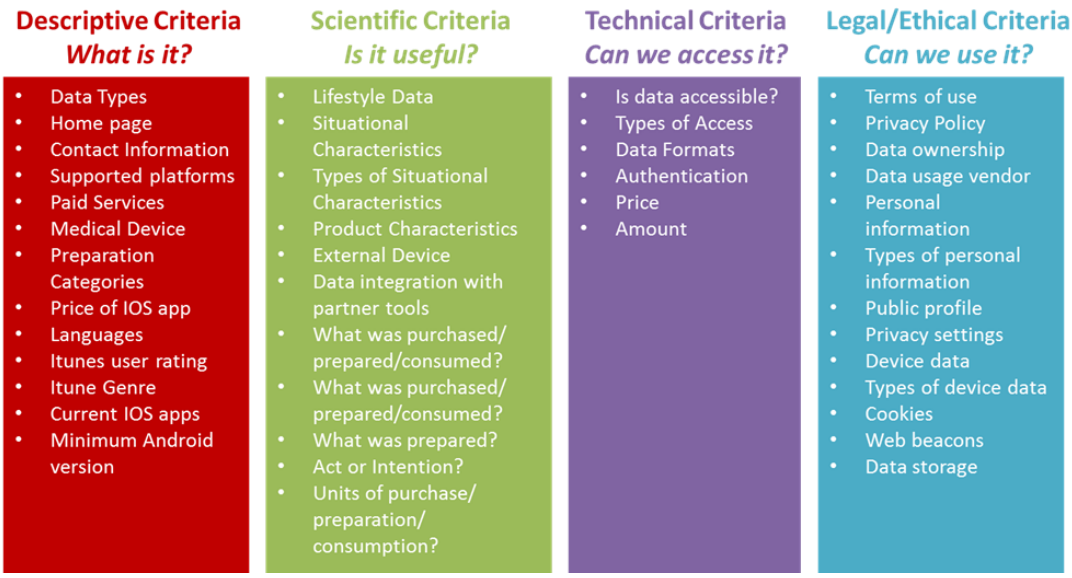

Fig 1. Quality criteria - Overview/summary of the variables collected for each APP in RIMS

3 Results – Synthesis of Phase 1 outcomes

3.1 Scientific usefulness of consumer-generated data

The activities performed in Phase 1 have identified both the possible limitations and potential opportunities with respect to consumer-generated data collected via the APPS evaluated:

3.1.1 Limitations for the scientific use of consumer-generated food purchase, preparation and consumption data in nutrition research

Food purchase data: The main limitation from a scientific perspective with respect to consumer-generated data in the purchasing domain is that it does not identify whether the purchased food is consumed or not, nor does it identify the individual that may actually consume the food. The food may well be consumed by someone other than the purchaser e.g. family or friends. In addition, this data does not typically differentiate between intention and actual purchase and as such is not really a proxy of consumption. As a result, strong connections to public health outcomes at an individual level are limited if this type of consumer-generated data is utilised in isolation of consumption data. Whilst food purchase data is able to provide some understanding of consumer preferences/habits i.e. the types of foods, food retailers and restaurants that may be “on a user’s mind” or that they utilise most frequently and can provide insight on food spend per week, month or year, its value is potentially limited: it cannot be used to track the behaviour associated with the purchase

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

(e.g. the extent to which the purchased food is consumed, shared with others or results as waste).

Food preparation data: Similarly to consumer-generated food purchase data, the degree to which consumer-generated food preparation data can act as a 'proxy' for intake is questionable. Whilst the data reflects consumers' motivation to gain knowledge and to develop skills in food preparation, 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 purchase or intake certain foods and/or meals. Nevertheless, if it is possible to link food preparation and food consumption data for a single individual through the data linkages, this may provide an invaluable insight into the complex relationship between intention and actual behaviour.

Food consumption data: In contrast to the consumer-generated food purchase and preparation data the majority of food consumption APPS do collect data at the individual level, on a daily basis, at a specific moment in time and over a period of time. Therefore, from a scientific perspective data collected by these APPS has the potential to provide insight into habitual food consumption behaviours and how these change over time at an individual level. However, the problem with many of these APPS is that similar to traditional food diaries, from a user perspective they rely on extensive commitment/high levels of individual motivation to maintain such a diary, good recall, time investment and a degree of expertise to identify and input appropriate food categories / products into the system. So, the issues of reliability, validity, perhaps social desirability (cheating) and drop out still exist.

From a scientific perspective, the unknown quality and validity of the food composition databases used to underpin these APPS and the non-standardised procedures for portion size estimation means that conclusions with respect to the relationship between food consumption and nutrition related diseases may be limited. Detailed research on the associations between specific nutrients and health outcomes may also be limited since majority of APPS in this domain focus only on energy and macronutrients.

Finally, and perhaps the most fundamental issue with consumer-generated food consumption data is that there is a particularly high prevalence of APPS with the aim of behavioural change. This intervention focus is likely to limit the ability to develop a true picture of people's habitual or typical food consumption behaviour because they have been primed towards a behaviour change goal that by definition, may change their habitual practices.

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

3.1.2 Opportunities for the scientific use of consumer-generated food purchase, preparation and consumption data in nutrition research

Whilst the above section highlights the limitations for the potential scientific use of consumer-generated data collected via APPS particularly for the three domains studied, there are still many opportunities for use of this data to help better understand food behaviour. In particular these include opportunities for RICHFIELDS to link with the existing AGGREGATORS established in the public domain (see technical considerations section 4.1) which are already linking consumer derived data across a range of different APPS into a personalised overview for a consumer.

Furthermore, the scientific limitations highlighted above for consumer-generated purchase and preparation data are potentially possible to overcome by linking to data from the consumption APPS identified in WP7 allowing a more extensive mapping of food choice and eating behaviour from preparation through to consumption for an individual. Although, it must be recognised that protocols for performing such linkages would need to be carefully developed. Unstandardized or undocumented food intake assessment procedures, data exchange protocols and formats, terms of use and privacy regulations, limit possibilities to integrate, process and share user-documented food consumption data in a scientifically robust way and therefore best practice guidelines, quality standards and protocols are needed for the effective integration of consumer-generated food purchase, preparation and particularly composition data in a scientifically meaningful way.

Consumer-generated food purchase, preparation and consumption data are not typically collected in isolation of other potentially relevant data. A vital source for better understanding the possible drivers and barriers for people's food purchase, preparation and consumption behaviour is likely to come from associations between these data and other relevant social, health and lifestyle data. This undoubtedly has the potential to give a more valid picture whereby different data sets corroborate each to her to create a fuller, more accurate picture overall and the interconnectedness of APPS/tools now presents new opportunities to further enrich the food-related data from external sources. For example, it may be useful to gain domestic food purchase, preparation and consumption data from dedicated APPS and link this with health and lifestyle APPS for an individual. This combined data could be further enriched with demographic, situational and social context data collected through APPS such as Facebook, Twitter and Instagram. However, it should be noted that the degree to which users would find this interlinkage acceptable and be willing to share this type of extensive data with the proposed RI will need to be carefully considered and governed (see section 4.3).

An overview of some of the potential opportunities and associated limitations for the scientific use of consumer generated data for RICHFIELDS are listed in Table 3. However, in the context of the future grand challenge of creating sustainable food systems these opportunities must be easily extendable to include the wider determinants of behaviour in

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

the food environment (i.e. the scope of the proposed FNH-RI).

3.2 Consumers' willingness to share data

Initial results from the analysis of the UK dataset are encouraging as they indicate that UK participants are on the whole willing to share their food related data with the three stakeholder groups; (1) universities or publicly funded research organisations, (2) governments and (2) commercial companies. However, the data demonstrates that they are statistically more willing to share with universities and this is due to a significantly higher level of trust associated with universities and lower perceived risk of sharing their data with this type of organisation when compared to the other two stakeholder groups. It should be noted however that this data was collected prior to the recent Cambridge Analytica and Facebook revelations, which are likely to have an impact on the stance of the public towards data sharing.

3.3 Legal and ethical considerations

The work carried out by Phase 1 highlights the many gaps with respect to the availability of publicly accessible data about the APPS/tools they evaluated. Due to the lack of available legal documents related to the terms and conditions and privacy statements, there is insufficient information available about the rules users must accept in order to use a service and the ways in which each APP gathers, uses, discloses, and manages their users' data. Hence, the legal limitations, organizational restrictions, confidentiality and privacy concerns related to collection, integration and dissemination of this consumer-generated data remain difficult to navigate other than on a specific case by case basis/detailed exploration with each individual APP of interest.

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

Table 3 – Potential opportunities and associated limitations for the scientific use of Purchase, Preparation and Consumption consumer-generated data

Domain	Potential opportunities	Limitations
Purchase	• Inferences about the trends at the population level linked to purchase intention/food spend etc • Trends linked to C2B interactions (which retailers/restaurants/outlets are most visited) • Trends in how preferences in different food groups /products are shifting i.e. atttitudinal changes re purchase intention	• Cannot directly link to an individual purchase • Cannot directly link to the individual's consumption • Cannot identify the unit of analysis (i.e. does the data refer to the individual or household?)
Preparation	• People's search behaviour online • Trends in recipe generation • Trends in social networking facilitated by food preparation knowledge/recipe sharing etc.	• Links to individual preparation behaviour • Cannot directly link to purchase or consumption at an individual level
Consumption	• People's individual food intake profiles • Understanding of habitual food consumption behaviours across groups of interest	• Quality/completeness of the underlying food composition databases questionable • Quality and completeness of the self-reports through diet intake/physical activity APPS • Level of detail of the estimated food composition values is low, with APPS typically focusing on energy and macronutrients. • Lack of information regarding the procedures for estimating portion sizes • High prevalence of behavioural change objective which might pose a barrier towards a better understanding of the real determinants of food consumption behaviours as well as the ability to provide an unbiased insight in peoples' habitual food consumption behaviours

4 Discussion

The following sections makes recommendations for the design of the proposed RICHFIELDS RI/data platform based on the synthesis of the Phase 1 outcomes. These are presented in terms of the three streams of work being undertaken in Phase 3 (ie WP11 technical, WP12 business models and WP13 Governance and ethical considerations.

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

4.1 Technical considerations (WP11)

In the first instance it is suggested that the RIMS database and the typologies developed by Phase 1 for the data collection in the purchase and preparation domains (see D5.1 & D6.1) should be utilised to inform the development of the RICHFIELDS ontology and linkage and harmonisation methods in WP12.

From the data collected in RIMS, it would appear to be the existing AGGREGATORS that stand out as offering the greatest potential for future data linkage as they already integrate with a number of other APPS in the RIMS sample and many implement a documented API for data access. These AGGREGATORS typically offer a framework designed to integrate healthcare and fitness APPS, allowing them to work together and collate data from an individual perspective. For instance, an exercise monitoring APP and dietary tracking APP that do not offer the option of exchanging data and services through their own infrastructure could still exchange data with an AGGREGATOR application.

There are 12 AGGREGATORS that integrate with food consumption data currently logged in RIMS (see appendix 1) and it is recommended that further exploration of these is performed by Phase 3 to establish what opportunities arise for the proposed RI in linking with these types of tools. There are also a number of food consumption APPS detailed within RIMS (such as MyFitnessPal and Fitbit) which, although in themselves are not AGGREGATORS, already connect to a large array of other health and fitness APPS within their existing functionality. Further exploration of these types of APPS by Phase 3 with a view to developing a data sharing relationship and technical protocols is also recommended.

As part of the RICHFIELDS design, consideration might also be given to the development of a RICHFIELDS specific APP that could not only act as an AGGREGATOR to link with other APPS used by an individual, but also as a means of collecting additional standardised data from a cohort of individuals that are of interest for research purposes. Furthermore, the establishment of a RICHFIELD APP and cohort could facilitate a standardized EU-survey allowing researchers and policy makers to compare data at a cross-sectional and longitudinal EU level. This will allow data connectivity, standardization and harmonization of food-nutrition-health data across Europe.

Due to the fast rate of change with respect to APPS/tools launched onto the market, it is recommended that the RI/data platform develops a means of capturing and describing new APPS and AGGREGATORS that are being used by consumers in the food and wellbeing domain so that researchers can easily see what data is being created by consumers that might best inform future research designs

4.2 Business Model (WP12)

There are four possible business models that are being explored in Phase 3 (see deliverable D12.1). These vary on the degree to which the four main stakeholder groups

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

(researchers, policy makers, businesses and consumers) are considered to be customers of the proposed RICHFIELDS RI/data platform being designed:

- Alternative 1 - researchers and policy makers as customers of the RI;
- Alternative 2 - researchers, policy makers and businesses as customers of the RI;
- Alternative 3 - researchers, policy makers, and consumers
- Alternative 4 - researchers, policy makers, businesses, and consumers.

Whilst alternative 4 may be the ideal scenario, the results from Phase 1 survey suggest that from the perspective of consumers there is more perceived risk and less trust associated with sharing their data with policy makers and businesses. Therefore, the wider the data is shared outside of the publicly funded research community, the more it is likely to impact on the amount of consumer data RICHFIELDS is able to obtain with the appropriate levels of consumer consent to share. However, this could be addressed by providing different levels of consent options to consumers (potentially via the RICHFIELDS APP) which would allow consumers to specify exactly those stakeholder categories they are willing for their data to be shared with and those that they are not. Alternatively by developing relationships with the AGGREGATORS already in the marketplace the appropriate levels of overarching consent could be built into their systems and then the data shared between the AGGREGATOR and RICHFIELDS with the consumer's consent.

4.3 Governance/ethical considerations

Within their exploration of the terms and conditions for the APPS contained in RIMS, it was frequently identified by Phase 1 that although users are defined as being the owners of the data, the APP vendors have retained the right to sell the data. In the commercial world this has facilitated the selling and buying of data between organisations. The recent implementation of the GDPR seeks to offer some protection to users by ensuring that their personal data cannot be traded in this way without their consent although, it does not curtail the selling of pseudo or fully anonymised data. (*GDPR definitions: Pseudonimisation is "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information." Anonymized data is "data rendered anonymous in such a way that the data subject is not or no longer identifiable.")*

However, the recent revelations about the degree to which users' various data could be linked despite being considered anonymous/pseudonymous and the insight this could generate (e.g. Cambridge Analytica scandal) highlighted the deeper ethical and legal implications of repurposing of such data. Whilst the initial results from the user survey (UK only) indicate that due to the increased trust and reduced risk associated with Universities when compared to government or commercial organisation, the RICHFIELDS RI Governance Model may benefit from clearly identifying a University/Universities as the lead rather than a more diverse model where all stakeholders have equal leadership, the social context and legal

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

framework within which these opportunities could be harnessed are constantly changing and will continue to be dynamic for some time.

Despite what is happening in the commercial world, within the research community, there are established ethical criteria that need to be met in relation to informed consent from data owners. Without this consent the data is not acceptable for use in studies destined for scientific publication. This means that without significant scrutiny on a case-by-case basis of each existing commercial dataset, the data they hold is not readily useable by researchers. This will ultimately limit the value of any commercial data that RICHFIELDS incorporates into the proposed data platform for scientific purposes unless the issues associated with consent are fully addressed. Looking forward, the required level of ethical consent for the re-use of consumers' data across all their APPS could be obtained via the previously proposed RICHFIELDS APP. In this way direct consent could be obtained from the consumer for the use of their data either for general research or even for specific purposes and that consent held as meta-data within the RICHFIELDS data platform which from a governance perspective is the most desirable scenario.

5 Conclusions and implications

The main **recommendations** arising from this synthesis report are summarised below:

Scientific usefulness of consumer generated data

- The scientific limitations for consumer-generated purchase and preparation data identified by Phase 1 are potentially possible to overcome by linking to data from consumer-generated consumption APPS allowing a more extensive mapping of food choice and eating behaviour from preparation through to consumption for an individual.
- Unstandardized or undocumented food intake assessment procedures, data exchange protocols and formats, terms of use and privacy regulations, limit possibilities to integrate, process and share user-documented food consumption data in a scientifically robust way. Therefore best practice guidelines, quality standards and protocols are needed for the effective integration of consumer-generated food purchase, preparation and particularly composition data in a scientifically meaningful way.
- A vital source for better understanding the possible drivers and barriers for people's food purchase, preparation and consumption behaviour is likely to come from associations between these data and other relevant social, health and lifestyle data. For example, to gain domestic food purchase, preparation and consumption data from

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

dedicated APPS and link this with health and lifestyle APPS for an individual. This combined data could be further enriched with demographic, situational and social context data collected through APPS such as Facebook, Twitter and Instagram.

Technical Considerations

- In the first instance it is suggested that the RIMS database and the typologies developed by Phase 1 for the data collection in the purchase an preparation domains (see D5.1 & D6.1) should inform the development of the RICHFIELDS ontology and linkage and harmonisation methods in WP12.
- It is key to connect to developers/owners of AGGREGATORS already in the marketplace for the further development of the RICHFIELDS technical data **infrastructure** and to **facilitate** access to a wider breadth of consumer data.
- As part of the RICHFIELDS design, consideration might also be given to the development of a RICHFIELDS specific APP that could not only act as an AGGREGATOR to link with other APPS used by an individual, but also as a means of collecting additional standardised data from a cohort of individuals that are of interest for research purposes.
- It is recommended that the RI/data platform develops a means of capturing and describing new APPS and AGGREGATORS that are being used by consumers in the food and wellbeing domain so that researchers can easily see what data is being created by consumers that might best inform future research designs.

Business model, governance and ethical considerations

- Providing different levels of consent options to consumers (potentially via the RICHFIELDS APP) would allow consumers to specify exactly those stakeholder categories they are willing for their data to be shared with and those that they are not. In this way direct consent could be obtained from the consumer for the use of their data either for research, policy development or commercial activities and that consent held as meta-data within the RICHFIELDS data platform which from a governance perspective is the most desirable scenario.
- In parallel, by developing relationships with the AGGREGATORS already in the marketplace the appropriate levels of overarching consent could be built into their systems and then the data shared between the AGGREGATOR and RICHFIELDS with the consumer's consent.
- The initial results from the user survey (UK only) indicate that due to the increased trust and reduced risk associated with universities when compared to government or commercial organisations the RICHFIELDS RI Governance Model may benefit from

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

clearly identifying a university/universities as the lead rather than a more diverse model where all stakeholders have equal leadership.

Acknowledgements

This deliverable has drawn on, and in some cases utilised verbatim content from the Phase 1 deliverables. The authors would therefore like to fully acknowledge and thank the following colleagues for their contribution to this deliverable:

SP/RISE: Susanne Ekman, Erik Baderstedt, Erik Kaunisto, Haris Hondo

Wageningen University & Research: Anouk Geelen, Muriel Verain

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

Appendix

Appendix 1 – AGGREGATORS (data extracted from RIMS database)

Name	Website	Description	Tools	Sensors
Nudge Health Tracking	www.nudgeyourself.com/	Nudge automatically imports data from APPS like Moves and RunKeeper. It then allows you to enter in data with what type of food you've eaten, how much water you drink, exercise, your energy level, and plenty more. From there, Nudge tracks that data, rates how healthy you're living, and gives you recommendations for what you need to do more of. You can also connect Nudge to social feeds called "Clubs" where you can talk with other Nudge users about exercise.		Fitbit, UPÂ® â€“ Smart Coach for Health
Validic	validic.com/	Validic is a healthcare technology platform for convenient, easy access to digital health data from clinical and remote-monitoring devices, sensors, fitness equipment, wearables and patient wellness applications.	Calorie Counter by FatSecret, Calorie Counter & Diet Tracker by MyFitnessPal	Garmin Connectâ„¢ Mobile, UPÂ® â€“ Smart Coach for Health, Fitbit, Health Mate - Steps tracker & Life coach, Misfit, Striiv Activity Tracker, TomTom MySports, Polar Flow â€“ Activity & Sports Analyzing, Microsoft Band
Fitnesssyncer	www.fitnesssyncer.com/	FitnessSyncer.com is an on-line dashboard and health and fitness synchronization system that centralizes fitness data.	Beddit Sleep Tracker, Calorie Counter by FatSecret, Calorie Counter -FDDB Ext. Pro (website not english), S Health	Fitbit, UPÂ® â€“ Smart Coach for Health, Health Mate - Steps tracker & Life coach, Garmin Connectâ„¢ Mobile, Misfit, TomTom MySports
Human Api	www.humanapi.co	The Human API is a data aggregation that provides a platform that stores and abstracts data from mobile devices, wearable sensors, and other data sources. The service aims to make available all the activity data generated by health tracking devices that monitor blood glucose, wight, physical activity, caloric intake, blood pressure, and many others data types. The API aims to provide the tools to make the data meaningful and trackable so that it can be utilize in research and applications. The service uses REST calls and returns JSON.	Calorie Counter by FatSecret, Calorie Counter & Diet Tracker by MyFitnessPal, Argus Calorie Counter Diet, Activity, Step Tracker	Fitbit, UPÂ® â€“ Smart Coach for Health, Health Mate - Steps tracker & Life coach, Garmin Connectâ„¢ Mobile, Microsoft Band, Striiv Activity Tracker, Misfit

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

HealthVau	www.healthvault.com/	Microsoft HealthVault is a web-based platform from Microsoft to store and maintain health and fitness information. This personal health record system started in October 2007, the website addresses both individuals and healthcare professionals. In June 2010, Microsoft HealthVault expanded its services to include the United Kingdom. A HealthVault record stores an individual's health information. Access to a record is through a HealthVault account, which may be authorized to access records for multiple individuals, so that a mother may manage records for each of her children or a son may have access to his father's record to help the father deal with medical issues. Authorization of the account can be through Windows Live ID, Facebook or a limited set of OpenID providers. HealthVault Connection Center allows health and fitness data to be transferred from devices (such as heart rate watches, blood pressure monitors and the Withings wifi bodyscale[6]) into an individual's HealthVault record. It can also be used to find and download drivers for medical devices.[7][8] In 2014 Microsoft introduced the Microsoft Band a fitness band that is powered by Microsoft Health, a service that supports the Microsoft HealthVault for aggregation and integration of different services such as MyFitnessPal.[9][10]	20/20 LifeStyles	Microsoft Band, Fitbit, Health Mate - Steps tracker & Life coach
Healthgraph	runkeeper.com/developer/healthgraph/	The Health Graph captures a user's fitness activities and health measurements over time. Each activity and measurement is a time-stamped node, and each node is linked to those other nodes closest in time. Devices can add new nodes to the graph by submitting health measurements; applications can follow the links in the graph from node to node to extract a user's progress towards his or her fitness goals. The Health Graph API presents the Health Graph as a collection of Web-based resources, one for each node of the graph. The Health Graph API also makes available resources that enable quick navigation of the Health Graph, as well as provide access to a user's profile, sharing and display settings, and friend information.	Lifesum - Healthier eating, better living, Lose It! â€“ Weight Loss Program and Calorie Counter, MealLogger, Calorie Counter & Diet Tracker by MyFitnessPal, FoodPrintâ„¢ Diet by Nutrino	UPÂ® â€“ Smart Coach for Health, Health Mate - Steps tracker & Life coach

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

GoogleFit	developers.google.com/fit/	Google Fit is a health-tracking platform developed by Google for the Android operating system. It is a single set of APIs that blends data from multiple APPS and devices. Google Fit uses sensors in a user's mobile device or activity tracker to record physical fitness activities (such as walking or cycling), which are measured against the user's fitness goals to provide a comprehensive view of their fitness.	Lose It! â€“ Weight Loss Program and Calorie Counter, Calorie Counter & Diet Tracker by MyFitnessPal, Noom Healthy Weight Loss Coach, Calorie Counter by FatSecret, Lifesum - Healthier eating, better living, FitWell Fitness, Health, Diet, Poundaweek - Calorie Counter, Track - Calorie Counter, HealthifyMe Weight Loss Coach, Hydro Coach - drink water, Mevo - Weight Loss & Fitness, Fitatu Calorie Counter, Weight Loss & Fitness Program, Diet Watch, Ultimate Food Value Diary Plus - Diet & Weight Tracker, Water Drink Reminder, Kiqplan, your complete health & fitness trainer, Calories Carb Prot Fat Counter	Health Mate - Steps tracker & Life coach, Record by Under Armour, connects with UA HealthBox, Pebble, Garmin Connectâ„¢ Mobile, Polar Flow â€“ Activity & Sports Analyzing
S Health	shealth.samsung.com/	Manage your fitness activities, track your weight and monitor your diet using Samsung Galaxy and Gear devices. You can manage your diet, health and exercise data from each App in one place.	Hydro Coach - drink water, Lifesum - Healthier eating, better living, Noom Healthy Weight Loss Coach, Water Drink Reminder, Healthy 365, Kiqplan, your complete health & fitness trainer	Garmin Connectâ„¢ Mobile, UPÂ® â€“ Smart Coach for Health, Misfit, Fitbit, Samsung Gear
IF	ifttt.com/	IF connects the APPS you love. Create simple connections between APPS like Facebook, Dropbox, Instagram, Twitter, and Gmail, as well as devices like your iPhone, Nest Thermostat, Fitbit, and Philips Hue.	Tesco Groceries, Instagram, Facebook, Twitter, Foursquare, Lifelog	UPÂ® â€“ Smart Coach for Health, Fitbit, Health Mate - Steps tracker & Life coach, Misfit
HealthKit	developer.apple.com/healthkit/	HealthKit is the accompanying developer application programming interface (API) included in the iOS SDK (Software Development Kit) for the Mac. It is used by software developers to design applications that have extensibility and that can interact with the Health application on iOS. The API allows other applications with the user's permission, to access health data. For example, a blood pressure application would share information with a doctor, or a nutrition application could inform a fitness application how many calories a user consumes each day.	Calorie Counter & Diet Tracker by MyFitnessPal, Noom Healthy Weight Loss Coach, Activ8rlives Health Monitoring and Food Diary App, Sugar Sense - Diabetes App, Blood Sugar Control, and Carb Counter, etc etc	Polar Flow â€“ Activity & Sports Analyzing, Misfit, Record by Under Armour, connects with UA HealthBox, Pebble, Microsoft Band, UPÂ® â€“ Smart Coach for Health, Garmin Connectâ„¢ Mobile, Health Mate - Steps tracker & Life coach

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

Tactio	www.tactiohealth.com	Tactio Health App helps you track and manage a wide range of health data from simple manual logging or syncing with connected health APPS and medical devices. It is packed with science-based rules that provide you a reference range on every data that it gets for weight, steps, nutrition, activity, sleep, mood, blood pressure, pulse, glucose, cholesterol, temperature and oximetry. HEALTH DASHBOARD See all your health trends, risks and indicators (such as your heart age) in a color coded dashboard from science-based rules. This customizable dashboard makes it possible for you to have an easy and quick overview of what you need to track, manage, and monitor in order to enjoy a healthy lifestyle or keep a chronic disease in control. CONNECTED HEALTH Embrace self-care with the Internet of Things (IoT) using popular cloud based apps such as Fitbit, MyFitnessPal, Garmin, & Apple Healthkit and Bluetooth medical devices such as Welch Allyn, Roche, Nonin & A&D Medical. HEALTH RISKS Your health risks for cardio-vascular, metabolic syndrome, and type 2 diabetes are computed with every new data. Parameters help you quickly see the health benefit of various lifestyle changes like “what if I stopped smoking?”. CONNECTED LOGBOOK See all your health data chronically ordered and colored from science-based rules. Filter your data in a couple of taps to see only the one you’d like to focus on. Generate a PDF report that can be shared to your health care professional. HEALTH COACHING Coaches will send you science-based feedback about your healthy behaviors, trends and chronic disease control. The coaches remind you about things you should do or measure, best preventive health practices and also analyze the data you log to coach you into making small behavior changes to improve your health or keep a health condition under control such as obesity, diabetes, hypertension, atherosclerosis, COPD & pregnancy.		Fitbit, Garmin Connectâ„¢ Mobile, UPÂ® â€“ Smart Coach for Health, Health Mate - Steps tracker & Life coach

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

Zenobase	zenobase.com/#/	Zenobase is service that provides users with the ability to store data, as well as aggregate and visualize personal time-series data. The Zenobase API allows users to store and retreive data. The API uses REST calls and returns JSON. The API exposes 4 sections of the API: the authentication, the buckets of data, events, and common data. Each of these categories includes a number of resources. An account is required with service.	Foursquare	UPÂ® â€“ Smart Coach for Health, Fitbit, Health Mate - Steps tracker & Life coach, Microsoft Band, Misfit

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